NPL REPORT IEA 6

ESTIMATING THE PRICE ELASTICITY OF DEMAND FOR NPL’S SERVICES

MIKE KING, EUGENIO RENEDO

JULY 2020

ESTIMATING THE PRICE ELASTICITY OF DEMAND FOR NPL'S SERVICES

DR MIKE KING, EUGENIO RENEDO
Analysis and Evaluation, Strategy Directorate

Abstract

This paper applies standard panel data analysis to a country-level panel dataset to estimate the price elasticity of demand for NPL's services. The data used consists of NPL internal administrative information, as well as data coming from reliable and widely used publicly available databases such as the World Bank's World Development Indicators database, or the CEPII database. The analysis finds that the price elasticity of demand for NPL's services is greater than one, meaning that these are elastic goods for which the quantity demanded will vary more than proportionally if the price changes.

© NPL Management Limited, 2020

ISSN 2633-4194

https://doi.org/10.47120/npl.IEA6

National Physical Laboratory

Hampton Road, Teddington, Middlesex, TW11 0LW

Extracts from this report may be reproduced provided the source is acknowledged and the extract is not taken out of context.

Approved on behalf of NPLML by
Fiona Auty, Head of Government Relations

Contents

List of Acronyms

A4I	Analysis for Innovators
BIMP	The International Bureau of Weights and Measures (usually referred by its French initialism, Bureau International des Poids et Mesures) is an intergovernmental organisation through which member states act together on matters related to measurement science
CEPII	The Centre d'Etudes Prospectives et d'Informations Internationales is a leading French centre for research and expertise on the world economy.
CMC	Calibration and Measurement Capabilities
NMI	National Measurement Institute
NMS	National Measurement System
NPL	National Physical Laboratory
PPP	Purchasing Power Parity

Preface

This paper has been prepared by NPL's Analysis and Evaluation team to support the evidence this organisation is submitting for the 2020 Comprehensive Spending Review. This document is addressed to NPL's government owner: The Department for Business, Energy & Industrial Strategy. Its objective is to support the case NPL is presenting to justify the benefits of its public funding.

Executive Summary

This work applies a two-step Heckman selection model to a country-level panel dataset to estimate the price elasticity of demand for NPL's services. The data used consists of NPL internal invoicing information, as well as data on financial and demographic variables coming from reliable external sources – namely, the World Bank's World Development Indicators database, the *Centre d'Etudes Prospectives et d'Informations Internationales*, the Bureau of Weights and Measures, and the United Nations. The analysis finds that:

- The central estimate for the price elasticity of demand is 1.24 with a 95% confidence interval given by $[1.09\,, 1.39]$.
- The four regressors in our model (the national price level, GDP per capita, distance and the CMC of the local NMI) turned out to be highly significant.

The rest of the paper is organised as follows. Section 1 provides context about NPL and outlines the rationale for its public funding; it also makes the case for estimating the price elasticity of demand using a country-level panel. Section 2 depicts the conceptual framework that synthetises the basic economic theory that explains NPL's international sales of services and identifies the variables of interest. Section 3 discusses the data and the variables used by showing some descriptive statistics. Section 4 provides a discussion of the model selection criteria and presents the estimation results. Section 5 shows how the results found in the econometric analysis can be used to estimate the welfare generated by NPL. Section 6 concludes.

1 Introduction

1.1 The National Physical Laboratory

NPL is a government owned and funded national laboratory that specialises in metrology (the science of measurement). It has over 380 state of the art laboratories and more than 800 scientists and engineers, as well as 200 co-supervised postgraduate students working in partnership with industry and academia. NPL's goal is to generate welfare for the UK's society. To that end, NPL carries out a multitude of activities across a wide range of areas – from quantum sensing and composite materials, to radiotherapy and emissions monitoring. In a nutshell, NPL's value chain can be summarised as follows. Firstly, NPL conducts fundamental research and performs international measurement comparisons that generate articles in peer-reviewed scientific journals. This enables the development of cutting-edge measurement capabilities that support the creation of primary standards and state-of-the-art instrumentation. This expertise is then used to deliver calibration, testing, and training services to private businesses, hospitals, and universities. In addition, NPL works closely with Innovate UK to offer grant-funded collaborative R&D projects which involve many firms and research organisations.

NPL is part of the National Measurement System (NMS). The NMS is the technical and organisation infrastructure which ensures a consistent and internationally recognised basis for measurement in the UK. It has two central objectives:

i. To enable individuals and organisations in the UK to make measurements competently and accurately and to demonstrate the validity of such measurement

ii. To coordinate the UK's measurement system with the measurement systems of other countries.

The following figures give a clearer picture of NPL[1]:

- 820 scientific and technical staff, as well as 280 administrative and managerial staff.
- A turnover of around £90m; £57m of that revenue in annual NMS[2] funding.
- Around 350 articles in peer reviewed journals each year; also, its scientists perform around £30m of public research work each year[3].
- £13m of revenue from sales of measurement services. The R&D performed by NPL supports the introduction of new and improved calibration services, whose benefits fan-out down the calibration chain.
- Sells services to around 500 UK-based firms each year. And, the lab's scientists collaborate on R&D projects with around 200 UK-based firms each year.

[1] These figures are as of 2019. They differ to the ones used in this study which relates to the period 2010-2017.
[2] The National Measurement System (NMS) is a network of national laboratories and processes that provide measurement standards and calibration testing facilities in the UK. It maintains the measurement infrastructure, represents the position of UK measurement internationally and influences the development of standards.
[3] Based on tax credits claims information.

NPL seeks to fulfill its objective of providing excellent metrology science to generate welfare for the UK. Hence, it is essential for NPL to know its impact. For that reason, NPL commissioned:

- A survey asking its users about the sales of new products that they feel would not have been achieved without the support of NPL and the other NMS laboratories. This survey found that users of the NMS laboratories believe that without NPL's support, their total annual sales of new products would decrease by at least £470M. Furthermore, they believe that about £2bn worth of new products might be at risk without this support. The self-reported nature of these estimates may make the exact size of these benefits doubtful, but it still provides evidence that such benefits exist.
- An independent econometric study by that found that companies supported by NPL grow more rapidly than unsupported comparators - on average, supported companies have around 20 additional employees 2-3 years after working with NPL, when compared to a matched control group of similar unsupported ones.

1.2 The public funding of NPL

The economic rationale for the existence of a publicly funded organisation like NPL is that measurement R&D is subject to market failure. Indeed, the private investment needed to generate innovative measurement capabilities, will always be below the socially optimal level. This occurs because the benefits that measurement R&D generates will always spill over to firms who did not contribute, and this creates a strong incentive to free ride. The problem is particularly acute in the case of the R&D that the NPL undertakes, because advances in metrology tend to have applications across many sectors. It is this wide applicability that makes the development of these new tools and techniques particularly susceptible to free riding. Consequently, it is argued that measurement should really be seen as a public *infratechnology*, that is, a technology that provides tools and techniques which can be widely applied across a number of sectors to enable further innovation. In short, NPL and its partners' scientific work generates a pool of knowledge that can be accessed and used by any firm. This fact carries a strong incentive to free ride, and thus, there is a clear need for public funding to complement measurement R&D funded through private spending.

Another key argument for publicly supporting a specialist laboratory like NPL is that the kind of metrology research it conducts requires the setting up of large facilities. In such cases, the fixed costs could be so high that they exceed the private gains to any one company. Therefore, the facility would never be developed on the basis of individual private funds alone, despite the total benefits from the capability outweighing the cost.

Lastly, there is an efficiency justification that supports the idea of a publicly funded metrology laboratory like NPL. Indeed, the high cost and difficulty of maintaining primary standards makes the calibration chain very efficient. NPL supplies a costly high-level calibration service to a commercial laboratory, which then calibrates the instruments of a vast number of users without the need for the calibration laboratory to establish their own primary standard.

1.3 Estimating the price elasticity of demand

Knowing the responsiveness of demand to changes in prices is essential for any business to assess its potential for revenue growth. Moreover, since NPL is a publicly funded organisation, a reliable estimate of the price elasticity of demand constitutes the cornerstone for further work aimed to predict the effect of increasing public support.

The analysis outlined in this paper makes use of a country-level panel dataset involving more than 100 countries over a period of 17 years (2001-2017) to estimate the price elasticity of demand for NPL’s measurement services. There are several reasons why this country-level approach was chosen to estimate the price elasticity of demand:

- Over half of NPL’s sales are made overseas. This means that a fairly significant fraction of NPL’s income is subject to international trade forces.
- Analyses of the market for high accuracy measurement services reveal that relative prices with respect to other NMIs have experienced hardly no variation over the last two decades. This means that any trend in prices is common to almost every NMI, and that price-driven substitution effects must have been almost insignificant. On the other hand, we know that, in an international trade context, the perceived price by consumers is different from the selling price. Indeed, the perceived price is also influenced by the consumer’s local purchasing power parity and the exchange rates between the local and the selling currency. Thus, variations in inflation and exchange rates can be exploited to measure the price elasticity of demand for NPL’s services.

2 Theory

This section outlines the conceptual framework that synthetises the basic economic theory that explains NPL's international sales of services. To that end, the global market for high accuracy measurement services is described first. Next, a few simple micromodels are presented. These capture the intuition behind the role played by the various variables of interest in the subsequent econometric analysis. Hence, the objective of these models is not to provide a comprehensive analytical framework to be econometrically tested, but to show the mechanisms whereby several key variables affect international sales of measurement services.

2.1 Key variables in the market for high accuracy measurement services

NPL and other NMIs offer high quality measurement services. These vary from one NMI to another. In particular, the accuracy delivered depends on how developed the measurement infrastructure is in each country. Although all users in the market value accuracy (since it ultimately constitutes valuable information[4]), it is not equally appreciated by all of them; for instance, a manufacturer of aircraft engines is more likely to require greater accuracy than a mass producer of chocolate bars.

Another important aspect of the market for high accuracy measurement services is the time it takes for an NMI to provide the service. Indeed, particularly in the case of calibration services, the assets to be calibrated are extremely valuable. This implies that, while waiting to be being calibrated, the instruments do not generate value for their owners. Thus, queues have a substantial effect on consumption patterns.

On the other hand, the global market of high accuracy measurement services is not frictionless – it is subject to significant transaction costs. Particularly two elements play a major role: geographical proximity and imperfect information. The former entails that transport costs make services from distant NMIs less attractive; the latter causes the existence of search costs, that is, the time and money spent by users who research the market to find a service that satisfies their needs.

Therefore, based on this theoretical framework, we can think of a worldwide, vertically differentiated, spatial market for high accuracy measurement services. Total demand in this market is assumed to depend mainly on:

- The price of the service: Like most goods, high accuracy measurement services are affected by the law of the demand which states that conditional on all else being equal, as the price of a good increases, the quantity demanded decreases. However, we are dealing with a multi-country, multi-currency, international trade framework; thus, exchange rates and monetary effects play a substantial role. The perceived price is influenced by the user's local purchasing power and the exchange rate.
- The time to delivery: Since measurement is at the heart of many production processes, the longer it takes for the customer to get the service, the higher the opportunity cost.
- The geographical proximity between the NMI and the user: Transport costs disincentivise sales abroad – the further the distance, the higher the transportation and

[4] Accurate measurement is often crucial to R&D activities to develop new products (and therefore the possibility of increasing market power and commanding a price premium), or new processes that typically entail a cost-saving improvement to the production process, which leads to a greater efficiency and an increase in the profit margin.

the insurance cost. Moreover, highly accurate measurement instruments are very sensitive, and their specification may deviate due to transportation. These two facts create a bias towards consumption from closer NMIs.

- The variety of services provided by the home NMI: The market of high accuracy measurement services is a very specialised market. Hence, users face substantial search costs. Rational consumers will continue to search for a better product until the marginal cost of searching exceeds the marginal benefit. Thus, users will be more inclined to buy services from their local NMI (provided that they can find a service that meets their needs), rather than a distant foreign NMI. This comes as no surprise, since search costs are lower at the local level because there are fewer barriers to knowing the NMI's portfolio (e.g. common language, the possibility to visit the facilities…).

The framework outlined above synthetises the main forces driving the decision of users worldwide regarding the purchase of high accuracy measurement services from NMIs. However, one typical element affecting demand is missing: real income. This assumption implies that the income elasticity of high accuracy measurement services is thought to be close to zero. The argument for this is that the quantity demanded of these services by any specific firm, tends to be constant over time. In other words, high accuracy measurement services are a basic need for most of the users. Indeed, there are effectively two reasons for demanding these services: either because it is a requirement of the production process or because it supports R&D activities. The former is consistent with the zero-income elasticity of demand assumption. The customer is forced to acquire the services of an NMI instead of those of a commercial laboratory because that level of accuracy is fundamental to run its operations. The latter is not as straightforward. Although the empirical evidence shows that R&D activities are highly procyclical , and therefore linked to income growth, it could be argued that research spending comprises a very small fraction of the company's resources. Moreover, measurement is just one of the activities within an R&D project. Hence, measurement R&D spending must be negligible compare to measurement spending in productive related activities. Consequently, the zero-income elasticity of demand is considered a valid approximation.

In addition, the forces that drive demand may not be the same when we aggregate the units of analysis. In other words, it is not the same to analyse the factors that determine the demand of a single user, than to study those at the country level. Notably, the number of companies that need high accuracy measurement services within a country, must be a good predictor of the quantity sold. In section 4 we will control for this by including GPD per capita as a time-varying covariate. The intuition behind this is twofold. On the one hand, the richer a country is, the more sophisticated tends to be its economy, and the higher the demand for high accuracy services. On the other hand, the number of companies fluctuates with income growth, *i.e.* the number of companies is certainly linked to the business cycle. Arguably, the first aspect (the relationship between economic sophistication and the need for high accuracy measurement services) is related to GDP, whereas the second one (the link between the volume of this kind of services sold and the number of companies), pertains to oscillations around the mean value of GDP per capita. In any case, both situations are addressed by adding the GPD per capita time-varying covariate in our regression model. Moreover, as a by-product of adding GDP per capita in the regression model, we are also accounting for any income effect, should our assumption of zero-income elasticity not be completely valid.

Lastly, note that queue times are impossible to account for in our analysis. Simply this information is not publicly available, since this consists of internal information of each NMI. Section 4.1 deals with the econometric issues around the lack of data for this relevant variable and details the assumptions made to work around it.

The rest of the section provides further theoretical basis to understand the relationship of sales with the rest of the magnitudes identified as drivers of the demand for high accuracy measurement services. A series of micromodels are presented in order to show the effect of the national price level, the distance, and the services provided by the local NMI on the international market of high accuracy measurement services.

2.2 National price level and elasticity of demand

Perceived price is key to understand consumption patterns in international trade; thus, international trade is driven by purchasing power parities and the exchange rates. A concise and convenient way to express the effect of these variables is the ratio between the two – which is commonly known as the national price level. The national price level allows us to compare the cost of the same bundle of goods across different countries.

We define the national price level as:

$$L = \frac{S}{r} = \frac{p'_h/p'}{r} \tag{2.1}$$

where S is the purchasing power parity relative to the UK, that is, the price ratio of the same basket of goods in local currency units of the buyer's *home* country, p'_h, and in sterling pounds in the UK, p'; r is the exchange rate expressed in local currency units per pound – the superscript $'$ in both pricing variables in equation 2.1 symbolise that both relate to the *numeraire good*, i.e. the representative basket of goods typically bought by the firm; this will play a role in the model below. Unsurprisingly, if the buyer's home country is in fact the UK, then L is equal to one.

From the definition of the national price level it follows that increases in the national price level can be driven either by a higher domestic inflation or a stronger local currency unit. Hence, a positive relationship between an NMI's international sales and the national price level is expected. To show this, the micromodel below is developed to find the analytical expressions that links the quantity sold of high accuracy measurement services and the national price level.

A rational agent is only interested in changes in the perceived price, no matter whether they come from fluctuations in selling prices or in the exchange rates. To show this, let's start by considering the payoff function[5] of the typical user of high accuracy measurement services[6]. Naturally, the money spent on measurement services is just a tiny fraction of a company's total spending. Hence, the following quasi-linear payoff function serves as a good first approximation of the value obtained by the user:

$$\Pi = cQ^{\theta} + X \tag{2.2}$$

[5] This implicitly states the productive nature of high-accuracy measurement services which provide valuable information that support the development of new products or processes which ultimately yield higher profits.

[6] Without any loss of generality, for the sake of simplicity, in this model we consider the choice faced by a rational agent when purchasing measurement services either in their local market or in the UK. Note that the choice of the UK is arbitrary, but irrelevant for the validity of the model. The reason why we set the choice between the UK and the local country, is because it simplifies the explanation of the model and it links with the later empirical analysis in this paper, in which this model is used to estimate the price elasticity of demand for NPL's measurement services.

Where c and θ are positive constants such that $c > 0$ and $0 < \theta < 1$, Q is the quantity demanded of high accuracy measurement services[7], and X is the number of units of the numeraire good. This functional form is considered suitable to model the firm's payoff function because spending on measurement services is typically quite small compare to the rest of spending by the firm. Therefore, the effect on the marginal utility of money is negligible. Moreover, another key property of the quasilinear function, is that the Marshallian demand does not depend on income, which is consistent with our previous assumption of zero income elasticity.

The agent's spending limit is given by the budget constraint:

$$M = rpQ + p'_h X \tag{2.3}$$

Where M is the income of the agent, r is the exchange rate, p is the market price of the high accuracy measurement service in the UK and p'_h is the price of the numeraire in the home country[8]. Note that this specification means that M and p'_h are measured in the customer's country currency, and p is expressed in sterling pounds.

If we substitute 2.3 in 2.2 and maximize the resulting objective function, we obtain the Marshallian demand function:

$$Q^* = \left(\frac{c\theta p'_h}{rp}\right)^{\frac{1}{1-\theta}} \tag{2.4}$$

Now, this equation can be expressed in terms of the national price level and UK prices using equation 2.1:

$$Q^* = \left(\frac{c\theta p' L}{p}\right)^{\frac{1}{1-\theta}} = \left(\frac{c\theta}{\mathcal{P}}\right)^{\frac{1}{1-\theta}} \tag{2.5}$$

Where $\mathcal{P} = p/p'L = rp/p'_h$ is the perceived price by the customer – i.e. the number of units of the numeraire good that the firm can buy in the home country for the amount it would spend on one unit of the measurement service in the UK.

If we take the logarithms at both sides of equation 2.5, we get:

$$\ln Q = A - \frac{1}{1-\theta} \ln \mathcal{P} \tag{2.6}$$

Where $A = \frac{1}{1-\theta} \ln c\theta$ is constant.

We can derive both sides of equation 2.6 to get:

[7] Implicitly, the heterogeneous nature of measurement services is being ignored here. However, for the purposes of the model and the empirical analysis later on, we assume that in effect what all NMIs sell is the time of highly trained scientists and engineers in metrology. More on this on section 3.1.

[8] The price notation is such that subscripts denote if the price is referred to the UK (no subscript) or the buyer's home country (subscript h), and the superscripts represent if the price is related to the measurement service (no superscript) or to the numeraire good (superscript $'$)

$$\frac{\Delta Q}{Q} = -\frac{1}{1-\theta}\frac{\Delta \mathcal{P}}{\mathcal{P}} \tag{2.7}$$

Where the percentage change in the perceived price is given by:

$$\frac{\Delta \mathcal{P}}{\mathcal{P}} = \frac{\Delta p}{p} - \frac{\Delta p'}{p'} - \frac{\Delta L}{L} \tag{2.8}$$

Now, NPL's measurement services prices have roughly moved directly in relation to the UK general price level (i.e. the price of the numeraire good). Consequently, both percentage changes in prices in equation 2.8 cancel each other out and we are left with the percentage change in the national price level. In other words, although by using the national price level as a proxy for the perceived price we might be suffering from measurement error, this is likely to be quite small because relative changes in the perceived price are very similar to relative changes in the national price level

Lastly, if we expand the equation 2.6 using the definition of the perceived price $\mathcal{P}$, we obtain:

$$\ln Q = A - \frac{1}{1-\theta}\ln p + \frac{1}{1-\theta}\ln p' + \frac{1}{1-\theta}\ln L \tag{2.9}$$

Therefore, this model yields two important conclusions:

- The price elasticity of demand for NPL's measurement services is equivalent in magnitude (but opposite sign) to the elasticity with respect to the national price level.
- Relative changes in the perceived price are mostly driven by relative changes in the national price level because the price of measurement services have approximately been in sync with the general price level.

These two facts will be exploited in the empirical analysis in section 4.

2.3 The effect of distance in international trade

The gravity model of trade has been very successful when ordering the observed variation in economic interaction across space in trade flows. The good fit and the tight clustering of coefficient estimates in the empirical literature, suggest that some underlying economic law must be at work. Although the driving forces of international trade have been a vivid debate since the emergence of Ricardo's comparative advantage theory, the gravity model has often been used to test hypotheses rooted in many different economic theories.

Normally, in econometric applications the gravity model of trade is specified as follows:

$$X_{ij} = A \cdot \frac{E_i^{\gamma_1} \cdot E_j^{\gamma_2}}{D_{ij}^{\gamma_3}} \cdot \vartheta_{ij} \tag{2.10}$$

where X_{ij} represents the volume of trade from country i to country j, A is a constant, E_i and E_j represent some proxy variable of the economic size of the countries, typically their gross domestic products, D_{ij} denotes the distance between the two countries, and ϑ_{ij} represents an error term with mean 1.

The most common approach to estimate equation 2.10 is to take the logarithm of both sides to construct a log-log model that can be estimated by OLS:

$$\ln X_{ij} = \gamma_0 + \gamma_1 \ln E_i + \gamma_2 \ln E_j + \gamma_3 \ln D_{ij} + \eta_{ij} \quad (2.11)$$

where $\gamma_0 = \log A$ and the negative sign before γ_3 has been included in the coefficient.

If we consider the worldwide trade pattern of just one country, the model becomes simpler because all the observations are referred to that country (effectively, we drop one of the indexes in equation 2.11, so it becomes):

$$\ln X_i = \beta_0 + \beta_1 \ln E_i + \beta_2 \ln D_i + \varepsilon_{ij} \quad (2.12)$$

Where X_i represents the volume of trade from country i to the country of reference, D_i denotes the distance between the two countries, and $\beta_0 = \gamma_0 + \ln E$ with E representing the proxy variable of the economic size of the country analysed.

2.4 Search costs and the home bias effect

Most markets in the economy are far from being commoditised. This is especially true for the market of high accuracy measurement services, where the prevalence of product (service) differentiation is caused by an ample heterogeneity in user preferences and the wide range of measurement capabilities of the different NMIs. This differentiation provides NMIs with profitable opportunities and makes them face important strategic decisions in terms of how and when specialise in certain areas of metrology. The acute differentiation in the market also leads to search frictions for users, since they have to consider not just price, but a large number of characteristics of the various services on offer and how these fit their needs. Hence, users incur a cost of resources (namely money and time) when having to find and compare the different options at their disposal. Thus, the decision to purchase a specific service or to keep researching the market is mostly determined by two factors: search costs and the suitability of the options already considered.

The behaviour of the user of high accuracy measurement services can be rationalised by considering the following setup. Let's assume a differentiated market of high accuracy measurement services where users have different tastes (horizontal differentiation[9]). In the market, any potential buyer knows all the NMIs, but does not know the characteristics of the services they offer (for the sake of simplicity, price is considered as one of the elements pondered by the user). Users can gather information by sequentially and randomly sampling the NMIs in the market (the assumption of random search suggests that they know little about the market before committing to search; thus, it is consistent with the assumption that users know all the NMIs in the market, but they do not know the characteristics of the services). The search process has associated a cost (typically it consumes time and resources). However,

[9] Horizontal differentiation refers to distinctions in products that cannot be easily evaluated in terms of quality. This stands in contrast to vertical differentiation, where the distinctions between products are objectively measurable and are based in the products' respective level of quality. Therefore, in a horizontally-differentiated market, in general consumers will buy different products even if these have the same price, since consumption decisions are motivated by individual preferences. In the case of vertically-differentiated markets all consumers agree on the preference order of goods in the market. Hence, consumption decisions are determined (fundamentally) by their budget constraints.

knowing the characteristics of the services of the first NMI comes at no cost. This assumption models the fact that it is easier to know what the home NMI offers. Lastly, at any point in the search process, users can always fall back and buy the variety of any of the NMIs already sampled at no additional cost.

Let's start by considering the net present value of future profits[10] that a user of high accuracy measurement services can obtain thanks to those. The objective function of a user j is given by:

$$\Pi_j(x_0; i) = x_0 + v_i \tag{2.13}$$

Where x_0 is the quantity of the numeraire good[11] the user j has, and v_i is the value the user gives to a unit of brand i.

Given that users are assumed to maximise the net present value of profits, they will try to find the variety for which v_i is maximal. In other words, they will try to find the service of the NMI that best suits them. In addition, in order to account for different tastes within users, it is assumed that the valuations v_i are realisations of an independent and identically distributed random variable with distribution function G. Without loss of generality, we can assume G is given by a continuous uniform distribution (rectangular distribution) with distribution support $[0,1]$.

A rational user will stop searching for another service when the marginal cost of searching exceeds the marginal benefit. Let's consider a buyer who has already sampled some NMIs and attaches a valuation $v' \epsilon [0,1]$ to the best match found. The marginal benefit of keep searching (i.e. sample one more NMI) is given by the expected benefit of doing so minus the valuation of the best match up until that point:

$$\begin{aligned} b &= \mathbf{E}[v|v > v'] \cdot P(v > v') + a \cdot P(v < a) - a = \int_{v'}^{1} (v - v')dv \\ &= \frac{1}{2}(v' - 1)^2 \end{aligned} \tag{2.14}$$

The marginal cost of one more sample is given by the search cost, s. Hence, the user will stop searching when the marginal benefit given by 2.14 is equal to the search cost:

$$\frac{1}{2}(v^* - 1)^2 = s \Leftrightarrow v^* = 1 - \sqrt{2s} \tag{2.15}$$

Therefore, the user will stop searching when he finds the service of an NMI that provides a value equal to or greater than v^*. Note that this means that the higher the valuation the user assigns to the initial NMI (remember that the information about this first NMI comes at no cost), the less likely is the participation in the international market. In other words, the better the local

[10] As in section 2.1, the fact that the objective function for the representative agent is the net present value of future profits suggests the productive nature of high-accuracy measurement services.
[11] The inclusion of the numeraire good simply seeks to resemble the linear functional form of consumer's utility that is most commonly used in the economic literature. The numeraire does not play any role in the model. It only denotes the fact that a separate and additive relationship is assumed between all the goods that the user can acquire. Effectively, this implies that the user looks for the variety that best satisfies his needs as a standalone service, without considering any relationship of complementarity or substitution with the rest of the goods.

NMI meets the needs of the user, the less likely it is that the user will decide to survey the portfolio of foreign NMIs.

3 Data

The following section presents the data used to construct the variables used in the econometric analysis. First the data sources and the variables are presented. Then the methodological issues around those variables are addressed.

3.1 Variables

This study uses NPL's internal managerial data as well as data coming from several external sources. Table 1 summarises the variables and the sources of data used in the econometric analysis:

Variable	Units	Symbol	Source	Description	Magnitude
Number of invoices normalised by the population of the country	invoices per inhabitant	*I*	NPL	Yearly number of invoices issued to all users in a country divided by the population of that country.	Quantity sold
National price level	. (index)	*PL*	World Bank	Relative national price with respect to the UK.	Price
GDP per capita	£	*GDP*		Market value of all the final goods and services produced in a year per person in millions of pounds.	Income
Distance	km	*D*	CEPII	Distance in km between London and the most populate city in the user's country.	Geographical proximity
Calibration and measurement capabilities	. (index 0-100)	*CMC*	BIPM	Number of calibration and measurement capabilities of the home NMI of each country.	Metrology expertise
Population	inhabitants	*POP*	United Nations	Population in millions of people at 1 July 2019.	Population

Table 1: Variables, data sources and magnitudes.

There are several methodological difficulties when it comes to associating suitable proxies with the magnitudes identified in Table 1. The rest of this subsection analyses those issues and discusses potential solutions.

3.1.1 Quantity sold

Although NPL's portfolio is quite varied, roughly speaking it can be considered to offer one unique good: the time and the expertise of highly trained scientists and engineers. In that regard, NPL can be thought to operate much like a professional services firm which sells the time and the knowledge of its workforce. However, in other aspects NPL is far from resembling a typical professional services firm. In particular, there is a fundamental difference when comparing both business models – say between a law firm and NPL. For the former, its core business consists of one sole activity: representing its clients in court and providing them with legal advice. For NPL it is not as straightforward, because NPL's staff need to maintain the measurement capabilities required to deliver high accuracy measurement services and engage in cutting-edge collaborative R&D projects; and this task requires a significant portion of their time. Hence, unlike a law firm, NPL must preserve a knowledge stock that depreciates over time in order to meet the requirements of its users. To do so, NPL's staff carry out a wide variety of activities. These include conducting international key comparisons, participating in proficiency testing schemes, maintaining UKAS accreditation for calibration and testing services, running audits, contributing to standards and protocols, and performing research that generates articles in peer-reviewed scientific journals. This knowledge is then used to meet the needs of users. In this sense, NPL could be considered to resemble an orchestra for which the day of the concert is only the tip of an iceberg of constant work over the months and years. The show would not be possible without the previous effort put into practice. Similarly, for NPL both activities (the maintenance of capabilities and the delivery of services) are deeply entwined. Thus, it is not easy to discriminate the time spent by the staff on one activity or the other.

On top of that, NPL's strategic managerial approach focuses on achieving specific revenue targets for each scientific group. The way this target is achieved varies from one group to another, and over time. This means that the decisions taken by the group leader on how to distribute the time of scientists and engineers in the group may be influenced by many factors, such as long-term perspectives or the idiosyncrasy of the area of metrology. Effectively, this means that NPL does not have timecard data broken down by customer. Therefore, any *ex-post* consideration on the distribution of the time of scientists and engineers would not have enough comparability between groups and will lack temporal consistency.

Since the ideal variable to proxy the quantity sold (the time distribution of the scientific staff among jobs) is not available, some other proxy is needed. There are two other alternatives available to us: the number of invoices issued, and the income generated[12]. Undoubtedly, both must hold a positive relationship with the time spent by scientific staff. However, the number of invoices variable is the preferred option. This is because income is related to quantity sold through price ($R = p \cdot Q$); and prices vary substantially between jobs despite the service delivered being the same. This is because NPL's prices are set on a cost-plus basis[13]. This means that, although the services provided are fairly standardised, these are not always delivered by technical staff of the same grade – i.e. depending on availability of the workforce sometimes the service may be delivered by some more senior scientists or engineers, and this effectively translates into prices. Thus, income does not relate well to time spent on each job.

Nevertheless, this is a fundamental choice of the analysis that requires further discussion. To that end, a robustness check has been carried out in Annex A. The test consists of executing the same model using income as the dependent variable instead of the number of invoices. We conclude that both models produce very similar estimates; however, the preferred

[12] Note that in reality any choice for the dependent variable is normalised by population. This ensures comparability between countries.
[13] For most services provided by NPL there is no competitive market in the UK; which is why NPL deliver those services. This means that there is no competitive process driving prices and therefore, NPL set prices on a cost-plus basis.

specification is the one that considers the number of invoices as the dependent variable because we have the aforementioned reasons to believe that these estimates are slightly more accurate.

3.1.2 Metrology expertise of the home NMI

Another methodological issue revolves around the way to model the CMCs of a country. BIPM reports a comprehensive table that details the number of CMCs by country broken down into different metrology areas. These areas encompass many different services depending on the physical quantities they measure; for example, *Acoustics, Ultrasound and Vibration*, *Photometry and Radiometry* or *Chemistry*. There are mainly two possibilities to assess the CMCs of an NMI based on the information reported in the table. Both have been tested and compared.

The first one adopts a simple probabilistic approach to account for the CMCs of NMIs. The intuition is straightforward: the more measurement services an NMI offers, the more likely it is for a company to find a service within the NMI's portfolio that meets it needs. This assumption has a major weakness though. Undoubtedly, some metrology areas are more in demand than others. Moreover, even within a specific area of metrology, some services are more in demand than others. Hence a probabilistic measure that weights all services equally is not completely accurate. On the other hand, this approach implicitly assumes that companies do not demand services from different areas at the same time. Although this may be true for small businesses, large companies may demand numerous services of different types from the same NMI. They do so both for getting better prices, and for technical reasons (the services may be entwined or depend of one another). Again, this means that the probabilistic approach to model CMCs is not entirely accurate. Nonetheless, keeping those conceptual concerns aside, there is also another methodological issue to overcome with this approach. Since the probability of matching – i.e. the probability that an NMI can meet the user's needs – is defined as the number services offered by an NMI over the total number of services offered worldwide, it is crucial to know the total number of distinct services offered by all NMIs together. This cannot be inferred from the table provided by BIPM, since the services are not uniquely identified; only the total number of CMCs by area of metrology is reported. In order to approximate this, a simple method has been used. For each of the nine metrology areas, the country that offers the greatest number of services has been identified, and this figure has been assumed to be the total number of services available in that area. The idea behind this reasoning is that if a country is the world's largest specialist in an area of metrology, its portfolio must be quite comprehensive. In other words, we are assuming that there are not too many services in that area that the NMI does not offer. Therefore, both conceptual and methodological difficulties show that this probabilistic measure has some disadvantages in order for it to be used as proxy of the CMCs of an NMI. In particular, NPL's internal sales data suggests that some users demand services from different scientific groups (or equivalently, they demand services from different areas of metrology). To what extent this is in fact relevant to the econometric results must be determined. That is why an alternative method that accounts for the versatility of an NMI (i.e. how varied its portfolio is) is tested and compared.

For that matter, the geometric mean of the number of CMCs in the nine areas of metrology reported in the table is also proposed as a convenient proxy for the CMCs of an NMI. This alternative measure constitutes a comparable measure that represents the metrology expertise and scope of the NMI's portfolio. It consists of a continuous dimensionless index that assesses the development of the NMI in all areas equally and assigns larger values to the institutes that offer greater variety of services, hence accounting for taste variety among users, who obtain a greater benefit from those NMIs that provide them with more complete service bundles.

Both measures of CMCs have been tested, yielding very similar (almost equivalent) estimation results. The first approach is the preferred one though; basically, because it allows for better interpretation of the estimated coefficients. In any case, the estimation results with respect the second approach based in the geometric mean is included in Annex A.

3.1.3 Geographical proximity

For each country in the dataset, geographical proximity is measured as the distance in kilometres between London and the most populated city in the purchasing country. However, around half of NPL's income comes from the UK. Hence, we need a measure of the average distance between NPL and users within the UK; here, we follow the commonly used guidelines by that define the internal distance as:

$$d = 0.67\sqrt{A/\pi} \tag{3.1}$$

where A is the area of the country.

3.2 Data description

The objective of this analysis is to measure the price elasticity of demand. For that matter, it is convenient to work with the logarithms of the variables above. Table 2 shows some basic descriptive statistics of the variables of interest. (Variables in levels are abbreviated in capital letters; variables in logarithmic form are lowercased).

Variable	Mean	Std. Dev	Relative Std. Dev	within variation	Skewness	Kurtosis	Min	Max
I	2.4E-06	8.0E-06	3.3	0.5	6.2	46.3	0.0	8.6E-05
PL	0.5	0.3	0.5	0.3	0.8	2.8	0.1	1.4
GDP	1.5E+04	1.4E+04	0.9	0.4	1.7	7.0	4.1E+02	9.7E+04
CMC	11.6	16.7	1.4	0.0	2.3	8.0	0.0	77.6
D	5167.5	3933.0	0.8	0.0	0.8	3.5	185.8	19147.1
POP	5.9E+07	1.8E+08	3.0	0.0	6.0	40.8	3.3E+04	1.3E+09
i	-14.1	2.1	0.2	0.3	-0.4	2.7	-20.3	-9.4
pl	-0.7	0.5	0.7	0.4	0.0	2.3	-2.1	0.4
gdp	9.2	1.1	0.1	0.3	-0.6	2.9	6.0	11.5
d	8.2	1.0	0.1	0.0	-0.7	2.9	5.2	9.9
pop	16.3	1.9	0.1	0.0	-0.4	3.9	10.4	21.0

Table 2: Summarised descriptive statistics.

Table 2 provides some insight to understand the dataset[14]. To begin with, the outcome variable $I_{i,t}$ (the number of invoices issued to a country normalised by the population of the country) is not defined for over 40% of the panel. This is because small and/or less developed countries lack enough companies that need high accuracy measurement services. This leads to NPL sales to those countries being rare events. Therefore, the number of invoices is zero for a

[14] Annex C contains a more detailed table with summary statistics for all the variables included in **Table 2**.

significant part of the panel, causing a substantial incidental truncation of the outcome variable once logarithms are taken. This has important consequences on the statistical techniques needed to analyse the data properly; section 4.1 analyses this matter in depth, and proposes a two-step Heckman model as the preferred way to account for any selection bias that would come from this structure of the panel. Moreover, Annex A analyses the effect of removing countries with too few observations from the sample. The logic behind checking the robustness of the results when trimming these observations, is that those countries might have fundamentally different needs than the rest of the sample. These unobserved characteristics could have a misleading effect when estimating the real price elasticity. As shown in this annex the effect is not considerable.

Another key feature of the dataset is the distinction between *within* and *between* variation. These measure how far are the variables of interest spread out from their average value, across countries and over time respectively. Obviously, since distance (D) is a time-unvarying variable, it will not show any within variation. In addition, although the measurement capabilities of the home NMI (CMC) and the population of the country (POP) do vary with time, they barely do so. For that reason, for both variables only the 2019 figures have been considered. Therefore, these two will not show any within variation either.

As expected, the time-variant variables show more variation from country to country than over time for the same unit. Nonetheless, the national price level presents significant variation over time. This is an important requirement for the analysis proposed to work. Note that a basic element of the analysis, and so one of the reasons why overseas sales are used to estimate the price elasticity, is that measurement services relative prices have stayed fairly constant over the period analysed. Hence, we require enough variation in relative national price levels to estimate the price elasticity of NPL's services.

Finally, it is important to assess if there are outliers in the panel and whether they affect the estimates or not:

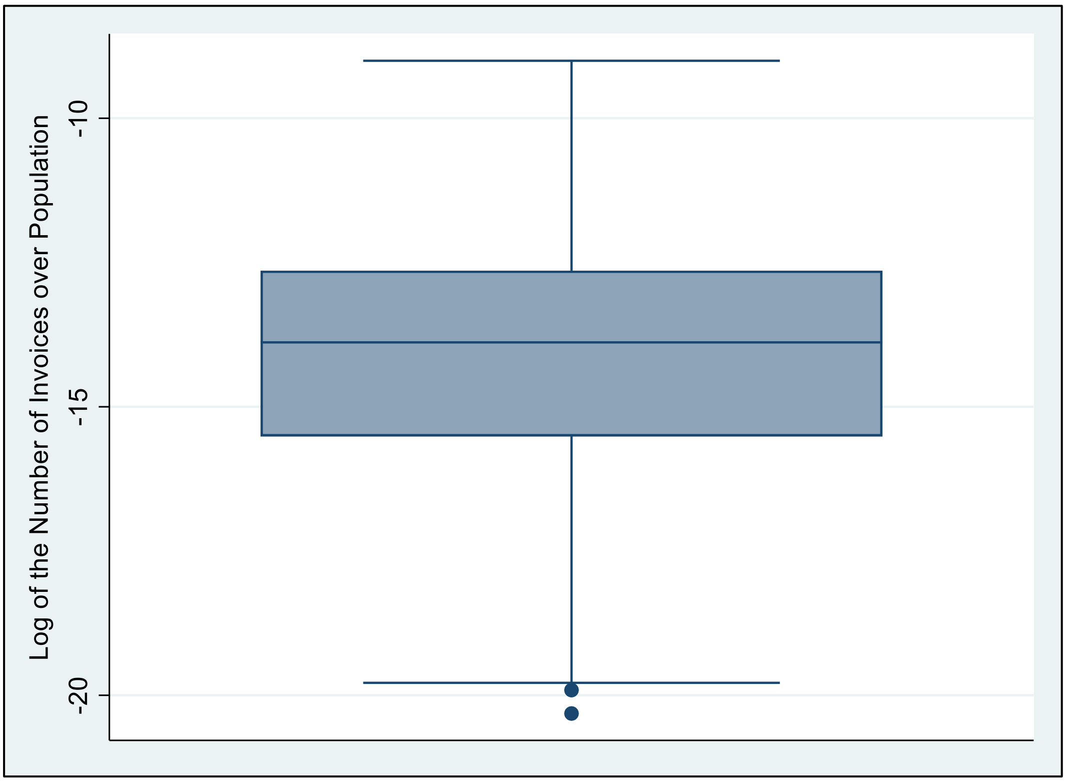


Figure 1: Box plot of the dependent variable.

The dependent variable shows a neat log-normal distribution. In any case, the identified outliers were dropped applying the standard approach by . However, the effect of these outliers in the estimation results is negligible. This is shown in Annex A.

3.3 Econometric analysis

The following section constitutes the nucleus of the analysis carried out in this paper. Firstly, the model of the data-generating process is presented, and the potential sources of bias are discussed. Secondly, the estimation results for the models proposed are presented; the models differ on the dynamics considered and whether the Heckman correction is applied or not – the goal is to build from simple to complex to allow the reader to understand the line of reasoning followed. Lastly, the postestimation test are run to find the preferred setup.

3.4 Specification

The theoretical population model to be analysed is:

$$\log I_{i,t} = \beta_0 + \beta_1 \log PL_{i,t} + \beta_2 \log GDP_{i,t} + \beta_3 CMC_i + \beta_4 \log D_i + \beta_5 \log POP_i + \gamma_t Y_t + \varepsilon_{i,t} \tag{4.1}$$

where $I_{i,t}$ is the number of invoices issued to country i in year t normalised by the population of the country, $PL_{i,t}$ is the national price level, $GDP_{i,t}$ is the gross domestic product per capita, CMC_i is a continuous index that rates the measurement capabilities of the country, D_i is the distance between the country and the UK, POP_i is the population of the country, Y_t is a year dummy variable, and $\varepsilon_{i,t}$ is the error term which is assumed to be independently and identically distributed as a normal distribution with mean of 0 and finite variance.

It could be argued that users in different countries might have fundamentally different needs. However, the proposed specification given by equation 4.1 does not contain an intercept which is dependent on the unit of analysis – a fixed effect that reflects the unobserved heterogeneity across countries. The reason for that is that we lack the statistical power as to assign each country its own intercept. This poses the question of whether we can use the elasticity found from this international dataset to characterise demand by UK users. We argue that the majority of sales occur in comparable countries such as OECD countries (94.8%) or EEA countries [15] (89.4%), and thus, there is reason to believe that ignoring the fixed effect is not crucial for the purpose of the analysis.

On the other hand, the inclusion of the year dummy variable must be discussed. This variable allows us to control for the effect of year-specific events.

Another important element worth highlighting is the fact that the variable that controls for the measurement capabilities of the home NMI is introduced in the model in level rather than logged. This makes the interpretation of the coefficient much more straightforward. Since this variable is a probabilistic measure of the local NMI offering a service that meets the user needs (as detailed in section 3.1), the coefficient estimate should be interpreted as the effect on sales (expected to be negative) of an increase of 1% on such probability. In any case, the correctness of the specification in level rather than logged has been tested using the approach by . All the detail can be found in Annex A.

Lastly, note that the ultimate goal of the model given by equation 4.1 is to estimate the price elasticity β_1. To ensure that the estimate we get from this specification is close to the real parameter, any potential source of bias needs to be accounted for. The rest of this subsection analyses systematically all potential issues that could be biasing the estimation results.

[15] European Economic Area.

3.4.1 Omitted variables bias

The specification defined by equation 4.1 clearly does not include all the variables that are responsible for all the variation in the dependent variable. By definition, the effect of these omitted variables is included in the error term. Consequently, the estimations provided by our model will not be biased, as long as the error term is not correlated with the regressors.

One relevant omitted variable identified in section 2 which is not explicitly included as a covariate, is delivery time. Indeed, international sales of high accuracy measurement services are influenced (presumably to a large extent) by the time that NMIs take to deliver their services. For instance, we can think of how VSL – the Netherlands' NMI – and NPL compete for an Australian customer. From the user's perspective, both are distant NMIs offering similar services and prices; hence, it is plausible that the decision between the two is mainly determined by how much time it would take for the customer to get the service.

So, the question is whether delivery times – or any other omitted variable for that matter – is correlated to our regressors. In particular, we are interested in knowing whether the national price level is endogenous, since the ultimate goal of the analysis is to determine the price elasticity of demand for NPL's services.

Endogeneity could be originated in either of the two dimensions of our panel. On the one hand, a certain degree of unobserved time-independent heterogeneity among the countries in our sample is expected. Unsurprisingly, customers from the US are likely to show different needs than Nigerian businesses. However, whether this fixed effect is sufficiently correlated with our regressors as to significantly bias our estimations is another matter. As it will be shown in section 4.3 this fixed effect is negligible; thus, the proposed approach of pooling all the observations ignoring any country-specific effect is deemed to be quite accurate.

Alternatively, endogeneity could be the result of the dynamics of the generating process. In other words, NPL's sales to other countries may depend not only on the contemporary values of the variables that determine these sales, but on their recent history too. This persistence can be caused by the explanatory variables, the unobservables, or both (i.e. we could introduce in our specification lagged regressors, past realisations of the residuals or lagged values of the dependent variable which effectively encompasses both). In our case, it is fair to assume that the inertia is just in the error term; that is, the dynamic persistence comes from the unobservables. There are two main arguments to motivate this assumption. Firstly, sales are supposed to react quickly to the time-varying regressors in our specification. That is, in most markets, a consumer is assumed to react very quickly to variations in relative prices and changes in their budget constraint. Secondly, NPL's sales worldwide are subject to macroeconomic forces that determine international trade; that is, the global business cycle. These in turn tend to show some persistence. Hence, the dynamics of the system must be caused by macroeconomic variables in the error term. For that reason, most of the models tested section 4.2 include lagged values of the residuals of the static regression model to account for this dynamic behaviour of the population process[16].

Endogeneity has been systematically checked in section 4.3 for all the models proposed. To that end, the approach suggested by is implemented[17]. The results of this test show that the assumption of exogeneity is correct (even if country-fixed-effects are ignored) as long as the dynamics of the population process are taken into account. This is very much expected in the sense that the idea behind using the Mundlak test is to test whether the time-invariant unobservable (unobserved heterogeneity) is correlated with our regressors of interest; in

[16] (Beck & Katz, 2011) deals with a variety of dynamic issues in the analysis of time-series–cross-section data that can be addressed by different kinds of dynamic specifications. In particular, the authors discuss the role of the serially correlated error model in panel data analysis governed by dynamic processes.

[17] A thorough explanation of this method is provided in Annex B.

particular with the national price level regressor which gives us in the end the price elasticity estimate. The fact that we pass the Mundlak test shows that the covariates are uncorrelated with the error term, which includes the unobserved heterogeneity. As mentioned, this is expected because although the fixed effect embodies fundamental determinants of trade such as historical links or language, these are almost certainly not correlated to the national price level, which is determined by macroeconomic forces at a global scale.

3.4.2 Errors-in-variables

Some of the independent variables might have been measured with errors. This would lead to inconsistent OLS estimations that underestimate the coefficient (attenuation bias). When dealing with country-level economic data the reliability of statistical information from less developed countries is always a concern. However, a significant measurement error seems unlikely due to the fact that the data utilised comes from a trusted external source such as the World Bank. In any case, even if some of those less developed countries include measurement error, they are not among the countries NPL sells the most, thus the relevance in our regression analysis is assumed minimal.

3.4.3 Simultaneity

Simultaneity bias can be ruled out from the get-go mainly because of the nature of the analysis. Our goal is to explain NPL sales worldwide through macroeconomic variables that are driven by forces at a global scale. Even if a hidden macroeconomic variable is affecting both the dependent variable and one of the regressors, as mentioned before, our specification controls for it by including a dummy variable for the year which should capture the effect of significant events.

3.4.4 Weighting of observations in the sample

An analytical weighting correction has been implemented across all estimations conducted, because the dependent variable consists of the number of invoices issued to a country divided by its population. Although the change in the results is not noticeable (the same values and confidence intervals are obtained up to several significant figures), the correction has been implemented as it is the standard in the empirical economic literature that uses country-level data.

3.4.5 Selectivity

Finally, selection bias deserves a more detailed explanation, since it is very relevant to our analysis. Many countries in the dataset, usually small and/or less developed countries, do not make any purchase in some years, and, thus, NPL does not issue any invoice to those countries. As a result, the dependent variable is zero for over 40% of the dataset. Moreover, since equation 4.1 is specified in terms of the logarithm of the dependent variable, a significant incidental truncation arises. In order to address this issue, a two-step selection correction has been used to model the individual sampling probability of each observation occurring (extensive effect), together with the conditional expectation of the dependent variable (intensive effect). In other words, the approach adopted allows us to disentangle the incidence of the national price level on the probability of purchase and the level of consumption. The first stage is represented by the dichotomous choice of whether to purchase NPL measurement services or not, and the second stage determines the level of consumption once the decision to purchase is made.

3.5 Estimation

Four different models have been tested based on the population equation 4.1. Models A1 to A3 consists of uncorrected for selectivity OLS regressions. Model A1 is static. Thus, it consists of a first naïve approximation to the population process that neglects any persistence.

However, as mentioned in section 4.1.1, the generating process is likely to exhibit some inertia, due to all the macroeconomic unobservables in the error term. For that reason, models A2 and A3 introduce one lag and three lags of the residuals of A1 respectively. These hold significant explanatory power over the dependent variable – the latter removing any trace of serial correlation. Thus, the presumed dynamic structure is confirmed. Lastly, Model B1 consists of the 2-step Heckman correction of A3. This is the preferred model because it accounts for the dynamics of the population process, as well as the potential bias coming from selectivity in the sample.

This subsection first introduces the estimations results for the four models, analysing the economic meaning of each estimated coefficient. Then the preferred model is presented and analysed in depth to show how the correction proposed by applies to our model and why it is a reliable method to this context.

3.5.1 Models tested and interpretation of the coefficients

Table 3 summarises the estimations results for models A1 to A3 and model B1:

DV = log *I*	**Model A1**		**Model A2**		**Model A3**		**Model B1**	
	Coeff.	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**
log *PL*	1.31 (0.10)	0.000	1.25 (0.08)	0.000	1.19 (0.07)	0.000	1.24 (0.08)	0.000
log *GDP*	1.15 (0.08)	0.000	1.23 (0.07)	0.000	1.25 (0.07)	0.000	1.33 (0.08)	0.000
CMC	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000
log *D*	-0.44 (0.03)	0.000	-0.45 (0.02)	0.000	-0.47 (0.02)	0.000	-0.48 (0.02)	0.000
log *POP*	-0.20 (0.04)	0.000	-0.20 (0.02)	0.000	-0.20 (0.02)	0.000	-0.17 (0.03)	0.000
u (t-1)	. .	.	0.76 (0.03)	0.000	0.36 (0.05)	0.000	0.35 (0.05)	0.000
u (t-2)	. .	.	. .	.	0.33 (0.04)	0.000	0.33 (0.04)	0.000
u (t-3)	.	.	.	.	0.22	0.000	0.22	0.000

	.		.		(0.04)		(0.04)	
Λ	.	.	.	.	.	.	0.28	0.047
	.		.		.		(0.14)	
	F-stat	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**
time dummies	7.13	0.000	11.23	0.000	18.59	0.000	17.87	0.000
R-squared	0.81		0.92		0.94		0.94	
Number of obs.	939		801		648		648	
Heckman	No		No		No		Yes	

Table 3: Estimation results.

The estimated coefficients for our variables of interest are highly significant and make perfect economic sense. Notably, the price level coefficient is positive. This is consistent with the definition of the national price level given by equation 2.1. With respect to GDP per capita, the coefficient is also coherent with the expected income effect, since NPL's services are supposed to be normal goods whose demand increases with real income. By contrast, the negative significant coefficients found for distance and CMCs of the local NMI tally with our assumption of transactional frictions in the form of transport and search costs. Furthermore, the statistical significance of the population coefficient suggests a non-linear relationship of the number of invoices and the population of the country. In addition, the significance of the coefficients of the lagged residuals confirms the presumption of persistence in the generation process.

Lastly, there is an additional regressor symbolised by λ which only applies to model B1. This is called the Inverse Mills Ratio (IMR), and results from a probit model that tries to reproduce the decision made by users worldwide about whether to buy NPL's services or not. Surely, there are unobserved factors that make the decision of buying NPL's services more likely, as well as being associated with higher levels of consumption. In other words, it is expected that those users who are more inclined to buy NPL's services, are also the ones that buy more services. By adding the IMR to the dynamic structure set by A3, we can control for those unobserved factors that affect both the decision of buying NPL's services and the level of consumption; thus, any bias in the rest of estimated parameters can be corrected.

The next subsection analyses in more depth the preferred setup Model B1. It depicts extensively the application of the Heckman correction to our model and justifies why it is a suitable method to address the potential selection bias in the analysis. Finally, the estimation results are interpreted focusing on the existing bias on the uncorrected setup, A1 to A3.

3.5.2 Model B1: 2-step Heckman selection model

The estimates found through the models A1 to A3 might be biased due to selectivity. The coefficients estimates may not be applicable to all countries (buyers and non-buyers) because we only observe those that actually make a purchase. In other words, the estimation results may not be representative of the whole population because of the non-random nature of the observed sample. To correct for this source of bias, we follow the approach proposed by . Formally, the correction to account for incidental truncation consists of adding an explicit selection equation to our population equation 4.1. Thus, the model is given by:

$$\log I_{i,t} = \beta_0 + \beta_1 \log PL_{i,t} + \beta_2 \log GDP_{i,t} + \beta_3 CMC_i + \beta_4 \log D_i + \beta_5 \log POP_i + \gamma_t Y_t + \delta_1 u_1 + \delta_2 u_2 + \delta_3 u_3 + \varepsilon_{i,t} \tag{4.2a}$$

$$s_{i,t} = 1\left[\eta_0 + \eta_1 \log PL_{i,t} + \eta_2 \log GDP_{i,t} + \eta_3 CMC_i + \eta_4 \log D_i + \eta_5 \log POP_i + \theta_i Y_t + \kappa_1 v_1 + \kappa_2 v_2 + \kappa_3 v_3 + \varphi_{i,t}\right] \tag{4.2b}$$

where $s_{i,t} = 1$ if a purchase is made, and zero otherwise, $I_{i,t}$ is the number of invoices issued to country i in year t normalised by the population of the country, $PL_{i,t}$ is the national price level, $GDP_{i,t}$ is the gross domestic product per capita, CMC_i is a continuous index that rates the measurement capabilities of the country, D_i is the distance between the country and the UK, POP_i is the population of the country, Y_t is a year dummy variable, u_1 to u_3 are the first three lagged residuals of the static uncorrected model (Model A1), v_1 to v_3 are the first three lagged residuals of the static specification of the probit model, and $\varepsilon_{i,t}$ and $\omega_{i,t}$ are the error terms in the population and selection equation respectively, which are assumed to be independently and identically distributed as a normal distribution with mean of 0 and finite variance.

An important remark should be made with regard to the differences between the sets of regressors used in both equations 4.2. Ideally, for the proposed method to work best, the set of regressors in the population equation should be a strict subset of the covariates in the selection equation. This means that two conditions must be satisfied: (1) any of the regressors in equation 4.2a should also be included in equation 4.2b, and (2) some of the regressors in equation 4.2b should not appear in equation 4.2a.

Technically, the first constraint is not met, whereas the second one is. Indeed, the first three lagged residuals of the static uncorrected model, u_1 to u_3, are not included in equation 4.2b. Conversely, these have been substituted by the first three lagged residuals of the static specification of the probit model. The reason for this is that, although both sets of residuals are linked to the unobservables, arguably the latter is more suitable to consider the effect of the omitted variables on the outcome variable since it does so in probabilistic terms. In any case, ideally, at least a true additional identifying variable should be included in the selection equation. This variable would affect the probability of making a purchase, but not (at least to a large extent) the number of invoices issued. However, all attempts to include identifying variables linked to the complexity of the country's exports or the relevance of the manufacturing sector have been unsuccessful. The reason for this is that these variables are highly correlated with the CMC regressor which assesses the measurement capabilities of the local NMI – and which holds more explanatory power. Nevertheless, the proposed setup happens to work quite well in terms of statistical significance and makes the most sense from a theoretical perspective.

The Heckman two-stage estimation procedure allows us to decompose the total effect of the covariates of interest (in particular the price level) into an extensive (probability of purchase – equation 4.2b) and an intensive (level of consumption – equation 4.2a) component.

Table 4 summarises the estimation results of the probit model that assesses the probability of purchasing NPL services (dummy variable $TREATED = 1$) as a function of the original regressors[18].

DV = ***TREATED***	**Probit Model**	
	Coeff.	**P-value**
log *PL*	0.91 (0.19)	0.000
log *GDP*	0.91 (0.09)	0.000
CMC	0.02 (0.01)	0.211
log *D*	-0.26 (0.07)	0.000
log *POP*	0.46 (0.06)	0.000
v (t-1)	-0.40 (0.34)	0.229
v (t-2)	-1.22 (0.33)	0.000
v (t-3)	-0.17 (0.32)	0.586
Pseudo R-squared	0.44	
Number of obs.	1134	

Table 4: Probit regression estimation results (extensive effect).

[18] The estimated coefficients for the year dummy variables have been omitted from this output for the sake of conciseness. These are included in Annex D.

These coefficients have no direct interpretation in terms of marginal effects; they are simply the values that maximise the likelihood function. Although the signs of the coefficient give the direction of the effect, their magnitude is in units of the standard-deviation of the errors. There are two commonly used approaches to approximate the marginal effects for non-linear models: the partial effects at the average (PEA) and the average partial effect (APE). The PEA is the marginal effect evaluated at the average value for each regressor; the APE is simply an estimate of a population-averaged marginal effect:

$$PEA_j = \left.\frac{\partial \mathbf{E}[y|\boldsymbol{x}]}{\partial \boldsymbol{x}}\right|_{x=\bar{x}} = \left.\frac{\partial \Phi[\boldsymbol{x\beta}]}{\partial \boldsymbol{x}}\right|_{x=\bar{x}} = \hat{\beta}_j \cdot \varphi\left(\bar{\boldsymbol{x}}\widehat{\boldsymbol{\beta}}\right) \tag{4.3a}$$

$$APE_j = \frac{\hat{\beta}_j}{N}\sum_{i=1}^{N} \varphi\left(\boldsymbol{x}\widehat{\boldsymbol{\beta}}\right) \tag{4.3b}$$

where PEA_j and APE_j are the partial effect at the average and the average partial effect for regressor j, y and $\boldsymbol{x}$ are denote our binary dependent variable and the set of covariates in the probit regression, N is the total number of observations, $\varphi(\cdot)$ is the standard normal density function, and $\widehat{\boldsymbol{\beta}}$ is the vector of the parameter estimates (including the intercept) of the probit regression – with $\hat{\beta}_j$ representing a component of that vector. Table 5 below shows both approximations of the marginal effects[19].

DV = *TREATED*	**Marginal Effect at Means**		**Average Marginal Effect**	
	Coeff.	**P-value**	**Coeff.**	**P-value**
log *PL*	0.18 (0.04)	0.000	0.17 (0.04)	0.000
log *GDP*	0.18 (0.03)	0.000	0.17 (0.01)	0.000
CMC	3.6E-03 (2.6E-03)	0.173	3.4E-03 (2.7E-03)	0.204
log *D*	-0.05 (0.01)	0.000	-0.05 (0.01)	0.000
log *POP*	0.09 (0.01)	0.000	0.09 (0.01)	0.000

[19] Again, the estimated coefficients for the year dummy variables have been omitted; these can be found in Annex D.

v (t-1)	-0.08 (0.07)	0.240	-0.08 (0.06)	0.230
v (t-2)	-0.24 (0.07)	0.000	-0.23 (0.06)	0.000
v (t-3)	-0.03 (0.06)	0.587	-0.03 (0.06)	0.586
Number of obs.	1134		1134	

Table 5: Partial effects for the extensive effect.

Note that both effects are very similar for all of the regressors. These is expected given the covariates are reasonably normally distributed, and thus, the mean is a good measure of centrality within the domain of definition of the covariates.

Once the extensive effect is estimated, we can turn our attention to the intensive effect. Our objective is to know how the variables of interest affect the level of consumption – in particular, the national price level. If we analyse this through a simple OLS regression, the results will be surely biased because the factors that determine if users buy NPL's services, also affect how much they consume. Hence, the sample we are working with is not representative of the whole population. In effect, we need to separate out the two decisions made by the user: whether to make a purchase or not (extensive), and how much to buy (intensive). To that end, we can use the results of the previous probit estimation by computing the Inverse Mills Ratio. The IMR is equivalent to the concept of hazard ratio in survival analysis, or force of mortality in demography and actuarial science. In the context of our analysis, the IMR shows the relative increase in the likelihood of buying NPL's services given a unit change in the level of consumption:

$$\lambda_{i,t} = \frac{\varphi(\boldsymbol{x}\widehat{\boldsymbol{\beta}})}{\Phi(\boldsymbol{x}\widehat{\boldsymbol{\beta}})} \qquad (4.3)$$

If the estimate of the linear parameter corresponding to the IMR is significant (expected positive), then we know the unobserved factors that make the decision of buying NPL's services more likely, tend to be associated with higher levels of consumption. Therefore, effectively, including the IMR in the OLS regression allows us to control for those unobserved factors and correct any bias in the rest of estimated parameters. Table 6 shows the results of the OLS regression including the IMR obtained from the previous probit setup.

DV = log *I*	**Model B1**	
	Coeff.	**P-value**
log *PL*	1.24 (0.08)	0.000

log *GDP*	1.33 (0.08)	0.000
CMC	-0.02 (0.00)	0.000
log *D*	-0.48 (0.02)	0.000
log *POP*	-0.17 (0.03)	0.000
u (t-1)	0.35 (0.05)	0.000
u (t-2)	0.33 (0.04)	0.000
u (t-3)	0.22 (0.04)	0.000
λ	0.28 (0.14)	0.047
	F-stat	**P-value**
time dummies	17.87	0.000
R-squared	0.94	
Number of obs.	648	

Table 6: Heckman's selection model estimation results (intensive effect).

As expected, the IMR is significant and positively signed – which suggests that the error terms in the selection and primary equations are positively correlated. Therefore, we confirm our suspicions that factors that make buying NPL's services more likely are associated with purchasing more of these goods. This justifies the use of Heckman's correction in our setup and provides confidence on the estimation results.

Finally, the estimated coefficient on the national price level is now larger than before – that is if we compare estimates for models A3 and B1. This indicates that the non-random nature of the sample biases the price elasticity down.

3.6 Postestimation

This subsection analyses the validity of the models tested, particularly that of the preferred 2-step Heckman selection model. Table 7 shows the results of a link test[20] which has been run to check for misspecification of the dependent variable for the four models tested. The motivation behind the link test is to assess if a regression model is affected by the so-called link error, that is, that the dependent variable needs a transformation or *link* function to properly relate to the independent variables. To verify this is not the case, the link test regresses the dependent variable against the original regression's predicted values and the squared values of this prediction. If the squared prediction regressor is significant, there is evidence that the model is mis specified; in addition, it is expected that the coefficient for the prediction regressor is highly significant with a coefficient close to one.

DV = log *I*	**Model A1**		**Model A2**		**Model A3**		**Model B1**	
	Coeff.	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**
predicted values	1.07	0.000	1.31	0.000	1.04	0.000	1.03	0.000
	(0.20)		(0.12)		(0.11)		(0.11)	
squared prediction	0.00	0.706	0.01	0.011	0.00	0.731	0.00	0.768
	(0.01)		(0.00)		(0.00)		(0.00)	
R-squared	0.82		0.92		0.94		0.94	
Number of obs.	939		801		648		648	

Table 7: Link test.

All the models pass the test. Even in the case of model A2 for which the coefficient of the squared prediction is statistically significant, the coefficient itself is very close to zero. However, clearly those models that include three lags of the error term of the static model (models A3 and B1) behave better. Therefore, the link test provides assurance in the specification of the preferred model.

In order to provide further evidence of the correct specification of the model, the Ramsey Regression Equation Specification Error Test (RESET) is conducted. This technique tests whether non-linear combinations of the fitted values have explanatory power over the dependent variable. The intuition behind this test is that if the second or the third power of any of the regressors, or any interaction term between them have any power in explaining the

[20] The test implemented here is based on an idea of (Tukey, 1949) which was further described by (Pregibon, 1980).

dependent variable, then the model omits relevant variables and may be better approximated by different functional form. Table 8 reports the results of the Ramsey RESET test for the four models considered.

Model A1		Model A2		Model A3		Model B1	
F-stat	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**
12.02	0.00	6.67	0.00	1.64	0.18	1.52	0.21

Table 8: Ramsey RESET test.

Like the link test, the Ramsey RESET test favours both specifications that account for three lags of the error term of the static regressions. Moreover, these are the only specifications that pass the test; hence, the test supports the preferred Heckman-corrected dynamic specification – model B1.

Two other important aspects to ensure the validity of the specification proposed are whether there are signs of endogeneity in the models tested, and whether the residuals are serially correlated. Both issues can be analysed together since they are linked. Indeed, endogeneity is likely to arise due to the dynamics of the generating process, hence leading to serially correlated errors if not taken into account. Table 9 shows the results of a test based on which allows us to test for endogeneity is an alternative to the Hausman test[21]. Unlike the Hausman test, the method developed by Mundlak may be used when the errors are heteroskedastic or have intragroup correlation. Hence, this alternative approach is especially suitable for the purpose of this analysis.

[21] A thorough explanation of the intuition and the algebra behind this method is depicted in Annex B.

DV = log *I*	Model A1		Model A2		Model A3		Model B1	
	Coeff.	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**
log *PL*	-0.04 0.19	0.816	1.00 0.17	0.000	0.74 0.27	0.006	0.77 0.26	0.004
mean (log *PL*)	1.83 0.34	0.000	0.28 0.22	0.218	0.50 0.30	0.094	0.52 0.28	0.066
log *GDP*	0.85 0.22	0.000	1.24 0.18	0.000	1.37 0.22	0.000	1.47 0.23	0.000
mean (log *GDP*)	-0.10 0.25	0.682	-0.02 0.19	0.911	-0.14 0.21	0.509	-0.16 0.21	0.457
CMC	-0.01 0.01	0.105	-0.02 0.00	0.000	-0.02 0.00	0.000	-0.02 0.00	0.000
log *D*	-0.40 0.12	0.001	-0.45 0.03	0.000	-0.46 0.02	0.000	-0.47 0.02	0.000
log *POP*	-0.32 0.10	0.002	-0.20 0.04	0.000	-0.20 0.02	0.000	-0.17 0.02	0.000
	F-stat	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**
Time dummies	240.33	0.000	93.91	0.000	44.45	0.000	48.04	0.000
Panel-level means	29.43	0.000	1.57	0.457	3.97	0.137	4.32	0.115
R-squared	within = 0.323 between = 0.861 overall = 0.808		within = 0.14 between = 0.982 overall = 0.917		within = 0.137 between = 0.988 overall = 0.938		within = 0.136 between = 0.991 overall = 0.939	

Number of obs.	939	801	648	648
Heckman	No	No	No	Yes

Table 9: Mundlak test.

After accounting for the dynamics of the generating process the cluster means of the two time-varying covariates become insignificant. Therefore, this means that our choice to explicitly model this inertia is appropriate, and particularly our preferred model B2 shows no signs of endogeneity. These conclusions are reinforced when we test whether the residuals are serially correlated:

DV = residuals	Model A1		Model A2		Model A3		Model B1	
	Coeff.	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**
residuals (t-1)	0.76	0.000	-0.29	0.000	-0.03	0.453	-0.03	0.444
	(0.02)		(0.04)		(0.04)		(0.04)	

Table 10: Serial correlation test.

It can be concluded that those models that incorporate three lags of the residuals of the static model better reflect the dynamics of the generating process.

Finally, it is necessary to make a brief mention about the normality of the residuals of the models tested. Although the sample is not very wide (just below a thousand observations) in all cases the residuals have a fairly normal distribution. This can be appreciated in Figure 2 where the histogram of the residuals for all models are plotted, and a kernel density estimation of the probability density function is compared against the expected normal distribution.

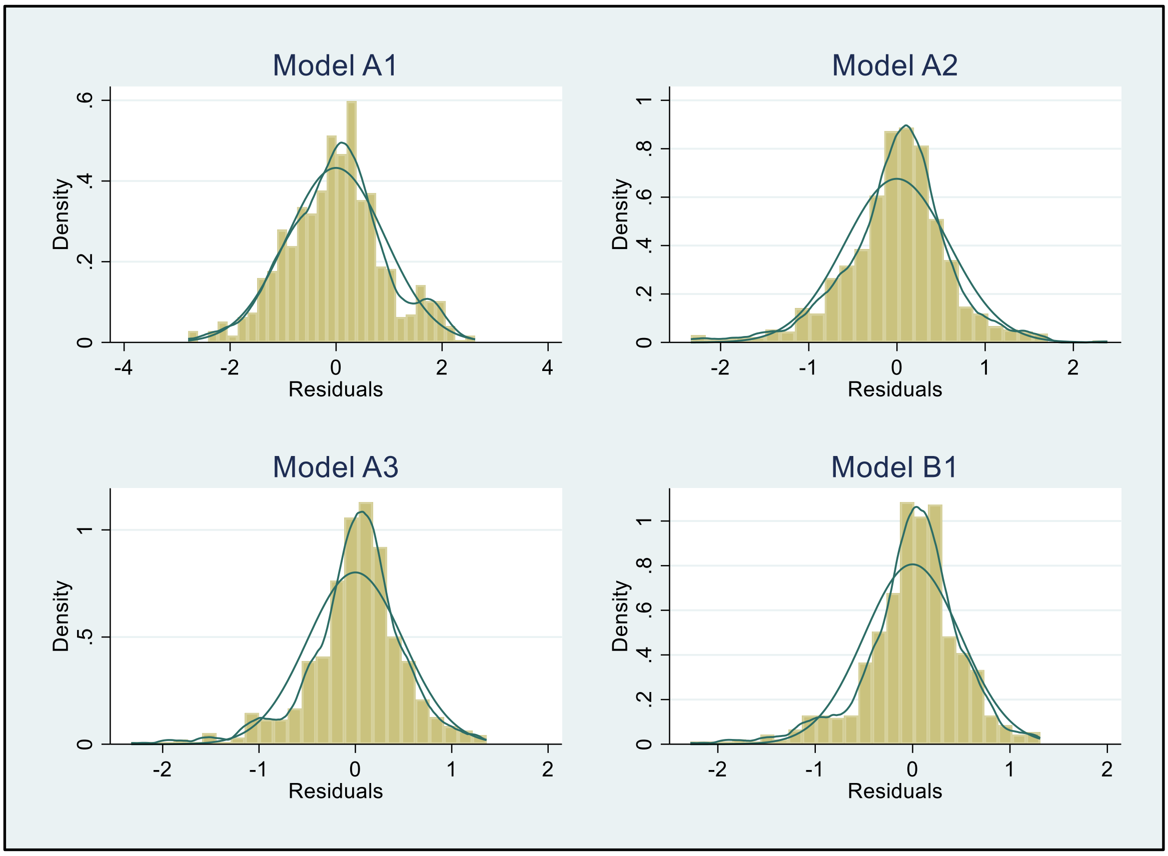


Figure 2: Normality of the residuals.

4 Economic Impact

This section shows how the price elasticity found in the previous econometric analysis can be used to estimate the welfare generated by NPL for the UK's society. It should be emphasized that the objective here is not to develop a comprehensive and accurate impact analysis, but to showcase that the price elasticity of demand is a useful tool to asses welfare generated. Therefore, the reader should not expect the following to be a detailed and complete analysis of the net present value of the public funding of NPL. In fact, all the figures pertaining to NPL's finances and public support received are estimative (although not very far off). Thus, any cost-benefit ratio presented should not be taken as representative of the current value for money of NPL public financing.

When it comes to assessing welfare created, there are two main types of benefits generated: direct and indirect benefits. Direct benefits are those that occur *within* supported companies. Effectively, NPL sells the time and the expertise of its scientists and engineers who help users to develop new products and processes. New products enable supported businesses to increase their market power and command a price premium. New processes enhance productivity and competitiveness. In any case, NPL's services lead to additional earnings, either through increased sales or costs savings. However, direct benefits are not limited to the increase in private profits which roughly accounts for about 25% of gross value added (GVA). The remaining three quarters are earned by workers in the form of higher salaries. Therefore, GVA is probably about four times the direct return to NPL's paying customers.

There are wider benefits that arise *beyond* the supported businesses. These are indirect benefits that occur because knowledge generated due to the collaboration between NPL and paying customers spills over to unsupported firms – normally in the same sector. The main channel through which this knowledge benefits companies that do not engage with NPL is the movement of workers among firms. The intuition behind this knowledge spillovers is as follows. First, a company engages with NPL. This allows the supported business to acquire knowledge which is then drawn on to promote additional sales and/or costs savings. Obviously, most of this knowledge is effectively obtained by the company's workers. If these workers decide to move to other firms (typically in the same sector or industry), they will carry over the acquired new knowledge. Frontier Economics found that the existing literature estimates that "social returns, based on spillover benefits from R&D conducted by one agent to the productivity or output of other agents, are typically 2 to 3 times larger than private returns." Therefore, in this analysis we will stick to the lower bound of these estimates and will approximate the indirect benefits by just doubling the direct benefits.

Lastly, given that 50% of NPL's income comes from customers who are based in other countries, the question of whether those sales have a positive impact in the UK arises. Regarding the direct benefits, it is clear that none of these stay in the UK – most of the GVA generated happens in the foreign country. However, that is not the case for indirect benefits. Indeed, NPL plays a fundamental role when it comes to any knowledge generated from collaborations with non-UK based users spilling over to UK businesses. This is due to the fact that NPL's scientific staff gain knowledge as a result of the collaboration. Moreover, these scientists and engineers tend to be young professionals who are likely to switch jobs given they are in the early stages of their career. In fact, the stability in NPL's workforce has decreased in recent times, as has the average age of workers. In this sense, NPL acts as a platform for all the knowledge generated (including the one developed as a result of collaborations with overseas companies) to reach all layers of the sectors involved. Hence, in the following impact analysis we will account for indirect benefits resulting from sales to users abroad, despite direct benefits being ignored.

4.1 Direct benefit to paying customers and welfare for the UK

Companies regard NPL's services as investment projects that generate profits over time. A rational user would prioritise projects with a higher payoff and, if no budget constraint is in place, would buy NPL's services to the point that the marginal benefit equals the marginal cost. This suggests a downward sloping aggregate demand curve for NPL's services: more profitable projects yield higher future earnings which increases customers' willingness to pay. This situation is depicted in Figure 3. The horizontal axis shows the number invoices issued by NPL, which is a proxy for the volume of services provided. The vertical axis of Figure 3 is the net present value of the investment projects supported by NPL.

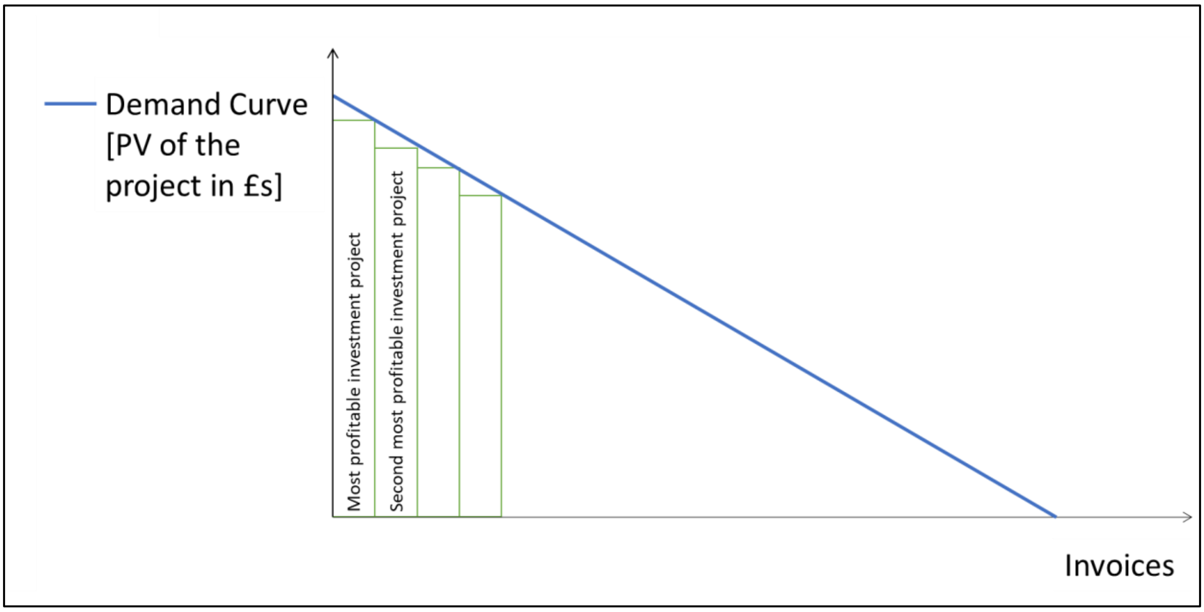


Figure 3: Demand curve for NPL's services.

Therefore, at first approximation the demand curve for NPL's services is given by:

$$p(Q) = A - BQ \tag{5.1}$$

Where $A > 0$ and $B > 0$ are constants.

The direct benefit to NPL's customers is given by the area beneath the curve:

$$\Pi = pQ + \frac{1}{2}(A - p)Q = AQ - \frac{1}{2}BQ^2 \tag{5.2}$$

Hence, in order to estimate the direct benefit to paying customers, we need to know the values of A and B. To that end, we can make use of the price elasticity of demand:

$$\varepsilon_p^0 = \frac{dQ}{dp}\frac{p^0}{Q^0} = -\frac{1}{B}\frac{p^0}{Q^0} \tag{5.3}$$

Note that in general the value of the elasticity depends on the level of sales. (In particular, this is true for the linear approximation we are considering). Thus, the superscripts in ε_p, p and Q denote the fact these are evaluated in the current market equilibrium.

Parameters A and B can be derived from the price elasticity of demand given the current level of sales:

$$B = -\frac{1}{\varepsilon_p^0}\frac{p^0}{Q^0} \tag{5.4a}$$

$$A = p^0 + BQ^0 = p^0(1 - \frac{1}{\varepsilon_p^0}) \tag{5.4b}$$

As mentioned in section 2.1, the elasticities of demand with respect to conventional price and the national price level are equivalent in magnitude but of opposite sign:

$$\varepsilon_p^0 = -\varepsilon_L^0 \tag{5.5}$$

Where ε_L^0 is the national price level elasticity of demand estimated in section 4.

If we substitute equations 5.4 and 5.5 into 5.2, we get:

$$\Pi^0 = AQ^0 - \frac{1}{2}B(Q^0)^2 = p^0Q^0\left(1 + \frac{1}{\varepsilon_L^0}\right) - \frac{1}{2}\frac{1}{\varepsilon_L^0}p^0Q^0 = R^0\left(1 + \frac{1}{2\varepsilon_L^0}\right) \tag{5.6}$$

Where R^0 is the revenue made by NPL.

The average commercial revenue made by NPL over the last three years was £35.6m[22] – 50% (£17.8m) of it was generated in the UK. Substituting this figure along with the national price level elasticity found in the econometric analysis of section 4 into equation 5.6, we get:

$$\Pi^D = 17.8 \cdot \left(1 + \frac{1}{2} \cdot \frac{1}{1.24}\right) = 25.05 \qquad \text{[£m]}$$

Hence, it is estimated that NPL's work with UK businesses generates £25.05m for supported companies. However, a mentioned at the beginning of this section, this direct return to NPL's paying customers corresponds to additional profits rather than gross value added (GVA). Hence, given that the return to capital (profit) accounts for about 25% of income, the corresponding increase in GVA is probably about four times the result given by equation 5.6, that is, £100.20m. Now, if we take into account that the current level of public funding received by NPL is £83.10m, we end up with a cost-benefit ratio of 1.21.

[22] Instead of using the average revenue generated throughout the whole period considered in the econometric analysis (2001-2017), only the average of the last three years of the dataset has been taken. This is considered to be more representative of the current situation. All revenue figures have been conveniently inflated to 2019 prices based on the GDP deflators provided by the ONS.

Up to this point, only the direct benefits that take place within supported companies have been calculated. Given the nature of the work carried out by the scientific staff of NPL, the benefits spill over to non-supported companies, thus carrying even greater benefits for the UK. As was also mentioned at the top of this heading, Frontier Economics finds that social returns are typically 2 to 3 times larger than private returns. Therefore, we can estimate the indirect benefits by doubling the direct benefits. Moreover, we argued that NPL allows for spillovers to occur even when the collaboration takes place with overseas users. Therefore, doubling the direct benefits to all customers is a reasonable approximation of the overall indirect benefits for the UK. Hence, given that average revenue made by NPL over the last three years was £35.6m, the direct benefit generated for all customers (UK and overseas) is given by:

$$\Pi^{I} = 35.6 \cdot \left(1 + \frac{1}{2} \cdot \frac{1}{1.24}\right) = 49.95 \qquad \text{[£m]}$$

Now, this figure multiplied by 4 gives us the additional gross value added (GVA), £199.82m. Lastly, if we double this number, we get the indirect benefits for the UK, £399.64m, which added to the direct benefits previously computed, yields a total welfare generated of £499.84m and a cost-benefit ratio of 6.01 when the current level of public funding received by NPL is considered.

4.2 The effect of shifts in public funding received by NPL on welfare generated

From an evidence-based policy perspective, it is key to know the effect of a change in the public funding of NPL and the welfare generated. Public funds allow NPL to hire scientific staff who support UK companies develop new product or processes. In other words, since NPL is an organisation that acts as a vehicle to efficiently allocate the public funds needed to complement private spending in measurement R&D, there is a strong relationship between public support and NPL's output. The estimation of such relationship is beyond the scope of this report though. Nonetheless, under the premise that this relationship exists and is stable, it is possible to assess the effect on welfare generated of a variation in NPL's output. Therefore, we will consider that a reduction in public funding triggers an equivalent reduction in output. That is, for every 1% less in public funding, a 1% reduction in output is expected. Figure 4 shows a schematic representation of a shift in output of 20% up (E'') or down (E').

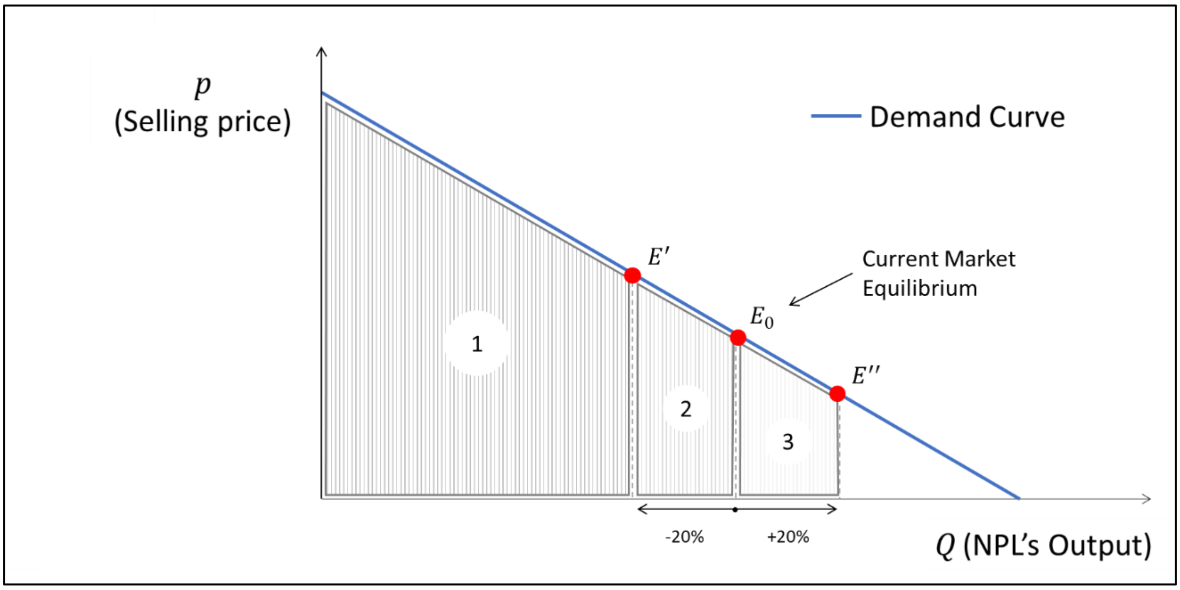


Figure 4: Changes in welfare generated.

The direct benefits to paying customers is currently given by areas 1 and 2 – this has been estimated in the previous section. Area 3 corresponds to the additional welfare that would be created if NPL's output increases by 20% (triggered by an equivalent increase in funding). Conversely, area 2 is what would be lost if NPL's reduces its output by 20% (because of a 20% reduction in funding). Analytically, this can be estimated using equation 5.6, for which parameters A and B stay constant. Thus, the first step is to compute both parameters using the current number of invoices issued in the UK ($3{,}860$ invoices), and the average size of an invoice (£$4{,}624$)[23]. These are given by Table 11:

Parameter	Value
A	£8,354
B	£0,97/Invoice

Table 11: Parameters for the demand curve.

Next, the output (number of invoices issued) in the new equilibria must be calculated (Q' for E', and Q'' for E''), as well as the corresponding level of public funding (F' and F'' respectively). These consist of a 20% positive and negative variation over the current situation. Table 12 shows these alternative equilibria.

Parameter	Value
Q'	3,088 invoices
F'	£66.5m
Q''	4,632 invoices
F''	£99.7m

Table 12: Alternative equilibria.

[23] Again, the last three years of the dataset are considered to be a representative period of the current situation.

Lastly, we can proceed the same way we did in the previous subsection. First, we calculate the change in private profits generated for the supported UK companies using equation 5.6 and multiply by 4 to get the variation in GVA. This gives us the change in direct benefits. Secondly, to get the change in indirect benefits we calculate the GVA variation for all users and multiply by 2. The sum of both variations (direct and indirect benefits) yields the overall change in welfare and cost-benefit ratios[24]. For the 20% decrease in funding scenario the overall welfare is reduced in £77.11m, yielding a decrease in the cost-benefit ratio of -1.16; for the 20% increase in public funds scenario the welfare increases by £65.60m and the cost-benefit ratio by 0.66.

[24] For the sake of simplicity all of these calculations have been omitted since they are equivalent to those made in section 4.1.

5 Conclusion

Our study estimates the price elasticity of demand for NPL's services, finding that it is well above 1 at 95% confidence. This supports the idea that NPL's services are elastic goods for which the quantity demanded will change more than proportionally if the price changes.

We apply standard panel data analysis to a country-level panel dataset to estimate the price elasticity of demand. The data utilised consists of NPL internal invoicing information, as well as data on financial and demographic variables coming from trusted external commonly-used sources (namely, the World Bank, the *Centre d'Etudes Prospectives et d'Informations Internationales*, the Bureau of Weights and Measures, and the United Nations). The proposed analytical setup allows us to control for several variables that are expected to affect the number of services sold. All controls in our model, GDP per capita, distance and the measurement capabilities of local competing NMIs, turned out to be highly significant.

6 References

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In R. R. Nelson, *The Rate and Direction of Inventive Activity: Economic and Social Factors* (pp. 609-625). NJ: Press: Princeton.

Barlev, G. (2007). On the cyclicality of research and development. *American Economic Review, 97*(4), 1131-1164.

Beck, N., & Katz, J. N. (2011). Modeling dynamics in time-series--cross-section political economy data. *Annual Review of Political Science*, 331-352.

Bureau International des Poids et Mesures. (2019, August 27). *BIPM Calibration and Measurement Capabilities - CMCs*. Retrieved September 15, 2019, from https://www.bipm.org/utils/common/pdf/KCDB/KCDB_CMCs.pdf

Davidson, R., & MacKinnon, J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, 781-793.

Frontier Economics. (2014). *Rates of return to investment in science and innovation.* London.

Frontier Economics. (2016). *The Impact of Public Support for Innovation on Firm Outcomes.*

Head, K., & Mayer, T. (2002). *Illusory border effects: Distance mismeasurement inflates estimates of home bias in trade.* Citeseer.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement, 5*(4), 475-492.

HM Treasury. (2018). *The Green Book: Central Government Guidance on Appraisal and Evaluation.*

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, 69-85.

Pregibon, D. (1980). Goodness of link tests for generalized linear models. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 29*(1), 15-24.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological), 31*(2), 350-371.

Tassey, G. (2004). Underinvestment in public good technologies. *The Journal of Technology Transfer, 3*(1-2), 89-113.

Tassey, G. (2008). and measuring the economic roles of technology infrastructure. *Econ. Innov. New Techn., 17*(7-8), 615-629.

Tukey, J. W. (1949). One degree of freedom for non-additivity. *Biometrics, 5*(3), 232-242.

Winning Moves. (2017). *National Measurement System: Customer Needs and Impact Survey.* NMS.

Annex A: Robustness tests

This annex weighs the effect of some of the decisions adopted throughout the analysis. Particularly with regard what countries to include in the analysis, and the methodological approaches taken as to how to deal with the distinct nature of the observations of the UK, the treatment given to outliers, and the definition of the CMC variable.

Countries that do not purchase NPL's services regularly

One of the fundamental features of the dataset is the notable differences among the countries included. For some of them, the size and commercial interrelations with the UK ensures a stable flow of purchases over time. For others, usually small less developed countries, buying services from the NPL is a rare event. This could lead to some sort of bias in the estimation results. Therefore, the effect of excluding countries that do not have a minimum number of observations has been verified. In particular, the effect of excluding countries that have only one observation, up to five observations (models C1 to C5), has been systematically tested. The results of this robustness test are included in Table 13, which shows that the effect is not very significant and that is why the main analysis is carried out with all observations.

DV = log *I*	Model B1		Model C1		Model C2		Model C3		Model C4		Model C5	
	Coeff.	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**	**Coeff.**	**P-value**
log *PL*	1.24 (0.08)	0.000	1.24 (0.08)	0.000	1.23 (0.08)	0.000	1.17 (0.08)	0.000	1.12 (0.07)	0.000	1.13 (0.07)	0.000
log *GDP*	1.33 (0.08)	0.000	1.33 (0.08)	0.000	1.35 (0.07)	0.000	1.40 (0.07)	0.000	1.46 (0.07)	0.000	1.46 (0.07)	0.000
CMC	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000
log *D*	-0.48 (0.02)	0.000	-0.48 (0.02)	0.000	-0.47 (0.02)	0.000	-0.48 (0.02)	0.000	-0.47 (0.02)	0.000	-0.47 (0.02)	0.000
log *POP*	-0.17 (0.03)	0.000	-0.17 (0.03)	0.000	-0.17 (0.03)	0.000	-0.16 (0.03)	0.000	-0.14 (0.03)	0.000	-0.14 (0.03)	0.000
u (t-1)	0.35	0.000	0.35	0.000	0.35	0.000	0.35	0.000	0.35	0.000	0.35	0.000

	(0.05)		(0.05)		(0.05)		(0.05)		(0.05)		(0.05)	
u (t-2)	0.33 (0.04)	0.000	0.33 (0.04)	0.000	0.33 (0.04)	0.000	0.33 (0.04)	0.000	0.33 (0.04)	0.000	0.33 (0.04)	0.000
u (t-3)	0.22 (0.04)	0.000	0.22 (0.04)	0.000	0.22 (0.04)	0.000	0.22 (0.04)	0.000	0.22 (0.04)	0.000	0.21 (0.04)	0.000
Λ	1.28 (0.43)	0.003	-1.24 (0.12)	0.000	-1.24 (0.12)	0.000	-1.26 (0.12)	0.000	-1.28 (0.12)	0.000	-1.27 (0.12)	0.000
	F-stat	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**	**F-stat**	**P-value**
time dummies	17.87	0.000	17.87	0.000	18.50	0.000	19.23	0.000	19.16	0.000	18.49	0.000
R-squared	0.94		0.94		0.94		0.94		0.94		0.94	
Number of obs.	648		648		648		648		648		648	

Table 13: Preferred model after trimming countries with too few observations.

The effect of UK sales in the estimated elasticity of demand

Around 50% of NPL sales take place within the UK. This could have a significant effect on the price elasticity estimation results. Hence, the preferred two-step Heckman model has been run with and without (model D1) the observations of the UK. The results are given by Table 14:

DV = log *I*	Model B1		Model D1	
	Coeff.	**P-value**	**Coeff.**	**P-value**
log *PL*	1.24 (0.08)	0.000	1.15 (0.08)	0.000
log *GDP*	1.33 (0.08)	0.000	1.23 (0.08)	0.000
CMC	-0.02 (0.00)	0.000	-0.01 (0.00)	0.000
log *D*	-0.48 (0.02)	0.000	-0.36 (0.02)	0.000
log *POP*	-0.17 (0.03)	0.000	-0.30 (0.03)	0.000
u (t-1)	0.35 (0.05)	0.000	0.34 (0.05)	0.000
u (t-2)	0.33 (0.04)	0.000	0.33 (0.04)	0.000
u (t-3)	0.22 (0.04)	0.000	0.21 (0.04)	0.000
λ	1.28 (0.43)	0.003	-1.13 (0.12)	0.000
	F-stat	**P-value**	**F-stat**	**P-value**

time dummies	17.87	0.000	14.31	0.000
R-squared	0.94		0.93	
Number of obs.	648		634	

Table 14: Preferred model with and without UK observations.

As seen, the effect of not considering the UK observations is not significant (around a 10% variation).

Outliers in the dependent variable

Figure 1 in section 3 shows a box plot of the dependent variable. Two outliers lay beyond Tukey's fences (represented by the whiskers); they are not far from the rest of the observations though. This suggest that they should not have a significant misleading effect on the estimation of the elasticity of demand. To check that, the preferred model has been tested with and without these two observations (see Table 15).

DV = log *I*	**Model B1**		**Model D1**	
	Coeff.	**P-value**	**Coeff.**	**P-value**
log *PL*	1.24 (0.08)	0.000	1.15 (0.08)	0.000
log *GDP*	1.33 (0.08)	0.000	1.23 (0.08)	0.000
CMC	-0.02 (0.00)	0.000	-0.01 (0.00)	0.000
log *D*	-0.48 (0.02)	0.000	-0.36 (0.02)	0.000
log *POP*	-0.17 (0.03)	0.000	-0.30 (0.03)	0.000
u (t-1)	0.35 (0.05)	0.000	0.34 (0.05)	0.000

u (t-2)	0.33 (0.04)	0.000	0.33 (0.04)	0.000
u (t-3)	0.22 (0.04)	0.000	0.21 (0.04)	0.000
λ	1.28 (0.43)	0.003	-1.13 (0.12)	0.000
	F-stat	**P-value**	**F-stat**	**P-value**
time dummies	17.87	0.000	14.31	0.000
R-squared	0.94		0.93	
Number of obs.	648		634	

Table 15: Preferred model with and without outliers.

As expected, the estimation results are almost equivalent.

Specification of the CMC variable

This subsection addresses the two methodological issues around the variable that controls for the capabilities of the local NMI:

- On the one hand, there is the question of which one of the two variables constructed using the table reported by BIPM to use; either the probabilistic measure of the NMI being able to fulfil the user's needs (based on the total number of services provided), or the geometric mean of the number of capabilities in different areas of metrology, that tends to favour those NMIs with more variety in their portfolios.
- On the other hand, once the appropriate CMC variable is selected, we have to decide whether to include this in levels or in logarithmic form in our model.

Therefore, firstly we test which one of the two alternative CMC measures is more suitable for our model. To that end, both have been used to estimate the national price level elasticity of demand through the preferred Heckman corrected setup – both models yield very similar results. This is shown by Table 16 which compares the preferred model B1 that uses probabilistic CMC variable, against model F1 which includes instead the geometric mean CMC variable.

DV = log *I*	**Model B1**		**Model F1**	
	Coeff.	**P-value**	**Coeff.**	**P-value**

log *PL*	1.24 (0.09)	0.000	1.22 (0.08)	0.000
log *GDP*	1.33 (0.09)	0.000	1.28 (0.08)	0.000
CMC	-0.02 (0.04)	0.000	-0.01 (0.00)	0.000
log *D*	-0.48 (0.02)	0.000	-0.49 (0.02)	0.000
log *POP*	-0.17 (0.03)	0.000	-0.22 (0.03)	0.000
u (t-1)	0.35 (0.05)	0.000	0.35 (0.05)	0.000
u (t-2)	0.33 (0.05)	0.000	0.33 (0.04)	0.000
u (t-3)	0.22 (0.05)	0.000	0.22 (0.04)	0.000
λ	0.28 (0.18)	0.136	0.25 (0.14)	0.047
	F-stat	**P-value**	**F-stat**	**P-value**
time dummies	26.69	0.000	17.71	0.000
R-squared	0.94		0.94	
Number of obs.	564		648	

Table 16: Preferred model using the two different definitions of the CMC variable.

The fact that both variables produce almost equivalent results reflects that those NMIs that have a more comprehensive portfolio (greater number of services available to the user) are also those that have a more complete portfolio (well distributed in all the areas of metrology). This was to be expected, because although some NMIs decide to focus on certain areas of metrology, it is usual to develop the measurement national infrastructure in all areas equally. In any case, the probabilistic measure of the CMC has been taken as a reference in the main analysis, since its interpretation is more immediate.

Once the CMC variable has been chosen, we need to address the question of whether the CMC variable should be introduced in the model in levels or logged. This affects the estimation results of the national price level elasticity substantially. Hence, it is essential to determine whether our specification in levels is appropriate or not. For that matter, once again, the preferred Heckman-corrected setup (model B1) is compared to an equivalent specification which only differs in that the CMC regressor is logged instead of in levels (model G1). In principal, running a link test should be a convenient way to detect any issue on the functional relation between the dependent variable and the regressor. However, as shown by Table 17, both specifications pass the test satisfactorily.

.DV = log *I*	Model B1		Model G1	
	Coeff.	**P-value**	**Coeff.**	**P-value**
predicted values	1.03	0.000	1.03	0.000
	(0.11)		(0.08)	
squared prediction	0.00	0.768	0.00	0.784
	(0.00)		(0.07)	
R-squared	0.94		0.94	
Number of obs.	648		648	

Table 17: Link test for the preferred model using the two different definitions of the CMC variable

Hence, another complementary approach to decide whether the CMC variable should appear in levels or in logarithmic form is needed. Following Davidson and MacKinnon we can test for this by verifying if the fitted values of the alternative model have any explanatory power in the original model. Hence, if our choice of including the CMC variable in levels is correct, then the fitted values of the model that considers this variable in logarithmic form should be insignificant when added to our model. Table 18 shows the results for this test.

DV = log *I*	Model B1		Model G1	
	Coeff.	**P-value**	**Coeff.**	**P-value**
log *PL*	-0.05	0.772	0.80	0.054
	(0.77)		(0.41)	
log *GDP*	0.50	0.470	1.19	0.032
	(0.47)		(0.56)	
CMC	-0.07	0.548	-0.02	0.033
	(0.55)		(0.01)	
log *D*	-0.06	0.646	-0.35	0.040
	(0.65)		(0.17)	

log *POP*	-0.05 (0.63)	0.632	-0.12 (0.08)	0.119
u (t-1)	0.09 (0.50)	0.496	0.28 (0.15)	0.063
u (t-2)	0.00 (1.00)	0.996	0.21 (0.11)	0.068
u (t-3)	0.01 (0.91)	0.908	0.14 (0.08)	0.079
λ	0.14 (0.47)	0.469	0.35 (0.22)	0.110
predicted for alternative model	0.81 (0.36)	0.023	0.23 (0.37)	0.534
	F-stat	**P-value**	**F-stat**	**P-value**
time dummies	0.28	0.995	0.37	0.979
R-squared	0.94		0.94	
Number of obs.	564		564	

Table 18: Davidson-MacKinnon test.

The results are highly clarifying. The fitted values of the alternative model are not significant when the CMC variable is included in the model in levels. Conversely, when the CMC variable is entered in logarithmic form, none of the variables of interest are significant and the predicted values of the alternative are highly significant. This provides assurance that our chosen specification is adequate.

Negative binomial regression

Although the gravity model of trade has been widely used traditionally to predict and explain trade flows across countries, there is no consensus about the optimal method to solve the existence of zero flows. Especially when analysing trade patterns at the product level, the possibility that two countries do not trade in that specific good is much higher, and therefore the problem of the existence of zeros in the dataset is more acute.

Count models are an alternative approach to the Heckman selection model used in this paper. In these models the dependent variable is introduced in levels rather than logged. The negative binomial regression is a generalisation of the Poisson regression model which loosens the restrictive assumption that the variance is equal to the mean. We can apply this model to our dataset given that the dependent variable in levels shows significant overdispersion. Table 19 compares the estimation results for the static specification using OLS and the negative binomial regression[25].

DV = log *I*	**Model A1**		**Model H1**	
	Coeff.	**P-value**	**Coeff.**	**P-value**
log *PL*	1.31 (0.10)	0.000	1.25 (0.13)	0.000
log *GDP*	1.15 (0.08)	0.000	1.28 (0.13)	0.000
CMC	-0.02 (0.00)	0.000	-0.05 (0.00)	0.000
log *D*	-0.44 (0.03)	0.000	-0.56 (0.04)	0.000
log *POP*	-0.20 (0.04)	0.000	0.33 (0.03)	0.000
	F-stat	**P-value**	**Chi2**	**P-value**
time dummies	7.13	0.000	112.70	0.000
Pseudo R-squared	0.81		0.10	
Number of obs.	939		1377	

Table 19: Negative binomial regression estimation results.

[25] Note that neither dynamics of the system (i.e. including past realisations of the error term) nor the excess of zeroes (by considering for example a zero-inflated negative binomial regression model) are accounted by this model. The purpose of model H1 is to serve as a robustness check that the estimation procedure does not have a significant impact on the estimation results, and not as a complete alternative approach to the Heckman selection model used in the main body of the paper.

We can see how the price elasticity found is quite similar to the ones found by the models tested in the main body of this paper. This provides further assurance that our empirical results are sound.

Split sample test

The macroeconomic and commercial conditions that govern the exchange of goods at the international level today are not the same as in 2001. Likewise, the market for high precision calibration services may have undergo significant changes during the last two decades. Both factors may influence NPL's sales in other countries and in the UK. Indeed, it could be that the estimated price elasticity differs substantially from the actual one at present. In other words, our analysis comes across an important trade-off. As more years are added to the sample, we have more variability that allows us to find reliable estimates. However, it could be that the initial years had little to do with the current situation.

For this reason, a simple robustness test is presented below. It consists of the static specification (model A1) for a subsample from 2010 onwards. We estimate the static model instead of the preferred Heckman corrected one because as soon as the dynamics of the system are taken into account, we inevitably lose observations; this fact plus the reduction in the sample will leave us with too few observations to play with.

Therefore, the objective of this robustness test is to compare the estimates from both static models and verify that both estimates are comparable. Table 20 shows the estimation results for the full sample static model (model A1) and the restricted sample static model (model I1).

DV = log *I*	**Model A1**		**Model I1**	
	Coeff.	**P-value**	**Coeff.**	**P-value**
log *PL*	1.31 (0.10)	0.000	1.23 (0.14)	0.000
log *GDP*	1.15 (0.08)	0.000	1.27 (0.12)	0.000
CMC	-0.02 (0.00)	0.000	-0.02 (0.00)	0.000
log *D*	-0.44 (0.03)	0.000	-0.47 (0.04)	0.000
log *POP*	-0.20 (0.04)	0.000	-0.24 (0.05)	0.000

	F-stat	P-value	F-stat	P-value
time dummies	7.13	0.000	1.60	0.121
R-squared	0.81		0.80	
Number of obs.	939		524	

Table 20: Split test for the static model.

It can be seen how both estimates for the coefficient of the national price level regressor are very similar and within the confidence interval of the other. On the other hand, the year dummies are not jointly significant (and in fact individually), which could show that the years in the second half of the sample (after the huge shock of 2008 and 2009) is more comparable to each other.

In any case, this simple check points to the validity of the estimates obtained.

Income as the dependent variable

A fundamental aspect of the econometric analysis is the choice of the dependent variable. We have two possible variables to use as our dependent variable: the number of invoices and the income generated (both normalised by population to ensure comparability across countries).

The objective of this robustness test is to check the impact on the estimation results of using income instead of the preferred dependent variable, the number of invoices. To that end, the following specification is tested:

$$\log R_{i,t} = \beta_0 + \beta_1 \log PL_{i,t} + \beta_2 \log GDP_{i,t} + \beta_3 CMC_i + \beta_4 \log D_i \\ + \beta_5 \log POP_i + \beta_5 \log AvPrice_{i,t} + \gamma_t Y_t + \varepsilon_{i,t}$$

where $R_{i,t}$ is income generated from customers of country i in year t normalised by the population of the country, $PL_{i,t}$ is the national price level, $GDP_{i,t}$ is the gross domestic product per capita, CMC_i is a continuous index that rates the measurement capabilities of the country, D_i is the distance between the country and the UK, POP_i is the population of the country, $AvPrice_{i,t}$ is the average income per invoice, Y_t is a year dummy variable, and $\varepsilon_{i,t}$ is the error term which is assumed to be independently and identically distributed as a normal distribution with mean of 0 and finite variance.

Note that this specification differs in two ways from the one considered in the main body of the document: (1) we are using income as the dependent variable instead of the number of invoices, and (2) we are including the average price per invoice as a regressor. The reason for including this other regressor is to get the Q component in income; i.e. given that $R = P \cdot Q$, we know that:

$$\log R_{i,t} = \log P_t + \log Q_{i,t}$$

Hence, by adding the average price per invoice to the regressors we can approximate $\log Q_{i,t}$.

We can compare the estimation results for the static version of the model both using the number of invoices (model A1) and income (model J1)

DV = log *I*	**Model A1**		**Model J1**	
	Coeff.	**P-value**	**Coeff.**	**P-value**
log *PL*	1.31 (0.10)	0.000	1.23 (0.10)	0.000
log *GDP*	1.15 (0.08)	0.000	1.09 (0.08)	0.000
CMC	-0.02 (0.00)	0.000	-0.01 (0.00)	0.000
log *D*	-0.44 (0.03)	0.000	-0.36 (0.03)	0.000
log *POP*	-0.20 (0.04)	0.000	0.70 (0.03)	0.000
log *AvPrice*	. .	.	1.01 (0.04)	0.000
	F-stat	**P-value**	**F-stat**	**P-value**
time dummies	7.13	0.000	7.04	0.000
R-squared	0.81		0.80	
Number of obs.	939		909	

Table 21: Static model using both available dependent variables.

Therefore, it is found that both estimates are very similar for all regressors of interest.

Annex B: Testing for endogeneity.

This annex discusses the alternative to the Hausman test proposed by . This method is used in section 4.3 to assess the whether the regressor of interest in our model, the price level, is correlated with the error term. If so, any the estimation results would be biased.

Firstly, the computation of the test proposed by Mundlak is presented. Then the intuition behind it is described.

In panel data analysis, the decision of using the fixed effects estimator or the random effects estimator depends on how time-invariant unobservables are related to the variables in the model. To assess this, Mundlak's test follows three steps:

1. The panel-level average of the time-varying covariates are computed.
2. The random effects estimator is used to regress the covariates and the panel-level means generated in the first step against the dependent variable.
3. Test whether the panel-level means are jointly zero.

If the test rejects the null hypothesis that the coefficients are jointly zero, it suggests that there is correlation between the time-invariant unobservables and the regressors, namely, the fixed-effects assumptions are satisfied. On the contrary, If the test does not reject the null hypothesis that the generated panel-level mean regressors are zero, there is evidence of no correlation between the time-invariant unobservable and the regressors; that is, the random effects assumptions are satisfied.

The intuition behind Mundlak's approach is straightforward. Suppose a linear panel-data model given by equation B.1:

$$y_{it} = \alpha_i + x_{it}\beta + \varepsilon_{it} \qquad \text{(B.1)}$$

Where the index i denotes the panel unit and the index t time. y_{it} is the outcome of interest, x_{it} is the set of regressors, ε_{it} is the time-varying unobservable (idiosyncratic error), and α_i is the time-invariant unobservable (unobserved heterogeneity).

Now, the mean of α_i conditional on the time-invariant part of the regressors is given by expression B.2:

$$\alpha_i = \bar{x}_i\gamma + \nu_i \Leftrightarrow \mathbf{E}(\alpha_i|x_i) = \bar{x}_i\gamma \qquad \text{(B.2)}$$

Where $\bar{x}_i$ is the panel-level mean of x_{it}, and ν_i is a time-invariant unobservable that is uncorrelated to the regressors.

Hence, if $\gamma = 0$, α_i and the covariates are uncorrelated. This is precisely what the Mundlak method tests. This can be seen by substituting B.2 into B.1:

$$y_{it} = \alpha_i + x_{it}\beta + \varepsilon_{it} = \overline{x_i}\gamma + \nu_i + x_{it}\beta + \varepsilon_{it} \Leftrightarrow \mathbf{E}(y_{it}|x_{it}) = x_{it}\beta + \overline{x_i}\gamma \qquad \text{(B.3)}$$

Where the last relation relies on the fact that the regressors and the unobservables are mean independent. Therefore, Mundlak's approach tests the null hypothesis of $\gamma = 0$.

Annex C: Detailed summary statistics

Table 22 provides detailed *overall*, *between* and *within* summary statistics for all the variables used in the analysis. Variables in level are abbreviated in capital letters; variables in logarithmic form are lowercased.

Variable	Variation	Mean	Std. Dev	Min	Max	Obs.
	overall		8.0E-06	0.0	8.6E-05	N =1717
I	between	2.4E-06	6.9E-06	7.5E-10	5.8E-05	n =101
	within		4.2E-06	-3.9E-05	6.6E-05	T =17
	overall		0.3	0.1	1.4	N =1649
PL	between	0.5	0.3	0.2	1.2	n =97
	within		0.1	0.2	1.0	T =17
	overall		1.4E+04	4.1E+02	9.7E+04	N =1649
GDP	between	1.5E+04	1.3E+04	6.2E+02	7.0E+04	n =97
	within		4.9E+03	-2.3E+03	4.7E+04	T =17
	overall		16.7	0	77.6	N =1394
CMC	between	11.6	16.8	0	77.6	n =82
	within		0.0	11.6	11.6	T =17
	overall		3933.0	185.8	19147.1	N =1683
D	between	5167.5	3951.9	185.8	19147.1	n =99
	within		0	5167.5	5167.5	T =17
	overall		1.8E+08	3.3E+04	1.3E+09	N =1717
POP	between	5.9E+07	1.8E+08	3.3E+04	1.3E+09	n =101
	within		0	5.9E+07	5.9E+07	T =17
	overall		2.1	-20.3	-9.4	N =1025
i	between	-14.1	2.1	-18.7	-9.8	n =101
	within		0.6	-16.7	-11.4	T-bar =10.15
	overall		0.5	-2.1	0.4	N =1649
pl	between	-0.7	0.5	-1.6	0.2	n =97
	within		0.2	-1.7	0.0	T =17
	overall		1.1	6.0	11.5	N =1649
gdp	between	9.2	1.0	6.4	11.1	n =97
	within		0.3	8.3	10.2	T =17
d	overall	8.2	1.0	5.2	9.9	N =1683
	between		1.0	5.2	9.9	n =99

	Within		0	8.2	8.2	T =17
	overall		1.9	10.4	21.0	N =1717
pop	between	16.3	1.9	10.4	21.0	n =101
	within		0	16.3	16.3	T =17

Table 23: Summary statistics.

Annex D: Stata outputs

Model A1

Estimation results

Linear regression

Number of obs	=	939
F(21, 917)	=	180.94
Prob > F	=	0.0000
R-squared	=	0.8066
Root MSE	=	.9328

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.314002	.1028418	12.78	0.000	1.112169	1.515835
LnWB_GDPpcgbp	1.153907	.0789686	14.61	0.000	.9989267	1.308887
BIPM_CMCs	-.0199687	.0026864	-7.43	0.000	-.0252409	-.0146964
LnCEPII_dist	-.4394035	.0329408	-13.34	0.000	-.5040515	-.3747554
LnISO_Population	-.2033567	.0354115	-5.74	0.000	-.2728538	-.1338596
YEAR						
2002	1.001832	.1848393	5.42	0.000	.6390748	1.364589
2003	.9818643	.194761	5.04	0.000	.5996353	1.364093
2004	1.21669	.1878141	6.48	0.000	.848095	1.585286
2005	1.40412	.1895101	7.41	0.000	1.032196	1.776043
2006	1.141488	.1852813	6.16	0.000	.7778637	1.505113
2007	1.264889	.1789166	7.07	0.000	.913755	1.616022
2008	.6823026	.1995518	3.42	0.001	.2906713	1.073934
2009	.5747391	.1848468	3.11	0.002	.2119672	.9375109
2010	.6944648	.187167	3.71	0.000	.3271395	1.06179
2011	.6846496	.197213	3.47	0.001	.2976083	1.071691
2012	.65513	.1865352	3.51	0.000	.2890444	1.021216
2013	.6602773	.1848684	3.57	0.000	.297463	1.023092
2014	.725009	.1901153	3.81	0.000	.3518973	1.098121
2015	.7314882	.184208	3.97	0.000	.3699701	1.093006
2016	.4271422	.1871181	2.28	0.023	.0599127	.7943716
2017	.2716416	.1943123	1.40	0.162	-.1097069	.65299
_cons	-18.15977	1.130125	-16.07	0.000	-20.3777	-15.94184

Wald test year dummies

```
( 1)  2002.YEAR = 0
( 2)  2003.YEAR = 0
( 3)  2004.YEAR = 0
( 4)  2005.YEAR = 0
( 5)  2006.YEAR = 0
( 6)  2007.YEAR = 0
( 7)  2008.YEAR = 0
( 8)  2009.YEAR = 0
( 9)  2010.YEAR = 0
(10)  2011.YEAR = 0
(11)  2012.YEAR = 0
(12)  2013.YEAR = 0
(13)  2014.YEAR = 0
(14)  2015.YEAR = 0
(15)  2016.YEAR = 0
(16)  2017.YEAR = 0

      F( 16,   917) =     7.13
           Prob > F =    0.0000
```

Link test

```
      Source |       SS           df       MS      Number of obs   =       939
-------------+----------------------------------   F(2, 936)       =   2093.06
       Model |  3535.3053         2  1767.65265   Prob > F        =    0.0000
    Residual | 790.482037       936  .844532091   R-squared       =    0.8173
-------------+----------------------------------   Adj R-squared   =    0.8169
       Total | 4325.78733       938  4.61171357   Root MSE        =    .91898

------------------------------------------------------------------------------
LnInvoices~p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   1.073418   .1950273     5.50   0.000     .6906766    1.456159
      _hatsq |   .0024892   .0065915     0.38   0.706    -.0104467    .0154251
       _cons |   .5315051   1.425031     0.37   0.709     -2.26512    3.328131
------------------------------------------------------------------------------
```

Ramsey RESET test

```
Ramsey RESET test using powers of the fitted values of LnInvoicesPop
       Ho:  model has no omitted variables
                 F(3, 914) =     12.02
                  Prob > F =      0.0000
```

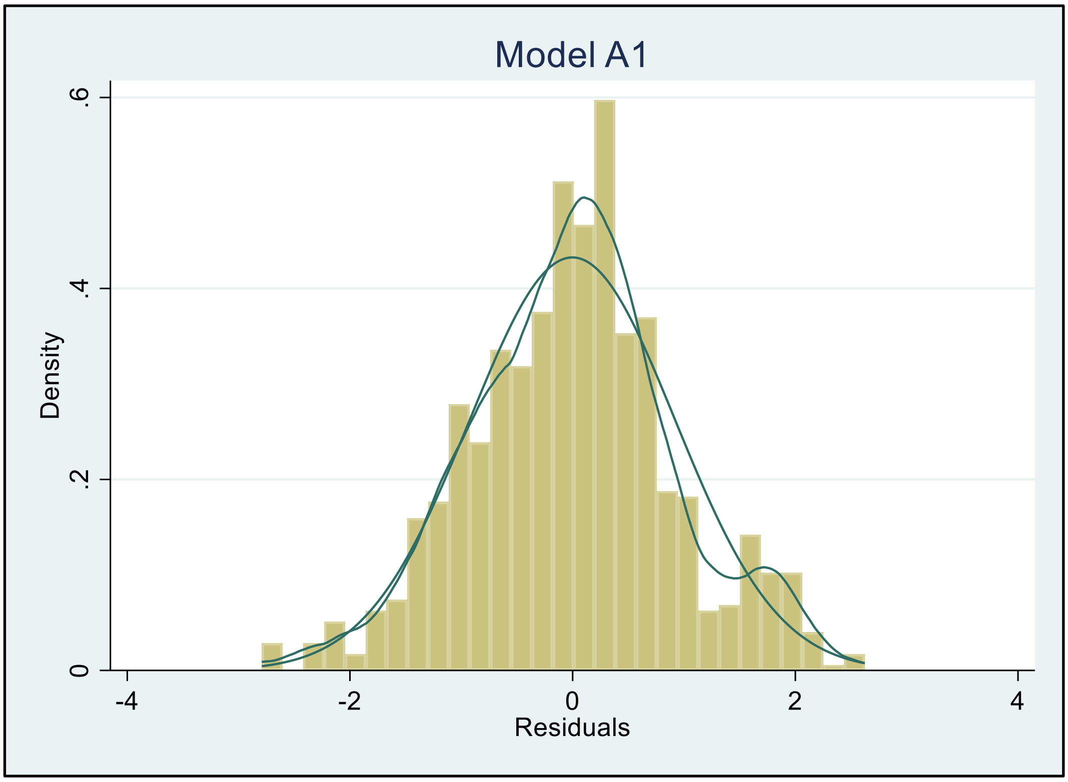
Serial correlation test

Source	SS	df	MS
Model	380.427527	1	380.427527
Residual	283.186566	800	.353983208
Total	663.614094	801	.828482014

Number of obs	=	802
F(1, 800)	=	1074.71
Prob > F	=	0.0000
R-squared	=	0.5733
Adj R-squared	=	0.5727
Root MSE	=	.59496

resid	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
resid						
L1.	.7566804	.0230817	32.78	0.000	.7113726	.8019883
_cons	.0164921	.0210185	0.78	0.433	-.0247658	.0577499

Normality of the residuals



Mundlak test

```
Random-effects GLS regression                   Number of obs     =        939
Group variable: ID                              Number of groups  =         81

R-sq:                                           Obs per group:
     within  = 0.3233                                         min =          1
     between = 0.8613                                         avg =       11.6
     overall = 0.8082                                         max =         17

                                                Wald chi2(23)     =    1242.79
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

                                        (Std. Err. adjusted for 81 clusters in ID)
------------------------------------------------------------------------------------
                   |               Robust
     LnInvoicesPop |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
            LnWB_PRL |   -.043811   .1883329    -0.23   0.816    -.4129366    .3253146
        LnWB_GDPpcgbp |   .8526772   .2162268     3.94   0.000     .4288805    1.276474
          BIPM_CMCs |  -.0149248   .0092068    -1.62   0.105    -.0329698    .0031203
        LnCEPII_dist |  -.3988639   .1181946    -3.37   0.001     -.630521   -.1672068
    LnISO_Population |  -.3157267   .1023604    -3.08   0.002    -.5163493    -.115104
        LnWB_PRLmean |   1.830448   .3376485     5.42   0.000     1.168669    2.492226
  LnWB_GDPpcgbpmean |  -.1043759   .2546585    -0.41   0.682    -.6034973    .3947455
                   |
              YEAR |
              2002  |   .9878086   .1098965     8.99   0.000     .7724155    1.203202
              2003  |    1.00638   .1266705     7.94   0.000     .7581108     1.25465
              2004  |   1.182101   .1288598     9.17   0.000       .92954    1.434661
              2005  |   1.389886   .1268374    10.96   0.000     1.141289    1.638482
              2006  |   1.227675   .1283833     9.56   0.000     .9760485    1.479302
              2007  |   1.245081   .1285922     9.68   0.000     .9930447    1.497117
              2008  |   1.050825   .1299461     8.09   0.000     .7961355    1.305515
              2009  |   1.007699   .1465433     6.88   0.000     .7204791    1.294918
              2010  |    1.21944   .1493779     8.16   0.000     .9266644    1.512215
              2011  |   1.190064   .1555207     7.65   0.000     .8852492    1.494879
              2012  |   1.158128   .1508161     7.68   0.000     .8625342    1.453722
              2013  |   1.223071   .1602709     7.63   0.000      .908946    1.537196
              2014  |   1.218597     .14362     8.48   0.000     .9371073    1.500087
              2015  |   1.138984    .150618     7.56   0.000     .8437777     1.43419
              2016  |   1.008738    .192704     5.23   0.000     .6310451    1.386431
              2017  |    .986852   .2419489     4.08   0.000      .512641    1.461063
                   |
             _cons |  -12.87396   2.407382    -5.35   0.000    -17.59234    -8.15558
-------------------+----------------------------------------------------------------
           sigma_u |  .78784645
           sigma_e |  .52782908
               rho |  .69020147   (fraction of variance due to u_i)
------------------------------------------------------------------------------------
```

```
 ( 1)  LnWB_PRLmean = 0
 ( 2)  LnWB_GDPpcgbpmean = 0

           chi2(  2) =   29.43
         Prob > chi2 =    0.0000
```

Model A2

Estimation results

Linear regression

Number of obs	=	801
F(21, 779)	=	401.90
Prob > F	=	0.0000
R-squared	=	0.9167
Root MSE	=	.59722

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.24861	.0806192	15.49	0.000	1.090354	1.406867
LnWB_GDPpcgbp	1.233787	.0651217	18.95	0.000	1.105953	1.361622
BIPM_CMCs	-.0203153	.001707	-11.90	0.000	-.0236662	-.0169645
LnCEPII_dist	-.4545122	.0181831	-25.00	0.000	-.4902059	-.4188186
LnISO_Population	-.1983297	.0235201	-8.43	0.000	-.2445	-.1521594
U						
L1.	.7585594	.0274461	27.64	0.000	.7046824	.8124364
YEAR						
2003	-.0478824	.1351532	-0.35	0.723	-.3131901	.2174252
2004	.2038653	.1333478	1.53	0.127	-.0578982	.4656288
2005	.3514206	.1253772	2.80	0.005	.1053034	.5975379
2006	.1823697	.1333442	1.37	0.172	-.0793869	.4441263
2007	.1840757	.1298663	1.42	0.157	-.0708537	.439005
2008	-.1434227	.1375788	-1.04	0.298	-.4134918	.1266464
2009	-.4572319	.1267269	-3.61	0.000	-.7059986	-.2084653
2010	-.2710897	.1276832	-2.12	0.034	-.5217336	-.0204458
2011	-.4548417	.1404845	-3.24	0.001	-.7306147	-.1790687
2012	-.3545824	.121594	-2.92	0.004	-.5932731	-.1158916
2013	-.3256319	.1303402	-2.50	0.013	-.5814915	-.0697722
2014	-.3018151	.1275169	-2.37	0.018	-.5521326	-.0514976
2015	-.2577654	.1246292	-2.07	0.039	-.5024142	-.0131165
2016	-.6442882	.1256325	-5.13	0.000	-.8909066	-.3976698
2017	-.8048086	.1420982	-5.66	0.000	-1.083749	-.5258679
_cons	-17.88848	.9017248	-19.84	0.000	-19.65858	-16.11838

Wald test year dummies

```
 ( 1)  2003.YEAR = 0
 ( 2)  2004.YEAR = 0
 ( 3)  2005.YEAR = 0
 ( 4)  2006.YEAR = 0
 ( 5)  2007.YEAR = 0
 ( 6)  2008.YEAR = 0
 ( 7)  2009.YEAR = 0
 ( 8)  2010.YEAR = 0
 ( 9)  2011.YEAR = 0
 (10)  2012.YEAR = 0
 (11)  2013.YEAR = 0
 (12)  2014.YEAR = 0
 (13)  2015.YEAR = 0
 (14)  2016.YEAR = 0
 (15)  2017.YEAR = 0

       F( 15,   779) =   11.23
            Prob > F =    0.0000
```

Link test

Source	SS	df	MS
Model	3246.09492	2	1623.04746
Residual	272.881765	798	.341957099
Total	3518.97668	800	4.39872085

Number of obs	=	801
F(2, 798)	=	4746.35
Prob > F	=	0.0000
R-squared	=	0.9225
Adj R-squared	=	0.9223
Root MSE	=	.58477

LnInvoices~p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_hat	1.311802	.1223507	10.72	0.000	1.071635	1.551969
_hatsq	.0107206	.0041919	2.56	0.011	.0024921	.0189491
_cons	2.22138	.8807734	2.52	0.012	.4924733	3.950286

Ramsey RESET test

```
Ramsey RESET test using powers of the fitted values of LnInvoicesPop
       Ho:  model has no omitted variables
                 F(3, 776) =      6.67
                  Prob > F =      0.0002
```

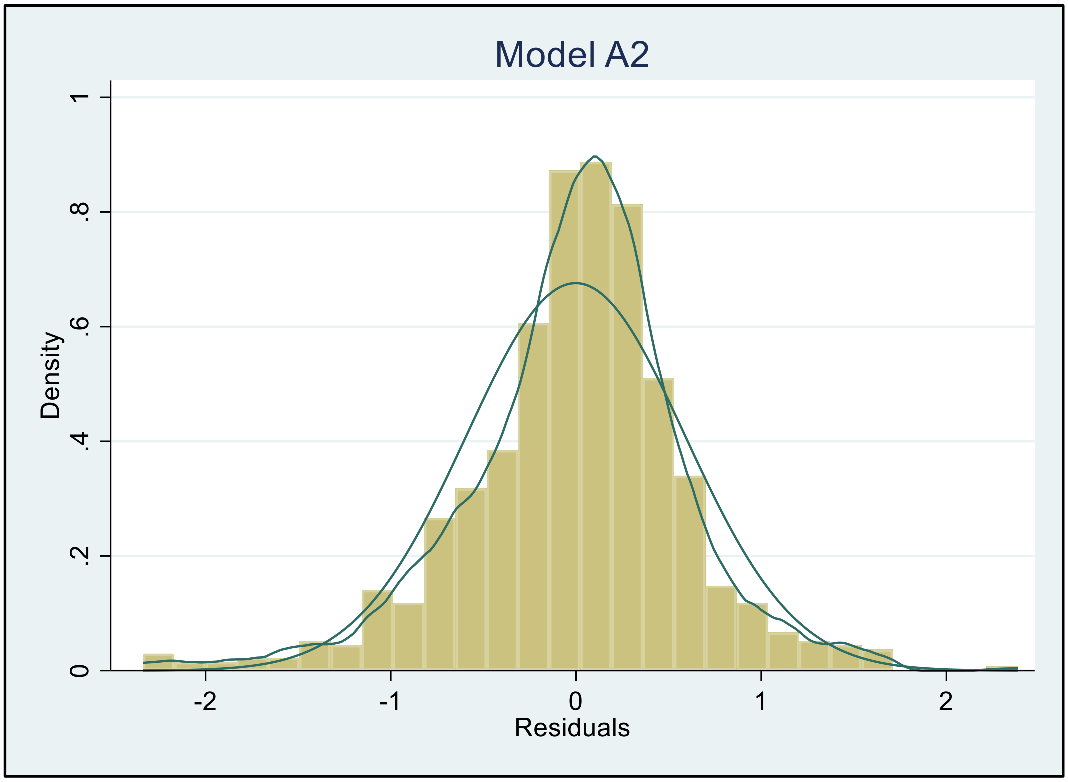
Serial correlation test

Source	SS	df	MS			
Model	19.7176045	1	19.7176045			
Residual	224.329231	714	.314186598			
Total	244.046835	715	.341324245			

Number of obs	=	716
F(1, 714)	=	62.76
Prob > F	=	0.0000
R-squared	=	0.0808
Adj R-squared	=	0.0795
Root MSE	=	.56052

resid	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
resid						
L1.	-.2875954	.0363035	-7.92	0.000	-.3588698	-.216321
_cons	.0056965	.0209481	0.27	0.786	-.0354307	.0468237

Normality of the residuals



Mundlak test

```
Random-effects GLS regression                   Number of obs     =        801
Group variable: ID                              Number of groups  =         69

R-sq:                                           Obs per group:
     within  = 0.1404                                         min =          1
     between = 0.9818                                         avg =       11.6
     overall = 0.9169                                         max =         16

                                                Wald chi2(23)     =    9475.47
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

                                       (Std. Err. adjusted for 69 clusters in ID)
```

LnInvoicesPop	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	1.002301	.1695931	5.91	0.000	.6699046	1.334697
LnWB_GDPpcgbp	1.241972	.1848001	6.72	0.000	.8797707	1.604174
BIPM_CMCs	-.0204574	.002707	-7.56	0.000	-.025763	-.0151518
LnCEPII_dist	-.450655	.0306665	-14.70	0.000	-.5107603	-.3905497
LnISO_Population	-.1987053	.0354847	-5.60	0.000	-.268254	-.1291566
U						
L1.	.7534812	.0435316	17.31	0.000	.6681608	.8388015
LnWB_PRLmean	.2766888	.2246772	1.23	0.218	-.1636705	.717048
LnWB_GDPpcgbpmean	-.021668	.1930202	-0.11	0.911	-.3999806	.3566446
YEAR						
2003	-.0452043	.1455296	-0.31	0.756	-.330437	.2400284
2004	.203408	.127875	1.59	0.112	-.0472224	.4540383
2005	.3543306	.1168089	3.03	0.002	.1253893	.583272
2006	.1875456	.1353537	1.39	0.166	-.0777427	.4528339
2007	.1839279	.12554	1.47	0.143	-.062126	.4299819
2008	-.1056411	.1415592	-0.75	0.456	-.3830921	.1718099
2009	-.401968	.1575953	-2.55	0.011	-.710849	-.0930869
2010	-.2021393	.1660991	-1.22	0.224	-.5276875	.1234088
2011	-.3824271	.156228	-2.45	0.014	-.6886284	-.0762258
2012	-.2894103	.1475682	-1.96	0.050	-.5786386	-.000182
2013	-.2576486	.1636314	-1.57	0.115	-.5783603	.0630631
2014	-.2524405	.141267	-1.79	0.074	-.5293187	.0244376
2015	-.2269928	.1714713	-1.32	0.186	-.5630704	.1090847
2016	-.5914459	.19774	-2.99	0.003	-.9790092	-.2038826
2017	-.7341057	.2287121	-3.21	0.001	-1.182373	-.2858383
_cons	-17.79782	1.296759	-13.72	0.000	-20.33942	-15.25622
sigma_u	0					
sigma_e	.49783632					
rho	0	(fraction of variance due to u_i)				

```
( 1)  LnWB_PRLmean = 0
( 2)  LnWB_GDPpcgbpmean = 0

          chi2(  2) =    1.57
        Prob > chi2 =    0.4568
```

Model A3

Estimation results

```
Linear regression                               Number of obs     =        648
                                                F(21, 626)        =     487.39
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9373
                                                Root MSE          =      .5058

------------------------------------------------------------------------------
             |               Robust
LnInvoicesPop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     LnWB_PRL |   1.192111   .0734575    16.23   0.000     1.047858    1.336364
 LnWB_GDPpcgbp |     1.2537     .06578    19.06   0.000     1.124524    1.382876
    BIPM_CMCs |  -.0202347   .0016867   -12.00   0.000     -.023547   -.0169223
  LnCEPII_dist |  -.4672374   .0161809   -28.88   0.000    -.4990128   -.4354621
LnISO_Population |  -.2025394   .0248424    -8.15   0.000    -.2513241   -.1537548
             |
           U |
         L1. |   .3603281   .0452495     7.96   0.000     .2714689    .4491873
         L2. |   .3347188   .0424106     7.89   0.000     .2514346     .418003
         L3. |   .2171507   .0407321     5.33   0.000     .1371626    .2971389
             |
        YEAR |
        2005 |   .1809868   .0957039     1.89   0.059    -.0069527    .3689263
        2006 |  -.0674422   .1123515    -0.60   0.549    -.2880737    .1531894
        2007 |  -.0468298   .1138695    -0.41   0.681    -.2704422    .1767826
        2008 |  -.4005389   .1108125    -3.61   0.000    -.6181481   -.1829298
        2009 |   -.628931   .1036749    -6.07   0.000    -.8325238   -.4253382
        2010 |  -.4978494   .1060187    -4.70   0.000    -.7060447    -.289654
        2011 |  -.7054619   .1182694    -5.96   0.000    -.9377148    -.473209
        2012 |  -.6617262    .103049    -6.42   0.000    -.8640897   -.4593626
        2013 |  -.5566118   .1134433    -4.91   0.000    -.7793874   -.3338362
        2014 |  -.5296157    .100073    -5.29   0.000    -.7261351   -.3330964
        2015 |  -.4985596   .0977432    -5.10   0.000     -.690504   -.3066153
        2016 |  -.8685235   .1078369    -8.05   0.000    -1.080289   -.6567576
        2017 |  -1.157288   .1141327   -10.14   0.000    -1.381418   -.9331588
             |
       _cons |  -17.70103    .949147   -18.65   0.000    -19.56493   -15.83713
------------------------------------------------------------------------------
```

Wald test year dummies

```
 ( 1)  2005.YEAR = 0
 ( 2)  2006.YEAR = 0
 ( 3)  2007.YEAR = 0
 ( 4)  2008.YEAR = 0
 ( 5)  2009.YEAR = 0
 ( 6)  2010.YEAR = 0
 ( 7)  2011.YEAR = 0
 ( 8)  2012.YEAR = 0
 ( 9)  2013.YEAR = 0
 (10)  2014.YEAR = 0
 (11)  2015.YEAR = 0
 (12)  2016.YEAR = 0
 (13)  2017.YEAR = 0

       F( 13,   626) =   18.59
            Prob > F =    0.0000
```

Link test

```
      Source |       SS           df       MS      Number of obs   =       648
-------------+----------------------------------   F(2, 645)       =   5172.08
       Model |  2543.23431         2  1271.61716   Prob > F        =    0.0000
    Residual |  158.580787       645  .245861686   R-squared       =    0.9413
-------------+----------------------------------   Adj R-squared   =    0.9411
       Total |   2701.8151       647  4.17591206   Root MSE        =    .49584

------------------------------------------------------------------------------
LnInvoices~p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   1.038406   .1122312     9.25   0.000     .8180236    1.258789
      _hatsq |    .001339   .0038978     0.34   0.731     -.006315     .008993
       _cons |   .2698725   .7976158     0.34   0.735    -1.296365     1.83611
------------------------------------------------------------------------------
```

Ramsey RESET test

```
Ramsey RESET test using powers of the fitted values of LnInvoicesPop
       Ho:  model has no omitted variables
                 F(3, 623) =      1.64
                  Prob > F =      0.1785
```

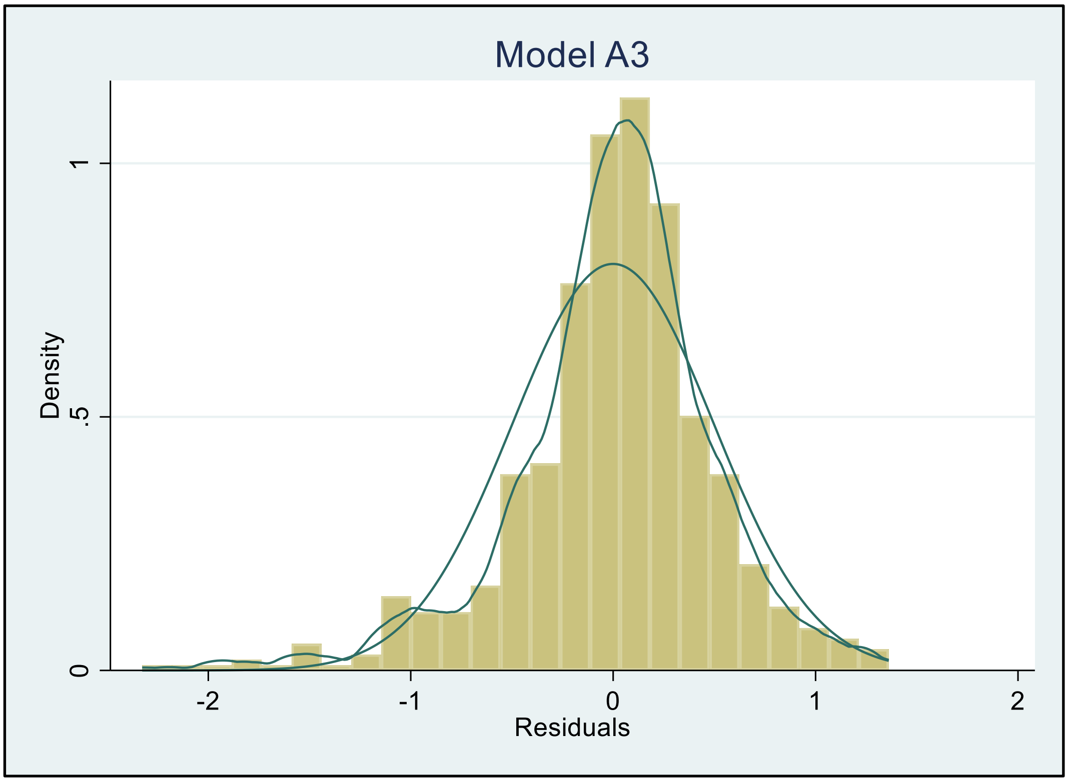
Serial correlation test

Source	SS	df	MS
Model	.138968391	1	.138968391
Residual	145.585577	590	.246755215
Total	145.724546	591	.246572835

Number of obs	=	592
F(1, 590)	=	0.56
Prob > F	=	0.4533
R-squared	=	0.0010
Adj R-squared	=	-0.0007
Root MSE	=	.49674

resid	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
resid						
L1.	-.0309829	.0412855	-0.75	0.453	-.1120673	.0501015
_cons	-.0021414	.0204165	-0.10	0.917	-.0422393	.0379565

Normality of the residuals



Mundlak test

```
Random-effects GLS regression                   Number of obs     =        648
Group variable: ID                              Number of groups  =         54

R-sq:                                           Obs per group:
     within  = 0.1372                                         min =          2
     between = 0.9878                                         avg =       12.0
     overall = 0.9379                                         max =         14

                                                Wald chi2(23)     =   33072.38
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

                                    (Std. Err. adjusted for 54 clusters in ID)
------------------------------------------------------------------------------
                  |               Robust
  LnInvoicesPop   |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
        LnWB_PRL  |   .7407621   .2699252     2.74   0.006     .2117185    1.269806
   LnWB_GDPpcgbp  |   1.369068   .2217878     6.17   0.000     .9343719    1.803764
      BIPM_CMCs   |  -.0205096   .0017628   -11.63   0.000    -.0239647   -.0170545
    LnCEPII_dist  |  -.4595426   .0170855   -26.90   0.000    -.4930296   -.4260556
 LnISO_Population |  -.2032983   .0236171    -8.61   0.000     -.249587   -.1570096
                  |
               U  |
             L1.  |   .3573989   .0479506     7.45   0.000     .2634175    .4513803
             L2.  |   .3311569   .0464437     7.13   0.000     .2401289    .4221849
             L3.  |   .2150397   .0348896     6.16   0.000     .1466574    .2834221
                  |
    LnWB_PRLmean  |   .4960139   .2961701     1.67   0.094    -.0844689    1.076497
LnWB_GDPpcgbpmean |  -.1394806    .211445    -0.66   0.509    -.5539051    .2749439
                  |
            YEAR  |
            2005  |   .1788467   .0887283     2.02   0.044     .0049423     .352751
            2006  |  -.0756216   .1177225    -0.64   0.521    -.3063534    .1551102
            2007  |  -.0610247   .1183016    -0.52   0.606    -.2928916    .1708422
            2008  |  -.3625741   .1128963    -3.21   0.001    -.5838467   -.1413015
            2009  |  -.5726524   .1534909    -3.73   0.000    -.8734891   -.2718157
            2010  |  -.4236139   .1710912    -2.48   0.013    -.7589464   -.0882813
            2011  |  -.6247922   .1788641    -3.49   0.000    -.9753594    -.274225
            2012  |  -.5944423   .1352508    -4.40   0.000     -.859529   -.3293555
            2013  |  -.4930519   .1771572    -2.78   0.005    -.8402735   -.1458302
            2014  |  -.4999567   .1473325    -3.39   0.001    -.7887232   -.2111902
            2015  |  -.5094565   .1776411    -2.87   0.004    -.8576266   -.1612864
            2016  |  -.8599917   .2316738    -3.71   0.000    -1.314064   -.4059194
            2017  |  -1.125954   .2506322    -4.49   0.000    -1.617184   -.6347235
                  |
           _cons  |  -17.51005   .8432178   -20.77   0.000    -19.16273   -15.85738
------------------+-----------------------------------------------------------
         sigma_u  |          0
         sigma_e  |  .45729096
             rho  |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
 ( 1)  LnWB_PRLmean = 0
 ( 2)  LnWB_GDPpcgbpmean = 0

           chi2(  2) =    3.97
         Prob > chi2 =    0.1374
```

Model B1

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -684.74634
Iteration 1:   log pseudolikelihood = -409.52578
Iteration 2:   log pseudolikelihood = -385.38575
Iteration 3:   log pseudolikelihood = -384.19913
Iteration 4:   log pseudolikelihood =  -384.1921
Iteration 5:   log pseudolikelihood =  -384.1921

Probit regression                               Number of obs     =      1,134
                                                Wald chi2(21)     =     265.44
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =  -384.1921               Pseudo R2         =     0.4389

------------------------------------------------------------------------------
             |               Robust
     TREATED |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    LnWB_PRL |   .9085269   .1943311     4.68   0.000      .527645    1.289409
LnWB_GDPpcgbp |   .9060121   .0904492    10.02   0.000     .7287351    1.083289
   BIPM_CMCs |    .018066   .0144459     1.25   0.211    -.0102474    .0463794
 LnCEPII_dist |  -.2644412   .0717865    -3.68   0.000    -.4051402   -.1237421
LnISO_Population |   .4579846    .055713     8.22   0.000      .348789    .5671801
             |
          pU |
         L1. |  -.4049774   .3369054    -1.20   0.229      -1.0653     .255345
         L2. |   -1.21919   .3341963    -3.65   0.000    -1.874203   -.5641774
         L3. |  -.1732434   .3183474    -0.54   0.586    -.7971929    .4507062
             |
        YEAR |
        2005 |  -.0255224   .3001165    -0.09   0.932    -.6137399     .562695
        2006 |  -.6309503   .2764285    -2.28   0.022     -1.17274   -.0891604
        2007 |  -.4416595   .3128439    -1.41   0.158    -1.054822    .1715033
        2008 |  -.4775795   .2849593    -1.68   0.094    -1.036089    .0809304
        2009 |   -.763232    .287055    -2.66   0.008     -1.32585   -.2006145
        2010 |  -.7262719   .2838726    -2.56   0.011    -1.282652   -.1698919
        2011 |   -1.15205   .2842461    -4.05   0.000    -1.709163   -.5949383
        2012 |  -1.020526   .2869221    -3.56   0.000    -1.582883   -.4581685
        2013 |  -.8162166   .2881362    -2.83   0.005    -1.380953   -.2514801
        2014 |  -.8320387   .2770438    -3.00   0.003    -1.375035   -.2890427
        2015 |  -.5971346   .2948917    -2.02   0.043    -1.175112   -.0191576
        2016 |  -1.029605   .2912724    -3.53   0.000    -1.600488   -.4587213
        2017 |  -1.252434   .3001771    -4.17   0.000     -1.84077   -.6640974
             |
       _cons |  -11.47245   1.332398    -8.61   0.000     -14.0839   -8.860996
------------------------------------------------------------------------------
```

Partial effects at the average

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	.180655	.042541	4.25	0.000	.0972762	.2640337
LnWB_GDPpcgbp	.1801549	.027308	6.60	0.000	.1266322	.2336776
BIPM_CMCs	.0035923	.002636	1.36	0.173	-.0015741	.0087587
LnCEPII_dist	-.0525825	.0147025	-3.58	0.000	-.0813988	-.0237661
LnISO_Population	.0910674	.0147548	6.17	0.000	.0621486	.1199862
pU						
L1.	-.0805273	.068542	-1.17	0.240	-.2148672	.0538127
L2.	-.2424284	.0685868	-3.53	0.000	-.3768561	-.1080008
L3.	-.0344484	.0634349	-0.54	0.587	-.1587784	.0898817
YEAR						
2005	-.0017894	.0210136	-0.09	0.932	-.0429753	.0393965
2006	-.0760453	.0351423	-2.16	0.030	-.144923	-.0071676
2007	-.0452918	.0360727	-1.26	0.209	-.115993	.0254093
2008	-.0505358	.0315134	-1.60	0.109	-.112301	.0112294
2009	-.102343	.043293	-2.36	0.018	-.1871957	-.0174904
2010	-.094581	.0410409	-2.30	0.021	-.1750196	-.0141423
2011	-.20384	.0583359	-3.49	0.000	-.3181763	-.0895037
2012	-.1654843	.0544368	-3.04	0.002	-.2721785	-.0587902
2013	-.1140402	.0462605	-2.47	0.014	-.2047092	-.0233713
2014	-.1176642	.0433788	-2.71	0.007	-.202685	-.0326434
2015	-.069973	.0403365	-1.73	0.083	-.1490312	.0090851
2016	-.1680028	.0560996	-2.99	0.003	-.2779559	-.0580497
2017	-.2357136	.0695933	-3.39	0.001	-.3721139	-.0993132

Average partial effects

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	.1734383	.0372017	4.66	0.000	.1005244	.2463523
LnWB_GDPpcgbp	.1729583	.0146125	11.84	0.000	.1443182	.2015983
BIPM_CMCs	.0034488	.0027168	1.27	0.204	-.001876	.0087736
LnCEPII_dist	-.050482	.0134677	-3.75	0.000	-.0768782	-.0240857
LnISO_Population	.0874295	.0099696	8.77	0.000	.0678894	.1069696
pU						
L1.	-.0773104	.0644451	-1.20	0.230	-.2036205	.0489996
L2.	-.2327441	.0624625	-3.73	0.000	-.3551685	-.1103198
L3.	-.0330723	.0607205	-0.54	0.586	-.1520822	.0859377
YEAR						
2005	-.0038493	.0452656	-0.09	0.932	-.0925682	.0848697
2006	-.1065187	.0452592	-2.35	0.019	-.1952252	-.0178123
2007	-.0723834	.0514069	-1.41	0.159	-.173139	.0283723
2008	-.0787425	.0460895	-1.71	0.088	-.1690764	.0115913
2009	-.1311629	.0475774	-2.76	0.006	-.2244129	-.037913
2010	-.1242222	.0467954	-2.65	0.008	-.2159394	-.032505
2011	-.2059109	.0472073	-4.36	0.000	-.2984355	-.1133863
2012	-.1803668	.0478305	-3.77	0.000	-.2741128	-.0866208
2013	-.1411784	.047801	-2.95	0.003	-.2348666	-.0474902
2014	-.144183	.0455096	-3.17	0.002	-.2333802	-.0549858
2015	-.1003141	.0489906	-2.05	0.041	-.196334	-.0042943
2016	-.1821243	.0482729	-3.77	0.000	-.2767375	-.087511
2017	-.2254859	.0499993	-4.51	0.000	-.3234828	-.127489

Estimation results

Linear regression						
			Number of obs	=	648	
			F(22, 625)	=	469.31	
			Prob > F	=	0.0000	
			R-squared	=	0.9380	
			Root MSE	=	.50351	

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.238868	.0765929	16.17	0.000	1.088457	1.389278
LnWB_GDPpcgbp	1.334916	.0755519	17.67	0.000	1.18655	1.483283
BIPM_CMCs	-.0203664	.00168	-12.12	0.000	-.0236655	-.0170673
LnCEPII_dist	-.475067	.0159356	-29.81	0.000	-.5063609	-.4437732
LnISO_Population	-.1712076	.028077	-6.10	0.000	-.2263442	-.116071
U						
L1.	.347696	.0455758	7.63	0.000	.2581957	.4371962
L2.	.3308802	.0429688	7.70	0.000	.2464996	.4152609
L3.	.216301	.0415533	5.21	0.000	.1347	.2979021
lambda	.283928	.142488	1.99	0.047	.0041147	.5637413
YEAR						
2005	.1819607	.0976722	1.86	0.063	-.0098448	.3737662
2006	-.0920482	.1124853	-0.82	0.413	-.312943	.1288466
2007	-.0619161	.1138929	-0.54	0.587	-.2855751	.1617429
2008	-.4259395	.1123042	-3.79	0.000	-.6464788	-.2054001
2009	-.6681301	.1053219	-6.34	0.000	-.8749577	-.4613025
2010	-.5348193	.104634	-5.11	0.000	-.7402961	-.3293426
2011	-.7635247	.1197042	-6.38	0.000	-.9985959	-.5284536
2012	-.7195368	.105622	-6.81	0.000	-.9269538	-.5121199
2013	-.6083887	.1154341	-5.27	0.000	-.8350744	-.3817029
2014	-.5725413	.1015011	-5.64	0.000	-.7718658	-.3732168
2015	-.5334067	.0992389	-5.37	0.000	-.7282888	-.3385247
2016	-.9362018	.1108656	-8.44	0.000	-1.153916	-.7184875
2017	-1.238251	.1201225	-10.31	0.000	-1.474143	-1.002358
_cons	-18.94411	1.10697	-17.11	0.000	-21.11794	-16.77027

Wald test year dummies

```
( 1)  2005.YEAR = 0
( 2)  2006.YEAR = 0
( 3)  2007.YEAR = 0
( 4)  2008.YEAR = 0
( 5)  2009.YEAR = 0
( 6)  2010.YEAR = 0
( 7)  2011.YEAR = 0
( 8)  2012.YEAR = 0
( 9)  2013.YEAR = 0
(10)  2014.YEAR = 0
(11)  2015.YEAR = 0
(12)  2016.YEAR = 0
(13)  2017.YEAR = 0

      F( 13,   625) =   17.87
           Prob > F =    0.0000
```

Link test

```
      Source |       SS           df       MS      Number of obs   =       648
-------------+----------------------------------   F(2, 645)       =   5235.06
       Model |  2545.03128         2  1272.51564   Prob > F        =    0.0000
    Residual |  156.783817       645  .243075685   R-squared       =    0.9420
-------------+----------------------------------   Adj R-squared   =    0.9418
       Total |   2701.8151       647  4.17591206   Root MSE        =    .49303

------------------------------------------------------------------------------
LnInvoices~p |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        _hat |   1.032845   .1115759     9.26   0.000     .8137496    1.251941
      _hatsq |   .0011452   .0038752     0.30   0.768    -.0064644    .0087548
       _cons |   .2307861    .792917     0.29   0.771    -1.326224    1.787796
------------------------------------------------------------------------------
```

Ramsey RESET test

```
Ramsey RESET test using powers of the fitted values of LnInvoicesPop
       Ho:  model has no omitted variables
                  F(3, 622) =      1.52
                   Prob > F =      0.2086
```

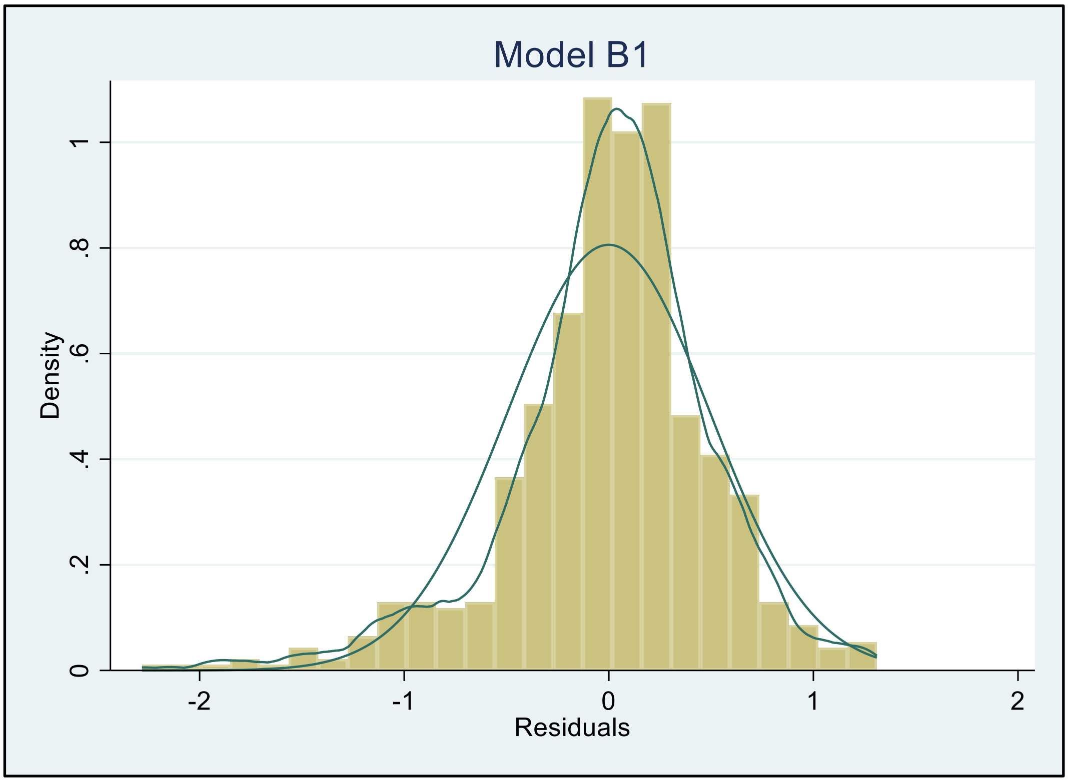
Serial correlation test

Source	SS	df	MS
Model	.144114257	1	.144114257
Residual	144.834366	590	.245481976
Total	144.97848	591	.245310457

Number of obs	=	592
F(1, 590)	=	0.59
Prob > F	=	0.4439
R-squared	=	0.0010
Adj R-squared	=	-0.0007
Root MSE	=	.49546

resid	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
resid						
L1.	-.0317344	.0414177	-0.77	0.444	-.1130785	.0496098
_cons	-.0015618	.0203639	-0.08	0.939	-.0415563	.0384327

Normality of the residuals



Mundlak test

```
Random-effects GLS regression                   Number of obs     =        648
Group variable: ID                              Number of groups  =         54

R-sq:                                           Obs per group:
     within  = 0.1359                                         min =          2
     between = 0.9906                                         avg =       12.0
     overall = 0.9386                                         max =         14

                                                Wald chi2(24)     =   21357.30
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

                                    (Std. Err. adjusted for 54 clusters in ID)
------------------------------------------------------------------------------
             |               Robust
LnInvoicesPop |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------------------------------------------------------------------
     LnWB_PRL |   .7651109   .2646211     2.89   0.004      .246463    1.283759
 LnWB_GDPpcgbp |   1.470836   .2340103     6.29   0.000     1.012184    1.929487
     BIPM_CMCs |  -.0206703   .0016967   -12.18   0.000    -.0239959   -.0173448
  LnCEPII_dist |  -.4672759   .0169279   -27.60   0.000     -.500454   -.4340978
LnISO_Population |  -.1704706   .0240287    -7.09   0.000     -.217566   -.1233752
             |
           U |
         L1. |   .3439932   .0468007     7.35   0.000     .2522655    .4357209
         L2. |    .326922   .0458537     7.13   0.000     .2370503    .4167936
         L3. |    .214052   .0345459     6.20   0.000     .1463433    .2817607
             |
      lambda |   .2978994   .1405654     2.12   0.034     .0223963    .5734025
  LnWB_PRLmean |   .5237407   .2848934     1.84   0.066    -.0346401    1.082121
LnWB_GDPpcgbpmean |  -.1573595   .2115587    -0.74   0.457    -.5720068    .2572879
             |
        YEAR |
        2005 |   .1792693   .0880072     2.04   0.042     .0067784    .3517601
        2006 |  -.1031179   .1179523    -0.87   0.382    -.3343001    .1280643
        2007 |  -.0786244   .1198505    -0.66   0.512     -.313527    .1562782
        2008 |  -.3894235    .111963    -3.48   0.001    -.6088668   -.1699801
        2009 |  -.6144575   .1527285    -4.02   0.000       -.9138   -.3151151
        2010 |   -.462623   .1671302    -2.77   0.006    -.7901922   -.1350538
        2011 |  -.6856868   .1777711    -3.86   0.000    -1.034112   -.3372619
        2012 |  -.6562717    .137658    -4.77   0.000    -.9260765   -.3864669
        2013 |  -.5493056    .180723    -3.04   0.002    -.9035162    -.195095
        2014 |  -.5485883   .1480064    -3.71   0.000    -.8386756    -.258501
        2015 |  -.5528964   .1808898    -3.06   0.002    -.9074338    -.198359
        2016 |  -.9383959    .235231    -3.99   0.000     -1.39944   -.4773517
        2017 |  -1.218002   .2581728    -4.72   0.000    -1.724011   -.7119926
             |
       _cons |  -18.80047   .8876373   -21.18   0.000    -20.54021   -17.06073
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  .45502543
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
 ( 1)  LnWB_PRLmean = 0
 ( 2)  LnWB_GDPpcgbpmean = 0

           chi2(  2) =    4.32
         Prob > chi2 =    0.1151
```

Model C1

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -684.74634
Iteration 1:   log pseudolikelihood = -409.52578
Iteration 2:   log pseudolikelihood = -385.38575
Iteration 3:   log pseudolikelihood = -384.19913
Iteration 4:   log pseudolikelihood =  -384.1921
Iteration 5:   log pseudolikelihood =  -384.1921

Probit regression                               Number of obs     =      1,134
                                                Wald chi2(21)     =     265.44
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =  -384.1921               Pseudo R2         =     0.4389
```

TREATED	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	.9085269	.1943311	4.68	0.000	.527645	1.289409
LnWB_GDPpcgbp	.9060121	.0904492	10.02	0.000	.7287351	1.083289
BIPM_CMCs	.018066	.0144459	1.25	0.211	-.0102474	.0463794
LnCEPII_dist	-.2644412	.0717865	-3.68	0.000	-.4051402	-.1237421
LnISO_Population	.4579846	.055713	8.22	0.000	.348789	.5671801
pU						
L1.	-.4049774	.3369054	-1.20	0.229	-1.0653	.255345
L2.	-1.21919	.3341963	-3.65	0.000	-1.874203	-.5641774
L3.	-.1732434	.3183474	-0.54	0.586	-.7971929	.4507062
YEAR						
2005	-.0255224	.3001165	-0.09	0.932	-.6137399	.562695
2006	-.6309503	.2764285	-2.28	0.022	-1.17274	-.0891604
2007	-.4416595	.3128439	-1.41	0.158	-1.054822	.1715033
2008	-.4775795	.2849593	-1.68	0.094	-1.036089	.0809304
2009	-.763232	.287055	-2.66	0.008	-1.32585	-.2006145
2010	-.7262719	.2838726	-2.56	0.011	-1.282652	-.1698919
2011	-1.15205	.2842461	-4.05	0.000	-1.709163	-.5949383
2012	-1.020526	.2869221	-3.56	0.000	-1.582883	-.4581685
2013	-.8162166	.2881362	-2.83	0.005	-1.380953	-.2514801
2014	-.8320387	.2770438	-3.00	0.003	-1.375035	-.2890427
2015	-.5971346	.2948917	-2.02	0.043	-1.175112	-.0191576
2016	-1.029605	.2912724	-3.53	0.000	-1.600488	-.4587213
2017	-1.252434	.3001771	-4.17	0.000	-1.84077	-.6640974
_cons	-11.47245	1.332398	-8.61	0.000	-14.0839	-8.860996

Estimation results

Linear regression		
Number of obs	=	648
F(22, 625)	=	469.31
Prob > F	=	0.0000
R-squared	=	0.9380
Root MSE	=	.50351

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.238868	.0765929	16.17	0.000	1.088457	1.389278
LnWB_GDPpcgbp	1.334916	.0755519	17.67	0.000	1.18655	1.483283
BIPM_CMCs	-.0203664	.00168	-12.12	0.000	-.0236655	-.0170673
LnCEPII_dist	-.475067	.0159356	-29.81	0.000	-.5063609	-.4437732
LnISO_Population	-.1712076	.028077	-6.10	0.000	-.2263442	-.116071
U						
L1.	.347696	.0455758	7.63	0.000	.2581957	.4371962
L2.	.3308802	.0429688	7.70	0.000	.2464996	.4152609
L3.	.216301	.0415533	5.21	0.000	.1347	.2979021
lambda	.283928	.142488	1.99	0.047	.0041147	.5637413
YEAR						
2005	.1819607	.0976722	1.86	0.063	-.0098448	.3737662
2006	-.0920482	.1124853	-0.82	0.413	-.312943	.1288466
2007	-.0619161	.1138929	-0.54	0.587	-.2855751	.1617429
2008	-.4259395	.1123042	-3.79	0.000	-.6464788	-.2054001
2009	-.6681301	.1053219	-6.34	0.000	-.8749577	-.4613025
2010	-.5348193	.104634	-5.11	0.000	-.7402961	-.3293426
2011	-.7635247	.1197042	-6.38	0.000	-.9985959	-.5284536
2012	-.7195368	.105622	-6.81	0.000	-.9269538	-.5121199
2013	-.6083887	.1154341	-5.27	0.000	-.8350744	-.3817029
2014	-.5725413	.1015011	-5.64	0.000	-.7718658	-.3732168
2015	-.5334067	.0992389	-5.37	0.000	-.7282888	-.3385247
2016	-.9362018	.1108656	-8.44	0.000	-1.153916	-.7184875
2017	-1.238251	.1201225	-10.31	0.000	-1.474143	-1.002358
_cons	-18.94411	1.10697	-17.11	0.000	-21.11794	-16.77027

Wald test year dummy variables

```
( 1)  2005.YEAR = 0
( 2)  2006.YEAR = 0
( 3)  2007.YEAR = 0
( 4)  2008.YEAR = 0
( 5)  2009.YEAR = 0
( 6)  2010.YEAR = 0
( 7)  2011.YEAR = 0
( 8)  2012.YEAR = 0
( 9)  2013.YEAR = 0
(10)  2014.YEAR = 0
(11)  2015.YEAR = 0
(12)  2016.YEAR = 0
(13)  2017.YEAR = 0

      F( 13,   625) =   17.87
           Prob > F =    0.0000
```

Model C2

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -598.32459
Iteration 1:   log pseudolikelihood = -368.71544
Iteration 2:   log pseudolikelihood = -346.48657
Iteration 3:   log pseudolikelihood = -345.30723
Iteration 4:   log pseudolikelihood = -345.30402
Iteration 5:   log pseudolikelihood = -345.30402

Probit regression                               Number of obs     =      1,064
                                                Wald chi2(21)     =     231.42
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -345.30402               Pseudo R2         =     0.4229
```

TREATED	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	.9734544	.2050084	4.75	0.000	.5716453	1.375264
LnWB_GDPpcgbp	.8021639	.0928346	8.64	0.000	.6202116	.9841163
BIPM_CMCs	.0120047	.0130965	0.92	0.359	-.0136639	.0376734
LnCEPII_dist	-.2177604	.0776437	-2.80	0.005	-.3699392	-.0655816
LnISO_Population	.4357047	.0577139	7.55	0.000	.3225875	.5488219
pU						
L1.	-.7017946	.3258374	-2.15	0.031	-1.340424	-.063165
L2.	-1.348467	.3217812	-4.19	0.000	-1.979146	-.717787
L3.	-.4648687	.311382	-1.49	0.135	-1.075166	.1454288
YEAR						
2005	-.0500267	.3111108	-0.16	0.872	-.6597927	.5597394
2006	-.6212117	.2843	-2.19	0.029	-1.178429	-.063994
2007	-.4435881	.3237331	-1.37	0.171	-1.078093	.1909171
2008	-.4818523	.2934968	-1.64	0.101	-1.057096	.093391
2009	-.7017645	.3008086	-2.33	0.020	-1.291338	-.1121906
2010	-.6423429	.2981859	-2.15	0.031	-1.226777	-.0579092
2011	-1.120724	.2956011	-3.79	0.000	-1.700092	-.5413568
2012	-1.01464	.296684	-3.42	0.001	-1.59613	-.4331496
2013	-.7790938	.3015406	-2.58	0.010	-1.370102	-.1880852
2014	-.7301093	.2932763	-2.49	0.013	-1.30492	-.1552984
2015	-.4385673	.3149661	-1.39	0.164	-1.05589	.178755
2016	-.8790883	.3077962	-2.86	0.004	-1.482358	-.2758188
2017	-1.201648	.3153783	-3.81	0.000	-1.819778	-.5835179
_cons	-10.29496	1.354829	-7.60	0.000	-12.95037	-7.639541

Estimation results

```
Linear regression                               Number of obs     =        648
                                                F(22, 625)        =     468.43
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9380
                                                Root MSE          =     .50334

------------------------------------------------------------------------------
             |               Robust
LnInvoicesPop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     LnWB_PRL |   1.232712   .0767565    16.06   0.000     1.081981    1.383444
 LnWB_GDPpcgbp |   1.348551   .0741122    18.20   0.000     1.203012    1.494091
    BIPM_CMCs |  -.0205216   .0016853   -12.18   0.000    -.0238311   -.0172121
  LnCEPII_dist |  -.4729609   .0159258   -29.70   0.000    -.5042355   -.4416864
LnISO_Population |  -.1684318   .0278516    -6.05   0.000    -.2231259   -.1137378
             |
           U |
         L1. |   .3478359   .0455298     7.64   0.000      .258426    .4372458
         L2. |   .3307852   .0430039     7.69   0.000     .2463355    .4152349
         L3. |   .2159888   .0415521     5.20   0.000     .1343902    .2975874
             |
      lambda |   .3017339   .1476431     2.04   0.041     .0117973    .5916705
             |
        YEAR |
        2005 |   .1843174   .0977495     1.89   0.060    -.0076398    .3762745
        2006 |  -.0934201   .1124526    -0.83   0.406    -.3142507    .1274106
        2007 |  -.0644678   .1138542    -0.57   0.571    -.2880509    .1591152
        2008 |  -.4281073   .1122763    -3.81   0.000    -.6485919   -.2076228
        2009 |  -.6767303   .1053928    -6.42   0.000    -.8836972   -.4697633
        2010 |  -.5429053   .1048186    -5.18   0.000    -.7487446   -.3370659
        2011 |  -.7687787   .1196001    -6.43   0.000    -1.003645    -.533912
        2012 |  -.7234247   .1053653    -6.87   0.000    -.9303375   -.5165119
        2013 |  -.6128461   .1151118    -5.32   0.000    -.8388988   -.3867934
        2014 |  -.5713248   .1011587    -5.65   0.000    -.7699768   -.3726727
        2015 |  -.5324065   .0989629    -5.38   0.000    -.7267465   -.3380664
        2016 |  -.9348863   .1099122    -8.51   0.000    -1.150728   -.7190444
        2017 |  -1.244838   .1192749   -10.44   0.000    -1.479066    -1.01061
             |
       _cons |  -19.13626   1.090025   -17.56   0.000    -21.27682   -16.99571
------------------------------------------------------------------------------
```

Wald test year dummy variables

```
 ( 1)  2005.YEAR = 0
 ( 2)  2006.YEAR = 0
 ( 3)  2007.YEAR = 0
 ( 4)  2008.YEAR = 0
 ( 5)  2009.YEAR = 0
 ( 6)  2010.YEAR = 0
 ( 7)  2011.YEAR = 0
 ( 8)  2012.YEAR = 0
 ( 9)  2013.YEAR = 0
 (10)  2014.YEAR = 0
 (11)  2015.YEAR = 0
 (12)  2016.YEAR = 0
 (13)  2017.YEAR = 0

       F( 13,   625) =   18.50
            Prob > F =    0.0000
```

Model C3

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -446.69777
Iteration 1:   log pseudolikelihood = -280.56498
Iteration 2:   log pseudolikelihood = -259.66367
Iteration 3:   log pseudolikelihood = -258.30506
Iteration 4:   log pseudolikelihood = -258.30098
Iteration 5:   log pseudolikelihood = -258.30098

Probit regression                               Number of obs     =        952
                                                Wald chi2(21)     =     200.27
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -258.30098               Pseudo R2         =     0.4218
```

TREATED	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	.6545887	.2252363	2.91	0.004	.2131337	1.096044
LnWB_GDPpcgbp	.7577269	.1062245	7.13	0.000	.5495307	.965923
BIPM_CMCs	.0153726	.0160751	0.96	0.339	-.016134	.0468792
LnCEPII_dist	-.2355645	.0892665	-2.64	0.008	-.4105237	-.0606053
LnISO_Population	.3702859	.0681741	5.43	0.000	.236667	.5039047
pU						
L1.	-1.12034	.3190567	-3.51	0.000	-1.74568	-.4950009
L2.	-1.458272	.3332166	-4.38	0.000	-2.111364	-.8051793
L3.	-.4344022	.3438771	-1.26	0.206	-1.108389	.2395846
YEAR						
2005	.0446718	.327767	0.14	0.892	-.5977397	.6870834
2006	-.5613187	.3043627	-1.84	0.065	-1.157859	.0352212
2007	-.3948577	.352433	-1.12	0.263	-1.085614	.2958983
2008	-.3227619	.3139229	-1.03	0.304	-.9380394	.2925156
2009	-.4468967	.3298895	-1.35	0.176	-1.093468	.1996748
2010	-.3969867	.3124455	-1.27	0.204	-1.009369	.2153953
2011	-.8465934	.319074	-2.65	0.008	-1.471967	-.2212198
2012	-.8312878	.3092257	-2.69	0.007	-1.437359	-.2252166
2013	-.2810815	.3311853	-0.85	0.396	-.9301929	.3680298
2014	-.4141301	.3335603	-1.24	0.214	-1.067896	.239636
2015	-.0178506	.37875	-0.05	0.962	-.7601869	.7244857
2016	-.4326574	.348879	-1.24	0.215	-1.116448	.2511328
2017	-.8441015	.3634953	-2.32	0.020	-1.556539	-.1316638
_cons	-8.926634	1.630389	-5.48	0.000	-12.12214	-5.73113

Estimation results

```
Linear regression                               Number of obs     =        648
                                                F(22, 625)        =     469.12
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9379
                                                Root MSE          =     .50385

------------------------------------------------------------------------------
             |               Robust
LnInvoicesPop|      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     LnWB_PRL|    1.17244   .0752526    15.58   0.000     1.024662    1.320219
 LnWB_GDPpcgbp|   1.401174   .0729221    19.21   0.000     1.257972    1.544376
    BIPM_CMCs|  -.0212586   .0016789   -12.66   0.000    -.0245555   -.0179617
 LnCEPII_dist|   -.476195    .015851   -30.04   0.000    -.5073227   -.4450673
LnISO_Population| -.1552831  .0271826   -5.71   0.000    -.2086633   -.1019028
             |
            U |
          L1. |   .3517025   .0456046     7.71   0.000     .2621457    .4412593
          L2. |   .3321039   .0428595     7.75   0.000     .2479378    .4162701
          L3. |   .2158438   .0416761     5.18   0.000     .1340016     .297686
             |
       lambda |   .3191935   .1834423     1.74   0.082    -.0410444    .6794314
             |
         YEAR |
         2005 |   .1807898   .0976225     1.85   0.065    -.0109181    .3724976
         2006 |  -.1118126   .1127697    -0.99   0.322     -.333266    .1096408
         2007 |  -.0866946   .1146688    -0.76   0.450    -.3118774    .1384882
         2008 |   -.440773   .1125196    -3.92   0.000    -.6617353   -.2198108
         2009 |  -.6725818   .1053786    -6.38   0.000    -.8795208   -.4656428
         2010 |  -.5416522   .1050011    -5.16   0.000      -.74785   -.3354544
         2011 |  -.7671354   .1197138    -6.41   0.000    -1.002225   -.5320454
         2012 |  -.7338291   .1056732    -6.94   0.000    -.9413466   -.5263116
         2013 |  -.6162592   .1151052    -5.35   0.000     -.842299   -.3902195
         2014 |  -.5827257   .1011512    -5.76   0.000    -.7813631   -.3840883
         2015 |   -.543442   .0990853    -5.48   0.000    -.7380223   -.3488617
         2016 |   -.939383   .1087017    -8.64   0.000    -1.152848   -.7259182
         2017 |  -1.255023   .1175631   -10.68   0.000     -1.48589   -1.024157
             |
        _cons |  -19.84282   1.066743   -18.60   0.000    -21.93766   -17.74799
------------------------------------------------------------------------------
```

Wald test year dummy variables

```
 ( 1)  2005.YEAR = 0
 ( 2)  2006.YEAR = 0
 ( 3)  2007.YEAR = 0
 ( 4)  2008.YEAR = 0
 ( 5)  2009.YEAR = 0
 ( 6)  2010.YEAR = 0
 ( 7)  2011.YEAR = 0
 ( 8)  2012.YEAR = 0
 ( 9)  2013.YEAR = 0
 (10)  2014.YEAR = 0
 (11)  2015.YEAR = 0
 (12)  2016.YEAR = 0
 (13)  2017.YEAR = 0

       F( 13,   625) =   19.23
            Prob > F =    0.0000
```

Model C4

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -362.04766
Iteration 1:   log pseudolikelihood = -235.56083
Iteration 2:   log pseudolikelihood =  -219.5452
Iteration 3:   log pseudolikelihood = -218.75292
Iteration 4:   log pseudolikelihood = -218.74944
Iteration 5:   log pseudolikelihood = -218.74944

Probit regression                               Number of obs     =        896
                                                Wald chi2(21)     =     176.91
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -218.74944               Pseudo R2         =     0.3958
```

TREATED	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	.4638484	.2246473	2.06	0.039	.0235477	.9041491
LnWB_GDPpcgbp	.8427734	.1551164	5.43	0.000	.5387508	1.146796
BIPM_CMCs	.0103968	.01431	0.73	0.468	-.0176504	.0384439
LnCEPII_dist	-.174327	.0902575	-1.93	0.053	-.3512283	.0025744
LnISO_Population	.3703882	.0713508	5.19	0.000	.2305432	.5102332
pU						
L1.	-.8866492	.3442407	-2.58	0.010	-1.561349	-.2119499
L2.	-1.525047	.3342729	-4.56	0.000	-2.18021	-.869884
L3.	-.6021976	.3502288	-1.72	0.086	-1.288633	.0842382
YEAR						
2005	.2411578	.3490726	0.69	0.490	-.443012	.9253276
2006	-.4560896	.3102442	-1.47	0.142	-1.064157	.1519778
2007	-.3781685	.3561479	-1.06	0.288	-1.076206	.3198686
2008	-.1608407	.3318999	-0.48	0.628	-.8113525	.4896711
2009	-.2831282	.3472252	-0.82	0.415	-.9636772	.3974208
2010	-.1517936	.3270087	-0.46	0.643	-.7927188	.4891317
2011	-.7232792	.3337917	-2.17	0.030	-1.377499	-.0690594
2012	-.7012854	.3199339	-2.19	0.028	-1.328344	-.0742265
2013	-.2020155	.3450203	-0.59	0.558	-.8782428	.4742118
2014	-.4213002	.3452476	-1.22	0.222	-1.097973	.2553727
2015	.0268572	.4141924	0.06	0.948	-.7849449	.8386593
2016	-.2908256	.3847097	-0.76	0.450	-1.044843	.4631915
2017	-.7521861	.3753773	-2.00	0.045	-1.487912	-.0164601
_cons	-10.36444	2.116554	-4.90	0.000	-14.51281	-6.216074

Estimation results

```
(821 missing values generated)

Linear regression                               Number of obs     =        648
                                                F(22, 625)        =     467.68
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9378
                                                Root MSE          =     .50431

------------------------------------------------------------------------------
                 |               Robust
   LnInvoicesPop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
       LnWB_PRL  |   1.123656   .0740769    15.17   0.000     .9781867    1.269126
  LnWB_GDPpcgbp  |   1.460542   .0747442    19.54   0.000     1.313761    1.607322
      BIPM_CMCs  |  -.0218875   .0016923   -12.93   0.000    -.0252107   -.0185642
   LnCEPII_dist  |  -.4746845    .015869   -29.91   0.000    -.5058476   -.4435214
LnISO_Population |  -.1437861   .0275922    -5.21   0.000    -.1979708   -.0896014
                 |
               U |
             L1. |   .3539822   .0456251     7.76   0.000     .2643852    .4435793
             L2. |    .332643   .0428589     7.76   0.000     .2484781     .416808
             L3. |    .215367   .0416386     5.17   0.000     .1335985    .2971354
                 |
          lambda |   .3137628   .2129724     1.47   0.141    -.1044652    .7319909
                 |
            YEAR |
           2005  |   .1615747   .0974753     1.66   0.098     -.029844    .3529934
           2006  |  -.1351552   .1128261    -1.20   0.231    -.3567194     .086409
           2007  |  -.1056242   .1151345    -0.92   0.359    -.3317215     .120473
           2008  |  -.4556213   .1121128    -4.06   0.000    -.6757846   -.2354579
           2009  |  -.6939795   .1049568    -6.61   0.000    -.9000901   -.4878689
           2010  |  -.5568172   .1052489    -5.29   0.000    -.7635016   -.3501329
           2011  |  -.7761306   .1194824    -6.50   0.000    -1.010766    -.541495
           2012  |  -.7426582    .105482    -7.04   0.000    -.9498002   -.5355161
           2013  |  -.6329711   .1149766    -5.51   0.000    -.8587583   -.4071839
           2014  |  -.6039764    .101283    -5.96   0.000    -.8028726   -.4050801
           2015  |  -.5662532    .099032    -5.72   0.000    -.7607289   -.3717775
           2016  |   -.963697   .1085828    -8.88   0.000    -1.176928   -.7504657
           2017  |  -1.278091   .1176928   -10.86   0.000    -1.509212    -1.04697
                 |
           _cons |  -20.61736   1.095647   -18.82   0.000    -22.76896   -18.46577
------------------------------------------------------------------------------
```

Wald test year dummy variables

```
 ( 1)  2005.YEAR = 0
 ( 2)  2006.YEAR = 0
 ( 3)  2007.YEAR = 0
 ( 4)  2008.YEAR = 0
 ( 5)  2009.YEAR = 0
 ( 6)  2010.YEAR = 0
 ( 7)  2011.YEAR = 0
 ( 8)  2012.YEAR = 0
 ( 9)  2013.YEAR = 0
 (10)  2014.YEAR = 0
 (11)  2015.YEAR = 0
 (12)  2016.YEAR = 0
 (13)  2017.YEAR = 0

       F( 13,   625) =   19.16
            Prob > F =   0.0000
```

Model C5

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -316.17231
Iteration 1:   log pseudolikelihood = -205.39911
Iteration 2:   log pseudolikelihood =  -189.9634
Iteration 3:   log pseudolikelihood =   -189.221
Iteration 4:   log pseudolikelihood = -189.21806
Iteration 5:   log pseudolikelihood = -189.21806

Probit regression                               Number of obs     =        868
                                                Wald chi2(21)     =     157.28
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -189.21806               Pseudo R2         =     0.4015

------------------------------------------------------------------------------------
                  |               Robust
          TREATED |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
        LnWB_PRL  |    .650153   .2414449     2.69   0.007     .1769296    1.123376
    LnWB_GDPpcgbp |   .8204127   .1603762     5.12   0.000     .5060812    1.134744
       BIPM_CMCs  |   .0071879   .0146195     0.49   0.623    -.0214658    .0358416
    LnCEPII_dist  |  -.1469719   .0978738    -1.50   0.133     -.338801    .0448571
LnISO_Population  |   .3869709   .0748883     5.17   0.000     .2401925    .5337493
                  |
               pU |
              L1. |  -.8645425   .3487202    -2.48   0.013    -1.548022   -.1810635
              L2. |  -1.555782   .3432901    -4.53   0.000    -2.228619   -.8829463
              L3. |   -.487099   .3621311    -1.35   0.179    -1.196863    .2226648
                  |
             YEAR |
             2005 |   .2415744   .3732191     0.65   0.517    -.4899216    .9730703
             2006 |  -.5629215   .3278195    -1.72   0.086    -1.205436    .0795928
             2007 |  -.4749399   .3797846    -1.25   0.211    -1.219304    .2694242
             2008 |   -.333615   .3436748    -0.97   0.332    -1.007205    .3399753
             2009 |  -.3379751   .3723018    -0.91   0.364    -1.067673     .391723
             2010 |  -.3297286   .3425226    -0.96   0.336    -1.001061    .3416034
             2011 |  -.8524577   .3544506    -2.41   0.016    -1.547168   -.1577474
             2012 |  -.8137174   .3347366    -2.43   0.015    -1.469789   -.1576458
             2013 |  -.2064948   .3694574    -0.56   0.576    -.9306181    .5176285
             2014 |   -.743584   .3501453    -2.12   0.034    -1.429856   -.0573118
             2015 |  -.0857151   .4540086    -0.19   0.850    -.9755557    .8041254
             2016 |  -.1994733   .4314772    -0.46   0.644    -1.045153    .6462064
             2017 |  -.9709024   .3884067    -2.50   0.012    -1.732165   -.2096393
                  |
            _cons |  -10.31847   2.226905    -4.63   0.000    -14.68313   -5.953818
------------------------------------------------------------------------------------
```

Estimation results

Linear regression

Number of obs	=	648
F(22, 625)	=	467.07
Prob > F	=	0.0000
R-squared	=	0.9379
Root MSE	=	.50389

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.134971	.0749253	15.15	0.000	.9878356	1.282107
LnWB_GDPpcgbp	1.455225	.0743798	19.56	0.000	1.30916	1.601289
BIPM_CMCs	-.0219057	.0016965	-12.91	0.000	-.0252372	-.0185742
LnCEPII_dist	-.4738119	.0159042	-29.79	0.000	-.5050441	-.4425797
LnISO_Population	-.144529	.0276398	-5.23	0.000	-.1988071	-.0902508
U						
L1.	.3534994	.0455302	7.76	0.000	.2640887	.4429101
L2.	.3317408	.0428759	7.74	0.000	.2475425	.4159392
L3.	.2149031	.0416885	5.15	0.000	.1330366	.2967696
lambda	.434563	.2553071	1.70	0.089	-.0668005	.9359265
YEAR						
2005	.1634459	.0977523	1.67	0.095	-.0285168	.3554085
2006	-.1429094	.1131174	-1.26	0.207	-.3650455	.0792268
2007	-.1124753	.1154409	-0.97	0.330	-.3391742	.1142237
2008	-.4620073	.1124252	-4.11	0.000	-.6827841	-.2412304
2009	-.6893391	.1049831	-6.57	0.000	-.8955015	-.4831766
2010	-.5554881	.1051958	-5.28	0.000	-.7620682	-.348908
2011	-.7736897	.1195253	-6.47	0.000	-1.00841	-.5389699
2012	-.7427924	.1056248	-7.03	0.000	-.9502148	-.53537
2013	-.6264322	.1149699	-5.45	0.000	-.8522063	-.4006582
2014	-.6152643	.1019575	-6.03	0.000	-.8154851	-.4150434
2015	-.5547638	.099307	-5.59	0.000	-.7497796	-.359748
2016	-.9444458	.1088167	-8.68	0.000	-1.158136	-.7307553
2017	-1.273729	.118714	-10.73	0.000	-1.506855	-1.040602
_cons	-20.55722	1.093192	-18.80	0.000	-22.704	-18.41045

Wald test year dummy variables

```
( 1)  2005.YEAR = 0
( 2)  2006.YEAR = 0
( 3)  2007.YEAR = 0
( 4)  2008.YEAR = 0
( 5)  2009.YEAR = 0
( 6)  2010.YEAR = 0
( 7)  2011.YEAR = 0
( 8)  2012.YEAR = 0
( 9)  2013.YEAR = 0
(10)  2014.YEAR = 0
(11)  2015.YEAR = 0
(12)  2016.YEAR = 0
(13)  2017.YEAR = 0

      F( 13,   625) =   18.49
           Prob > F =   0.0000
```

Model D1

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -679.87823
Iteration 1:   log pseudolikelihood = -408.18716
Iteration 2:   log pseudolikelihood = -385.38351
Iteration 3:   log pseudolikelihood = -384.19386
Iteration 4:   log pseudolikelihood = -384.18698
Iteration 5:   log pseudolikelihood = -384.18698

Probit regression                               Number of obs     =      1,120
                                                Wald chi2(21)     =     264.92
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -384.18698               Pseudo R2         =     0.4349
```

TREATED	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	.9082056	.1945689	4.67	0.000	.5268576	1.289554
LnWB_GDPpcgbp	.9059529	.0904852	10.01	0.000	.7286052	1.083301
BIPM_CMCs	.0180796	.0144627	1.25	0.211	-.0102668	.046426
LnCEPII_dist	-.2642824	.0719508	-3.67	0.000	-.4053035	-.1232614
LnISO_Population	.4578624	.0558522	8.20	0.000	.3483941	.5673307
pU						
L1.	-.4051436	.336928	-1.20	0.229	-1.06551	.2552232
L2.	-1.219299	.3342141	-3.65	0.000	-1.874347	-.5642515
L3.	-.1732405	.3183714	-0.54	0.586	-.7972369	.4507559
YEAR						
2005	-.0254701	.3001058	-0.08	0.932	-.6136667	.5627264
2006	-.6308604	.2764176	-2.28	0.022	-1.172629	-.0890919
2007	-.4415866	.3128331	-1.41	0.158	-1.054728	.171555
2008	-.4774517	.2849545	-1.68	0.094	-1.035952	.0810489
2009	-.7630788	.2870625	-2.66	0.008	-1.325711	-.2004466
2010	-.7260875	.283887	-2.56	0.011	-1.282496	-.1696792
2011	-1.151889	.284269	-4.05	0.000	-1.709046	-.5947319
2012	-1.020341	.2869505	-3.56	0.000	-1.582754	-.4579287
2013	-.8160219	.2881686	-2.83	0.005	-1.380822	-.2512219
2014	-.8318718	.2770649	-3.00	0.003	-1.374909	-.2888345
2015	-.5969697	.294884	-2.02	0.043	-1.174932	-.0190077
2016	-1.029416	.2913139	-3.53	0.000	-1.600381	-.4584511
2017	-1.252228	.3002398	-4.17	0.000	-1.840687	-.6637689
_cons	-11.47172	1.332712	-8.61	0.000	-14.08378	-8.859647

Estimation results

```
Linear regression                               Number of obs     =        634
                                                F(22, 611)        =     388.53
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9322
                                                Root MSE          =     .50691

------------------------------------------------------------------------------
                 |               Robust
   LnInvoicesPop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
        LnWB_PRL |   1.147715   .0768503    14.93   0.000     .9967926    1.298638
   LnWB_GDPpcgbp |   1.226853   .0774174    15.85   0.000     1.074816    1.378889
       BIPM_CMCs |  -.0111483   .0019352    -5.76   0.000    -.0149487    -.007348
     LnCEPII_dist |  -.3593276   .0170198   -21.11   0.000     -.392752   -.3259032
LnISO_Population |  -.3039163   .0322507    -9.42   0.000    -.3672521   -.2405806
                 |
               U |
             L1. |   .3413457   .0453473     7.53   0.000     .2522903    .4304011
             L2. |   .3274964   .0430124     7.61   0.000     .2430264    .4119664
             L3. |   .2138388   .0418296     5.11   0.000     .1316915    .2959861
                 |
          lambda |   .2671742   .1439183     1.86   0.064    -.0154604    .5498088
                 |
            YEAR |
            2005 |     .18594   .0998805     1.86   0.063    -.0102106    .3820907
            2006 |  -.0869085   .1149942    -0.76   0.450    -.3127404    .1389233
            2007 |  -.0645406   .1169112    -0.55   0.581    -.2941372    .1650561
            2008 |  -.4025957   .1143387    -3.52   0.000    -.6271401   -.1780512
            2009 |  -.6200695    .106941    -5.80   0.000    -.8300861   -.4100529
            2010 |  -.4724193   .1060936    -4.45   0.000    -.6807716    -.264067
            2011 |  -.7035685   .1227088    -5.73   0.000    -.9445507   -.4625862
            2012 |  -.6555865   .1081226    -6.06   0.000    -.8679234   -.4432495
            2013 |  -.5369498   .1175432    -4.57   0.000    -.7677874   -.3061122
            2014 |  -.5093392   .1043534    -4.88   0.000    -.7142742   -.3044043
            2015 |  -.4638597   .1010315    -4.59   0.000    -.6622708   -.2654486
            2016 |   -.847744   .1133406    -7.48   0.000    -1.070328   -.6251596
            2017 |   -1.13372    .123136    -9.21   0.000    -1.375541   -.8918988
                 |
           _cons |  -16.87623   1.159872   -14.55   0.000    -19.15405   -14.59841
------------------------------------------------------------------------------
```

Wald test year dummy variables

```
 ( 1)  2005.YEAR = 0
 ( 2)  2006.YEAR = 0
 ( 3)  2007.YEAR = 0
 ( 4)  2008.YEAR = 0
 ( 5)  2009.YEAR = 0
 ( 6)  2010.YEAR = 0
 ( 7)  2011.YEAR = 0
 ( 8)  2012.YEAR = 0
 ( 9)  2013.YEAR = 0
 (10)  2014.YEAR = 0
 (11)  2015.YEAR = 0
 (12)  2016.YEAR = 0
 (13)  2017.YEAR = 0

       F( 13,   611) =   14.31
            Prob > F =    0.0000
```

Model E1

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -684.74634
Iteration 1:   log pseudolikelihood = -409.55789
Iteration 2:   log pseudolikelihood = -385.42056
Iteration 3:   log pseudolikelihood = -384.23362
Iteration 4:   log pseudolikelihood = -384.22658
Iteration 5:   log pseudolikelihood = -384.22658

Probit regression                               Number of obs     =      1,134
                                                Wald chi2(21)     =     265.54
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -384.22658               Pseudo R2         =     0.4389
```

TREATED	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	.9086213	.1943051	4.68	0.000	.5277903	1.289452
LnWB_GDPpcgbp	.9056653	.0904226	10.02	0.000	.7284403	1.08289
BIPM_CMCs	.0180889	.0144484	1.25	0.211	-.0102295	.0464072
LnCEPII_dist	-.2644694	.0717729	-3.68	0.000	-.4051417	-.1237971
LnISO_Population	.4578372	.0556985	8.22	0.000	.3486701	.5670043
pU						
L1.	-.4043971	.3368633	-1.20	0.230	-1.064637	.255843
L2.	-1.218282	.3341524	-3.65	0.000	-1.873209	-.5633555
L3.	-.1728158	.3183533	-0.54	0.587	-.7967768	.4511451
YEAR						
2005	-.0255025	.3001084	-0.08	0.932	-.6137042	.5626992
2006	-.6308183	.2764301	-2.28	0.022	-1.172611	-.0890253
2007	-.4415715	.3128403	-1.41	0.158	-1.054727	.1715841
2008	-.4775647	.2849589	-1.68	0.094	-1.036074	.0809446
2009	-.7631265	.2870537	-2.66	0.008	-1.325741	-.2005117
2010	-.7261184	.2838711	-2.56	0.011	-1.282496	-.1697412
2011	-1.151814	.2842389	-4.05	0.000	-1.708912	-.5947158
2012	-1.020335	.2869176	-3.56	0.000	-1.582683	-.4579863
2013	-.816073	.2881358	-2.83	0.005	-1.380809	-.2513373
2014	-.8319086	.2770377	-3.00	0.003	-1.374893	-.2889246
2015	-.59693	.2948821	-2.02	0.043	-1.174888	-.0189717
2016	-1.029355	.2912548	-3.53	0.000	-1.600204	-.4585058
2017	-1.25216	.3001407	-4.17	0.000	-1.840424	-.6638946
_cons	-11.46723	1.332091	-8.61	0.000	-14.07808	-8.856379

Estimation results

```
Linear regression                               Number of obs     =        648
                                                F(22, 625)        =     469.25
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9379
                                                Root MSE          =     .50353
```

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.239995	.0766003	16.19	0.000	1.08957	1.39042
LnWB_GDPpcgbp	1.336032	.0755381	17.69	0.000	1.187693	1.484371
BIPM_CMCs	-.0204606	.0016798	-12.18	0.000	-.0237594	-.0171618
LnCEPII_dist	-.4750386	.0159368	-29.81	0.000	-.5063349	-.4437424
LnISO_Population	-.1712238	.0280765	-6.10	0.000	-.2263595	-.1160881
U						
L1.	.3477257	.0455771	7.63	0.000	.2582229	.4372285
L2.	.3308797	.0429685	7.70	0.000	.2464995	.4152599
L3.	.2162795	.0415512	5.21	0.000	.1346826	.2978765
lambda	.2837195	.1424895	1.99	0.047	.0039034	.5635355
YEAR						
2005	.179506	.0978287	1.83	0.067	-.0126067	.3716187
2006	-.0918432	.112821	-0.81	0.416	-.3133972	.1297109
2007	-.0565486	.1140949	-0.50	0.620	-.2806043	.1675071
2008	-.4208652	.1123959	-3.74	0.000	-.6415846	-.2001458
2009	-.6633315	.1053443	-6.30	0.000	-.8702031	-.4564598
2010	-.5301543	.10457	-5.07	0.000	-.7355054	-.3248032
2011	-.7589205	.1197563	-6.34	0.000	-.9940938	-.5237471
2012	-.7149478	.1058302	-6.76	0.000	-.9227736	-.5071219
2013	-.6038617	.1155701	-5.23	0.000	-.8308144	-.376909
2014	-.5679046	.1016586	-5.59	0.000	-.7675384	-.3682707
2015	-.528842	.099399	-5.32	0.000	-.7240384	-.3336455
2016	-.9318848	.1110406	-8.39	0.000	-1.149943	-.7138269
2017	-1.234148	.1202439	-10.26	0.000	-1.470279	-.9980169
_cons	-18.9574	1.106662	-17.13	0.000	-21.13062	-16.78417

Wald test year dummy variables

```
 ( 1)  2005.YEAR = 0
 ( 2)  2006.YEAR = 0
 ( 3)  2007.YEAR = 0
 ( 4)  2008.YEAR = 0
 ( 5)  2009.YEAR = 0
 ( 6)  2010.YEAR = 0
 ( 7)  2011.YEAR = 0
 ( 8)  2012.YEAR = 0
 ( 9)  2013.YEAR = 0
 (10)  2014.YEAR = 0
 (11)  2015.YEAR = 0
 (12)  2016.YEAR = 0
 (13)  2017.YEAR = 0

       F( 13,   625) =   17.71
            Prob > F =    0.0000
```

Model F1

Probit model estimations results

```
Iteration 0:   log pseudolikelihood = -684.74634
Iteration 1:   log pseudolikelihood = -409.51636
Iteration 2:   log pseudolikelihood = -384.63265
Iteration 3:   log pseudolikelihood = -383.57267
Iteration 4:   log pseudolikelihood = -383.56968
Iteration 5:   log pseudolikelihood = -383.56968

Probit regression                               Number of obs     =      1,134
                                                Wald chi2(21)     =     264.67
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -383.56968               Pseudo R2         =     0.4398

------------------------------------------------------------------------------------
                   |               Robust
           TREATED |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
         LnWB_PRL |   .8229866   .1871868     4.40   0.000     .4561072    1.189866
     LnWB_GDPpcgbp |   .8983547   .0857304    10.48   0.000     .7303262    1.066383
BIPM_GeomCMCs_2019 |   .0141068   .0061784     2.28   0.022     .0019974    .0262161
      LnCEPII_dist |  -.2632319   .0694796    -3.79   0.000    -.3994094   -.1270544
  LnISO_Population |   .4387313   .0484111     9.06   0.000     .3438474    .5336152
                   |
                pU |
               L1. |  -.3673151   .3355828    -1.09   0.274    -1.025045    .2904151
               L2. |  -1.148176   .3382731    -3.39   0.001    -1.811179    -.485173
               L3. |  -.1402035   .3203951    -0.44   0.662    -.7681663    .4877593
                   |
              YEAR |
             2005  |  -.0221187   .3016949    -0.07   0.942    -.6134299    .5691925
             2006  |  -.6262168   .2788602    -2.25   0.025    -1.172773   -.0796607
             2007  |  -.4312465   .3143712    -1.37   0.170    -1.047403    .1849098
             2008  |  -.4567975   .2893741    -1.58   0.114     -1.02396    .1103653
             2009  |  -.7379481   .2925832    -2.52   0.012    -1.311401   -.1644956
             2010  |  -.6873472   .2880807    -2.39   0.017    -1.251975   -.1227194
             2011  |  -1.107572   .2892938    -3.83   0.000    -1.674577   -.5405666
             2012  |  -.9805239   .2917756    -3.36   0.001    -1.552394   -.4086542
             2013  |  -.7746445   .2931961    -2.64   0.008    -1.349298   -.1999908
             2014  |  -.7971529     .28125    -2.83   0.005    -1.348393    -.245913
             2015  |  -.5576459   .2989487    -1.87   0.062    -1.143575    .0282827
             2016  |  -.9876264     .29445    -3.35   0.001    -1.564738   -.4105151
             2017  |  -1.200028   .3020728    -3.97   0.000     -1.79208   -.6079761
                   |
             _cons |  -11.31658    1.25856    -8.99   0.000    -13.78331   -8.849846
------------------------------------------------------------------------------------
```

Estimation results

```
Linear regression                               Number of obs     =        648
                                                F(22, 625)        =     467.88
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9379
                                                Root MSE          =     .50375
```

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.216917	.0755912	16.10	0.000	1.068473	1.36536
LnWB_GDPpcgbp	1.275696	.0729066	17.50	0.000	1.132524	1.418868
BIPM_GeomCMCs_2019	-.0108429	.0009704	-11.17	0.000	-.0127486	-.0089371
LnCEPII_dist	-.4881393	.0157778	-30.94	0.000	-.5191231	-.4571554
LnISO_Population	-.216086	.0249955	-8.64	0.000	-.2651713	-.1670006
U						
L1.	.3497241	.0453659	7.71	0.000	.260636	.4388122
L2.	.3330051	.0430575	7.73	0.000	.2484503	.4175599
L3.	.219649	.0414944	5.29	0.000	.1381637	.3011344
lambda	.2524964	.1413125	1.79	0.074	-.0250085	.5300013
YEAR						
2005	.1800435	.0969131	1.86	0.064	-.0102713	.3703582
2006	-.0917324	.1126793	-0.81	0.416	-.3130083	.1295435
2007	-.0654124	.1140195	-0.57	0.566	-.28932	.1584953
2008	-.4188464	.1121239	-3.74	0.000	-.6390316	-.1986613
2009	-.6501224	.1053664	-6.17	0.000	-.8570374	-.4432075
2010	-.5146298	.1045431	-4.92	0.000	-.719928	-.3093316
2011	-.7416724	.1196057	-6.20	0.000	-.9765501	-.5067947
2012	-.6964487	.1051337	-6.62	0.000	-.9029068	-.4899906
2013	-.582492	.1151121	-5.06	0.000	-.8085452	-.3564387
2014	-.5482957	.1010734	-5.42	0.000	-.7467802	-.3498111
2015	-.5057929	.0988661	-5.12	0.000	-.6999429	-.3116429
2016	-.8972401	.1104117	-8.13	0.000	-1.114063	-.6804172
2017	-1.191867	.1191627	-10.00	0.000	-1.425875	-.9578592
_cons	-17.52871	1.026036	-17.08	0.000	-19.54361	-15.51382

Wald test year dummy variables

```
 ( 1)  2005.YEAR = 0
 ( 2)  2006.YEAR = 0
 ( 3)  2007.YEAR = 0
 ( 4)  2008.YEAR = 0
 ( 5)  2009.YEAR = 0
 ( 6)  2010.YEAR = 0
 ( 7)  2011.YEAR = 0
 ( 8)  2012.YEAR = 0
 ( 9)  2013.YEAR = 0
 (10)  2014.YEAR = 0
 (11)  2015.YEAR = 0
 (12)  2016.YEAR = 0
 (13)  2017.YEAR = 0

       F( 13,   625) =   17.19
            Prob > F =    0.0000
```

Link test

Source	SS	df	MS			
Model	2544.91068	2	1272.45534	Number of obs	=	648
Residual	156.90442	645	.243262667	F(2, 645)	=	5230.79
				Prob > F	=	0.0000
				R-squared	=	0.9419
				Adj R-squared	=	0.9417
Total	2701.8151	647	4.17591206	Root MSE	=	.49322

LnInvoices~p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_hat	1.030556	.1117716	9.22	0.000	.8110754	1.250036
_hatsq	.0010654	.0038823	0.27	0.784	-.0065581	.0086889
_cons	.2146846	.7942384	0.27	0.787	-1.344921	1.77429

Model G1

Probit model estimations results

```
Iteration 0:   log pseudolikelihood =  -502.7507
Iteration 1:   log pseudolikelihood = -267.42327
Iteration 2:   log pseudolikelihood = -237.43985
Iteration 3:   log pseudolikelihood = -234.94123
Iteration 4:   log pseudolikelihood = -234.93177
Iteration 5:   log pseudolikelihood = -234.93177

Probit regression                               Number of obs     =        896
                                                Wald chi2(21)     =     171.50
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -234.93177               Pseudo R2         =     0.5327
```

TREATED	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	1.288559	.3180387	4.05	0.000	.6652149	1.911904
LnWB_GDPpcgbp	1.57834	.1713821	9.21	0.000	1.242437	1.914243
LnBIPM_ProbCMCs_2019	-.1377743	.0618693	-2.23	0.026	-.2590358	-.0165128
LnCEPII_dist	-.5003829	.0934655	-5.35	0.000	-.683572	-.3171938
LnISO_Population	.8802888	.0991039	8.88	0.000	.6860487	1.074529
pU						
L1.	-.4771806	.4116271	-1.16	0.246	-1.283955	.3295938
L2.	-1.379842	.4044985	-3.41	0.001	-2.172644	-.5870391
L3.	-.3093736	.3641698	-0.85	0.396	-1.023133	.4043862
YEAR						
2005	.0337844	.376637	0.09	0.929	-.7044105	.7719792
2006	-.6591003	.380549	-1.73	0.083	-1.404963	.086762
2007	-.9389165	.4104708	-2.29	0.022	-1.743424	-.1344085
2008	-.8462657	.3994197	-2.12	0.034	-1.629114	-.0634174
2009	-1.199542	.4038257	-2.97	0.003	-1.991026	-.4080583
2010	-1.051187	.404565	-2.60	0.009	-1.84412	-.2582538
2011	-1.690581	.3995006	-4.23	0.000	-2.473588	-.9075741
2012	-1.474223	.4127151	-3.57	0.000	-2.28313	-.6653164
2013	-1.335644	.4126961	-3.24	0.001	-2.144514	-.5267745
2014	-1.195563	.3920083	-3.05	0.002	-1.963886	-.4272412
2015	-1.055583	.3810894	-2.77	0.006	-1.802504	-.3086613
2016	-1.827495	.3998056	-4.57	0.000	-2.6111	-1.043891
2017	-1.767026	.4216751	-4.19	0.000	-2.593494	-.9405577
_cons	-21.42177	2.214983	-9.67	0.000	-25.76305	-17.08048

Estimation results

Linear regression		Number of obs	=	564
		F(22, 541)	=	381.14
		Prob > F	=	0.0000
		R-squared	=	0.9394
		Root MSE	=	.49362

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	.3531163	.0888604	3.97	0.000	.1785625	.52767
LnWB_GDPpcgbp	2.050616	.0858666	23.88	0.000	1.881944	2.219289
LnBIPM_ProbCMCs_2019	-.2999818	.0363754	-8.25	0.000	-.3714361	-.2285275
LnCEPII_dist	-.3440262	.0168584	-20.41	0.000	-.3771422	-.3109102
LnISO_Population	-.2596521	.0276372	-9.40	0.000	-.3139415	-.2053627
U						
L1.	.3813392	.0488852	7.80	0.000	.2853111	.4773672
L2.	.2696926	.0495894	5.44	0.000	.1722813	.3671039
L3.	.1870145	.0474517	3.94	0.000	.0938023	.2802267
lambda	.2631043	.1762271	1.49	0.136	-.083069	.6092776
YEAR						
2005	.2073651	.0983897	2.11	0.036	.0140925	.4006378
2006	-.100305	.1205786	-0.83	0.406	-.3371645	.1365546
2007	-.1361158	.1240081	-1.10	0.273	-.3797123	.1074806
2008	-.3741386	.1130257	-3.31	0.001	-.5961617	-.1521156
2009	-.7027645	.1051438	-6.68	0.000	-.9093046	-.4962244
2010	-.5946361	.1055619	-5.63	0.000	-.8019976	-.3872746
2011	-.7316469	.1277446	-5.73	0.000	-.9825831	-.4807106
2012	-.8340134	.1081834	-7.71	0.000	-1.046524	-.6215025
2013	-.7306132	.1189322	-6.14	0.000	-.9642387	-.4969877
2014	-.7601733	.1061207	-7.16	0.000	-.9686323	-.5517142
2015	-.8752069	.1053666	-8.31	0.000	-1.082185	-.6682291
2016	-1.31973	.1149759	-11.48	0.000	-1.545584	-1.093876
2017	-1.583862	.1304663	-12.14	0.000	-1.840144	-1.327579
_cons	-25.2942	1.126147	-22.46	0.000	-27.50636	-23.08204

Wald test year dummy variables

```
( 1)  2005.YEAR = 0
( 2)  2006.YEAR = 0
( 3)  2007.YEAR = 0
( 4)  2008.YEAR = 0
( 5)  2009.YEAR = 0
( 6)  2010.YEAR = 0
( 7)  2011.YEAR = 0
( 8)  2012.YEAR = 0
( 9)  2013.YEAR = 0
(10)  2014.YEAR = 0
(11)  2015.YEAR = 0
(12)  2016.YEAR = 0
(13)  2017.YEAR = 0

      F( 13,   541) =   26.69
           Prob > F =    0.0000
```

Link test

Source	SS	df	MS			
Model	2143.86052	2	1071.93026	Number of obs	=	564
Residual	131.74827	561	.234845401	F(2, 561)	=	4564.41
				Prob > F	=	0.0000
Total	2275.60879	563	4.0419339	R-squared	=	0.9421
				Adj R-squared	=	0.9419
				Root MSE	=	.48461

LnInvoices~p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_hat	1.025649	.1336776	7.67	0.000	.763079	1.288219
_hatsq	.0008781	.0045625	0.19	0.847	-.0080835	.0098397
_cons	.1836985	.9659626	0.19	0.849	-1.713647	2.081044

Davidson and McKinnon test

Linear regression

Number of obs	=	564
F(23, 540)	=	373.49
Prob > F	=	0.0000
R-squared	=	0.9401
Root MSE	=	.49094

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	-.053928	.1857533	-0.29	0.772	-.4188155	.3109596
LnWB_GDPpcgbp	.5027104	.6956891	0.72	0.470	-.8638781	1.869299
LnBIPM_ProbCMCs_2019	-.0667173	.1109031	-0.60	0.548	-.2845716	.151137
LnCEPII_dist	-.0582354	.1267691	-0.46	0.646	-.3072565	.1907857
LnISO_Population	-.0458663	.0957251	-0.48	0.632	-.2339056	.142173
U						
L1.	.0934597	.1373007	0.68	0.496	-.1762493	.3631687
L2.	.00057	.1238784	0.00	0.996	-.2427726	.2439126
L3.	.010232	.0880899	0.12	0.908	-.1628089	.1832729
lambda	.1361465	.1879553	0.72	0.469	-.2330666	.5053596
YEAR						
2005	.0884302	.1077586	0.82	0.412	-.1232473	.3001076
2006	.0311728	.1330813	0.23	0.815	-.2302476	.2925933
2007	-.0325792	.1333002	-0.24	0.807	-.2944297	.2292714
2008	.008334	.1972017	0.04	0.966	-.3790425	.3957105
2009	-.1346089	.2734298	-0.49	0.623	-.6717253	.4025075
2010	-.1301134	.2389437	-0.54	0.586	-.5994864	.3392595
2011	-.0736479	.3299094	-0.22	0.823	-.7217109	.5744152
2012	-.1582627	.3194775	-0.50	0.621	-.7858336	.4693082
2013	-.1076353	.3087559	-0.35	0.728	-.7141451	.4988745
2014	-.1404245	.2973409	-0.47	0.637	-.7245111	.4436622
2015	-.1834525	.3182648	-0.58	0.565	-.8086413	.4417363
2016	-.2953846	.4624266	-0.64	0.523	-1.20376	.6129908
2017	-.3211824	.5756842	-0.56	0.577	-1.452037	.8096726
hat	.8124062	.3552309	2.29	0.023	.1146025	1.51021
_cons	-6.024281	8.643885	-0.70	0.486	-23.00404	10.95548

```
 ( 1)  2005.YEAR = 0
 ( 2)  2006.YEAR = 0
 ( 3)  2007.YEAR = 0
 ( 4)  2008.YEAR = 0
 ( 5)  2009.YEAR = 0
 ( 6)  2010.YEAR = 0
 ( 7)  2011.YEAR = 0
 ( 8)  2012.YEAR = 0
 ( 9)  2013.YEAR = 0
 (10)  2014.YEAR = 0
 (11)  2015.YEAR = 0
 (12)  2016.YEAR = 0
 (13)  2017.YEAR = 0

       F( 13,   540) =    0.28
            Prob > F =    0.9946
```

```
Linear regression                               Number of obs     =        564
                                                F(23, 540)        =     378.21
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9401
                                                Root MSE          =     .49095
```

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	.7974128	.4136427	1.93	0.054	-.0151332	1.609959
LnWB_GDPpcgbp	1.194412	.5552636	2.15	0.032	.1036707	2.285153
BIPM_CMCs	-.0162823	.0076036	-2.14	0.033	-.0312186	-.001346
LnCEPII_dist	-.3545565	.1723232	-2.06	0.040	-.6930625	-.0160504
LnISO_Population	-.1227573	.0785469	-1.56	0.119	-.2770523	.0315376
U						
L1.	.2793594	.150092	1.86	0.063	-.0154764	.5741952
L2.	.2086622	.1141562	1.83	0.068	-.0155825	.4329069
L3.	.1435273	.0815319	1.76	0.079	-.0166312	.3036858
lambda	.3468526	.2163862	1.60	0.110	-.0782092	.7719144
YEAR						
2005	.1861038	.1350021	1.38	0.169	-.0790898	.4512974
2006	-.0363622	.1195494	-0.30	0.761	-.271201	.1984766
2007	-.0587587	.1210213	-0.49	0.628	-.296489	.1789715
2008	-.2547488	.1668369	-1.53	0.127	-.5824778	.0729801
2009	-.5284917	.2717653	-1.94	0.052	-1.062338	.0053552
2010	-.4357111	.2246478	-1.94	0.053	-.8770017	.0055796
2011	-.5374838	.2691439	-2.00	0.046	-1.066181	-.0087866
2012	-.574859	.2864681	-2.01	0.045	-1.137587	-.0121305
2013	-.4576488	.2348011	-1.95	0.052	-.9188843	.0035867
2014	-.4611261	.2329202	-1.98	0.048	-.9186668	-.0035853
2015	-.4536222	.2393071	-1.90	0.059	-.9237092	.0164647
2016	-.7978794	.3953984	-2.02	0.044	-1.574587	-.0211719
2017	-1.013126	.4822558	-2.10	0.036	-1.960454	-.0657988
hat2	.2325329	.3738366	0.62	0.534	-.5018194	.9668851
_cons	-16.46988	7.64537	-2.15	0.032	-31.48819	-1.451568

```
( 1)  2005.YEAR = 0
( 2)  2006.YEAR = 0
( 3)  2007.YEAR = 0
( 4)  2008.YEAR = 0
( 5)  2009.YEAR = 0
( 6)  2010.YEAR = 0
( 7)  2011.YEAR = 0
( 8)  2012.YEAR = 0
( 9)  2013.YEAR = 0
(10)  2014.YEAR = 0
(11)  2015.YEAR = 0
(12)  2016.YEAR = 0
(13)  2017.YEAR = 0

      F( 13,   540) =    0.37
           Prob > F =    0.9794
```

Model H1

```
Fitting Poisson model:

Iteration 0:   log pseudolikelihood = -.04372071
Iteration 1:   log pseudolikelihood = -.04195329
Iteration 2:   log pseudolikelihood =  -.0411505
Iteration 3:   log pseudolikelihood = -.04103759
Iteration 4:   log pseudolikelihood = -.04103737
Iteration 5:   log pseudolikelihood = -.04103737

Fitting constant-only model:

Iteration 0:   log pseudolikelihood =  -.0474305
Iteration 1:   log pseudolikelihood = -.04546309
Iteration 2:   log pseudolikelihood = -.04546225
Iteration 3:   log pseudolikelihood = -.04546211
Iteration 4:   log pseudolikelihood = -.04546211  (backed up)

Fitting full model:

Iteration 0:   log pseudolikelihood = -.04546211
Iteration 1:   log pseudolikelihood = -.04244636
Iteration 2:   log pseudolikelihood = -.04111914
Iteration 3:   log pseudolikelihood = -.04103873
Iteration 4:   log pseudolikelihood =  -.0410374  (not concave)
Iteration 5:   log pseudolikelihood = -.04103738

Negative binomial regression                    Number of obs     =      1,377
                                                Wald chi2(21)     =    3050.04
Dispersion           = mean                     Prob > chi2       =     0.0000
Log pseudolikelihood = -.04103738               Pseudo R2         =     0.0973
```

InvoicesPop	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
LnWB_PRL	1.254849	.1273218	9.86	0.000	1.005303	1.504395
LnWB_GDPpcgbp	1.280785	.1277083	10.03	0.000	1.030482	1.531089
BIPM_CMCs	-.049127	.0021096	-23.29	0.000	-.0532617	-.0449922
LnCEPII_dist	-.558762	.0384018	-14.55	0.000	-.6340281	-.4834959
LnISO_Populat~n	.3325169	.0344214	9.66	0.000	.2650522	.3999816
YEAR						
2002	1.193434	.215571	5.54	0.000	.7709228	1.615946
2003	1.259559	.2182692	5.77	0.000	.8317593	1.687359
2004	1.55385	.2046693	7.59	0.000	1.152706	1.954995
2005	1.587963	.20764	7.65	0.000	1.180996	1.99493
2006	1.378669	.212164	6.50	0.000	.962835	1.794502
2007	1.473126	.1928031	7.64	0.000	1.095239	1.851013
2008	1.136466	.2214286	5.13	0.000	.7024738	1.570458
2009	.8965622	.2368459	3.79	0.000	.4323528	1.360772
2010	1.119067	.1909451	5.86	0.000	.7448213	1.493313
2011	.9678055	.2108571	4.59	0.000	.5545333	1.381078
2012	1.006595	.1932551	5.21	0.000	.6278219	1.385368
2013	1.024209	.1948271	5.26	0.000	.6423548	1.406063
2014	1.05371	.1941153	5.43	0.000	.6732514	1.434169
2015	1.006142	.2031869	4.95	0.000	.6079032	1.404381
2016	.7788933	.1852894	4.20	0.000	.4157327	1.142054
2017	.5680915	.2217141	2.56	0.010	.1335399	1.002643
_cons	-26.92236	1.807223	-14.90	0.000	-30.46445	-23.38026
/lnalpha	-28.77903	.			.	.
alpha	3.17e-13	.			.	.

Model I1

Linear regression						
				Number of obs	=	524
				F(13, 510)	=	172.82
				Prob > F	=	0.0000
				R-squared	=	0.7969
				Root MSE	=	.93909

LnInvoicesPop	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.234759	.1433983	8.61	0.000	.9530354	1.516483
LnWB_GDPpcgbp	1.272517	.1189786	10.70	0.000	1.038768	1.506265
BIPM_CMCs	-.0162184	.0039194	-4.14	0.000	-.0239186	-.0085183
LnCEPII_dist	-.4702591	.0440446	-10.68	0.000	-.5567902	-.383728
LnISO_Populat~n	-.2424955	.0505897	-4.79	0.000	-.3418852	-.1431057
YEAR						
2010	.1127343	.1730359	0.65	0.515	-.2272166	.4526851
2011	.105014	.185326	0.57	0.571	-.2590822	.4691103
2012	.0739921	.1717113	0.43	0.667	-.2633564	.4113406
2013	.0815179	.1705558	0.48	0.633	-.2535605	.4165962
2014	.1321456	.1776032	0.74	0.457	-.2167782	.4810695
2015	.1358086	.1722977	0.79	0.431	-.2026919	.4743092
2016	-.1947744	.1818467	-1.07	0.285	-.5520351	.1624864
2017	-.3414605	.1882499	-1.81	0.070	-.7113012	.0283802
_cons	-17.92705	1.702443	-10.53	0.000	-21.27172	-14.58239

Model J1

Linear regression				Number of obs	=	909
				F(22, 886)	=	168.94
				Prob > F	=	0.0000
				R-squared	=	0.7990
				Root MSE	=	.88004

LnAmount	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
LnWB_PRL	1.227039	.1015138	12.09	0.000	1.027803	1.426274
LnWB_GDPpcgbp	1.093873	.0755286	14.48	0.000	.945637	1.242109
BIPM_CMCs	-.0135464	.0025653	-5.28	0.000	-.0185812	-.0085117
LnCEPII_dist	-.355838	.0307832	-11.56	0.000	-.4162546	-.2954214
LnISO_Populat~n	.6993046	.0334827	20.89	0.000	.6335899	.7650193
LnAvPrice	1.00568	.0364522	27.59	0.000	.9341374	1.077223
YEAR						
2002	.9811902	.1767436	5.55	0.000	.6343052	1.328075
2003	.9848605	.1890606	5.21	0.000	.6138017	1.355919
2004	1.185591	.1830361	6.48	0.000	.8263565	1.544826
2005	1.259619	.1835897	6.86	0.000	.8992971	1.61994
2006	1.017596	.1795931	5.67	0.000	.6651183	1.370073
2007	1.141101	.1735722	6.57	0.000	.8004407	1.481762
2008	.6147978	.1917575	3.21	0.001	.2384458	.9911498
2009	.4780578	.1794569	2.66	0.008	.1258475	.8302681
2010	.6568226	.1790258	3.67	0.000	.3054585	1.008187
2011	.6006337	.1925906	3.12	0.002	.2226466	.9786208
2012	.5496334	.1828935	3.01	0.003	.1906784	.9085884
2013	.5882491	.1836467	3.20	0.001	.2278158	.9486824
2014	.6182118	.1883186	3.28	0.001	.2486092	.9878143
2015	.6223945	.1807776	3.44	0.001	.2675922	.9771968
2016	.3331467	.1836443	1.81	0.070	-.0272819	.6935752
2017	.1888625	.1911476	0.99	0.323	-.1862925	.5640175
_cons	-16.74989	1.058168	-15.83	0.000	-18.8267	-14.67308

More information

Contact us

For more information, contact us:

Senior Economist Mike King , mike.king@npl.co.uk
Economist Eugenio Renedo, eugenio.renedo@npl.co.uk

National Physical Laboratory
Hampton Road, Teddington, Middlesex, TW11 0LW
Web: https://www.npl.co.uk/digital
Tel: 020 8977 3222