Evaluating demand for letter price elasticities and technology impacts in an evolving communications market*

Frédérique Fève[†], Jean-Pierre Florens[†], Frank Rodriguez[#], Soterios Soteri[‡] and Leticia Veruete-McKay[‡]

1. INTRODUCTION

A rich econometrics literature exists on the demand for mail. In recent times the focus of these studies, be they on individual countries, or across different countries, has tended to concentrate on three key aspects of the demand for mail: the impact of the economy; the substitution of letter mail by e-communications; and price elasticities. Estimates for each of these factors vary between studies but the overall evidence, with a few exceptions, yield broadly similar conclusions. In particular, the findings of econometric analysis of the demand for letters over the past few years have yielded the following three conclusions, amongst others. First, economic conditions have a significant impact on the demand for letters[1]. Second, the net impact of technology is having a negative impact on the overall demand for letters.[2] Third, price elasticities of letter demand differ by traffic stream or segment, but in total (that is, for all segments of traffic and net of any switching between different letter products) the aggregate letter price elasticity of demand is relatively low and somewhat below unity[3].

However, in contrast to the largely consistent findings of the econometrics literature, there has been a healthy and at times vigorous discussion in the UK that these findings could potentially be misleading. In particular, a question of some considerable debate is whether standard econometric time series techniques could be systematically attributing too much of the observed decline in mail volumes to structural factors and too little to price related effects when modelling the demand for mail in an evolving communications market.

We have examined issues concerning the demand for letters in an evolving communications market in two previous papers. In Fève et al. (2010) we explored how changes in the communication market can impact postal demand and how, via iterative analysis of outcomes

---

* The analysis contained in this paper reflects the views of the authors and not necessarily those of Royal Mail Group.
† Toulouse School of Economics (IDEI), University of Toulouse Capitale.
‡ Royal Mail Group
# Oxera

[1] Econometric estimates of the economic activity elasticity of the demand for letters include: Boldron et al. (2010) estimate this could lie in the range 0.4 to 0.7 for France; Meschi et al. ((2011), using panel data on 13 countries of the European Union, estimates this to be, on average, around 0.8 for the countries included in their study; Veruete-Mckay et al. (2011) estimate this to be around 1.1 in the UK.
[2] See Meschi et al. (2011), Nikali (2008) and Veruete-McKay (2011)
[3] Boldron et al, (2010), Thress (2006) and Veruete-McKay (2011) estimate price elasticities for France, the USA and the UK respectively. Estimates from this body of research suggest that the total letter price elasticity in these countries is likely to be in the range -0.2 to -0.7. Price elasticity estimates for Finland contained in Nikali (2008) for B2B and C2X segment are low and consistent with other international results. However, in contrast, Nikali (2008) estimate that Finland's B2C letter price elasticity could be around -1.4, which is somewhat higher than estimates found in most other international studies.

compared to projections, it is possible to learn about the nature of structural changes. In Fève et al. (2012) we provided an empirical example of the scale of uncertainty inherent in long term projections of the demand for mail and the need to explicitly assess and incorporate the possibility of regime change, or of structural breaks, into the forecast process when using data based models to project the future in an evolving market environment.

In this paper we examine how an applied econometrician may attempt to estimate the demand for mail using standard econometric techniques and partial information at different stages of the evolution of the communications market. In particular we construct a dynamic model of the demand for communications in an evolving environment and simulate outputs for this model which the econometrician only partially observes and uses to estimate demand for letter models at different points in time. This analysis explores a hypothetical econometrician's choice of variables and model specification and the potential effect these can have on the estimated elasticities for price and economic activity and estimates of the impact of esubstitution on the demand for mail.

Section 2 sets out a simple stylised model, where the evolution of technology depends on the characteristics of the population which is independent of letter prices, but the demand for postal versus ecommunications is dependent on the price of letters and digital alternatives (as in De Donder et al. (2012). Section 3 describes how a hypothetical econometrician may proceed to estimate a demand for letter model in the light of observed outcome data but without knowledge of the true underlying model. This section provides estimated regression outputs for different model specifications that the econometrician could estimate at different stages in the evolution of the communication market and compares them with the "true" results from the underlying model. This approach provides insights into the nature and possible size of errors that might be made by the hypothetical econometrician. Finally, section 4 draws a number of conclusions of the extent to which standard econometric time series could be over or under estimating letter price elasticities.

2. THE MODEL

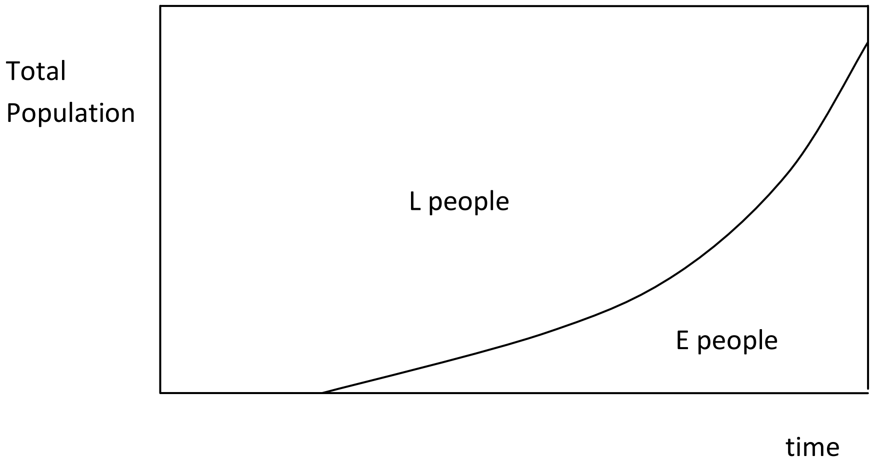
A simple stylised model of the demand for mail is adopted for ease of exposition. This model considers demand through a sequence of time periods of $t = 1,...T$ where at each point in time t there exist two categories of communication receivers, letter mail receivers ("Lpeople") and digital or electronic mail receivers ("Epeople"). The model assumes that Lpeople do not use the internet and may only be contacted by or send letter mail whereas Epeople use the internet and may be contacted by, or in some cases send, letter or digital mail. At time t it is assumed that there exist N_t^L Lpeople and N_t^E Epeople.

The model is based on the demand for mail by senders such that they match receiver preferences to communicate by post or electronic means. Underpinning the model is an assumption that senders of mail, which are mainly businesses sending transactional mail, are to some extent able to influence receiver communication preferences (see De Donder et al. 2012). This is captured in the model in two key ways: first, via the impact of the price variables for letters and digital communications (P_L and P_E respectively); and second, through the increasing number of individuals over time who have access to technology that could potentially be used to substitute letter mail (that is, via the rising number of Epeople variable, N_t^E).

The evolution of Epeople and Lpeople

For simplicity we assume that the population is constant, such that $N_t^L + N_t^E = N_t$ (that is, *N* is fixed for all time period *t=1, …,T*) but the proportion of Epeople (N_t^E/N_t) increases over time such that the number of Lpeople who can only be contacted by letter mail declines over time and progressively reduces. For example, the evolution of Epeople and Lpeople can be considered to follow a profile such as that displayed in Figure 1.

Figure 1. The evolution of Epeople and Lpeople



The demand for mail model

It is assumed that the sender demand for mail at each point in time, *t*, comprises of three relationships. First, the demand for letter mail sent to people who can only receive messages by letter (that is, mail sent to and received by Lpeople, Q_L^L, equation (1))), which depends on the number of people who do not have access to the internet (N^L), the price of letters (P_L) and economic activity (*Y*). Second, the demand for letter mail sent to people who could receive ecommunications (that is, letter mail sent to Epeople, Q_L^E, equation (2)), which is a function of the number of Epeople (N^E), the price of ecommunications (P_E), current and lagged price of letters (that is, P_{Lt} and P_{Lt-1}) to account for the possibility that it takes time for senders of mail to react to changes in letter prices, and economic activity (*Y*). Third, the demand for ecommunications sent to people with internet access (that is, Q_E^E, equation (3)) which is determined by the number of Epeople (N^E), the price of ecommunications (P_E) and economic activity (*Y*).

This simple theoretical demand for mail model can be expressed as follows:

$$Q_{Lt}^L = N_t^L\, C_0\, P_{Lt}^{\beta}\, Y_t^{\beta l} \varepsilon_t^u \qquad (1)$$

$$Q_{Lt}^E = N_t^E\, C_1\, P_{Lt}^{\beta 0} P_{Lt-1}^{\beta 1}\, P_{Et}^{\gamma}\, Y_t^{\beta le} \varepsilon_t^v \qquad (2)$$

$$Q_{Et}^E = N_t^E\, C_2\, P_{Et}^{\delta}\, Y_t^{\beta ee} \varepsilon_t^w \qquad (3)$$

where C_0, C_1 and C_2 are constant terms; α, β, βl, $\beta 0$, $\beta 1$, βle, βee γ, and δ, are model parameters; and ε_t^u , $\varepsilon_t^v$and ε_t^w are random disturbance terms. The number of Epeople versus Lpeople, economic activity, and the price of ecommunications and letters are assumed to be exogenous variables. In particular, the evolution of Epeople at each point in time (N_t^E) is assumed to follow a logistic curve, such as that shown in Figure 1 and described in (4); as the population is assumed to be fixed, the proportion of individuals only able to receive mail by post at each period in time (N_t^L) is part of an identity, as noted in (5); and the price of ecommunications is assumed to follow a stochastic function, such as (6).

$$N_{Et} = 0 \text{ for } t<T, \text{ otherwise } N_{Et} = C_3 \frac{e^{\propto(t-T)}}{1+e^{\propto(t-T)}} + C_4 \tag{4}$$

$$N_t^L + N_t^E \equiv N_t \tag{5}$$

$$P_{Et} = \phi P_{Et-1} + \varepsilon_t^x \tag{6}$$

where C_3 and C_4, are constant terms; α and ϕ are model parameters; $e^{\propto(t-T)}$ is the exp(α(t-T)); and ε_t^x is a random disturbance term.

The total demand for letters (Q_{Lt}) in this model is assumed to be observable and equal to (1) plus (2), that is:

$$Q_{Lt} = Q_{Lt}^L + Q_{Lt}^E \tag{7}$$

but none of the individual variables on the left hand side of the relationships (1) to (6) are assumed to be observable, although on the right hand side of these equations P_L, P_{Lt-1} and Y are observable.

Within this simple theoretical model the elasticities of the demand for mail prior to the evolution of new technology having a net negative impact on mail volumes (that is, where N_t^E is equal to zero) are equal to those in equation (1). In this case, the own-price elasticity of the demand for letters is β and the elasticity for economic activity is βl. However, as new technology evolves throughout the population and creates opportunities for substituting letters by ecommunications (that is, where $N_t^E > 0$) the demand for letter elasticities are non-linear time dependent functions of (1) and (2) that are influenced by the profile of letter prices (P_{Lt}) and ecommunication prices (P_{Et}) and the evolution of the mix of N_t.

In the very long run, as the proportion of Epeople tends towards unity the own-price elasticity of letter demand would tend towards $\beta 0+\beta 1$ and the economic activity elasticity for letter demand would tend towards βle. However, over the medium to long term, where the proportion of the population able to receive ecommunications is not equal to unity (that is, $N_{Lt}^L > 0$) the model price and income elasticities for letters are a non-linear and time dependent function of equations (1) and (2) and depend on the evolution of Epeople (N_{Et}) and the relative price of sending communications via letters (P_{Lt}) or digital alternatives (P_{Et}). The model therefore exhibits time varying price elasticities that depend on the evolution of technology in the population over time and is similar to the concepts raised by Nikali (2011). By contrast, in the model, the price and income elasticities for ecommuications (in equation (3)) are assumed to be unchanged.

The simulation model

Having specified a simple theoretical model of the demand for mail we proceed to calibrate it with parameter values and assess its properties via simulation techniques over 50 time periods (that is, t = 1, 2, …., 50) prior to exploring how an applied econometrician may attempt to estimate the demand for mail at different points in time using standard econometric techniques.

The hypothetical *true* parameter values used to populate the model are reported in Table 1. In some cases, these are informed by publicly available studies. However, it should be noted that the values are stylized and should not be considered to be values to any particular national postal operator. For simplicity the simulation model assumes all the economic activity parameters in equations (1), (2) and (3) are equal to unity (that is, $\beta l=\beta le=\beta ee$ =1) which results in the economic activity elasticity of demand for letters and ecommunications remaining constant over time. Economic activity (Y_t) and the price of letters (P_{Lt}) at each point in time are assumed to be fixed and determined outside of the model. In particular, the profile for Y_t is set equal UK GDP and letter prices (P_{Lt}) are assumed to equal those for total addressed inland letter prices in the UK deflated by the retail prices index.

The properties of the model using these parameter values are summarized in Figure 2. In general, the model assumes that the total volume of communications ($Q_L^L + Q_L^E + Q_E^E$) is increasing over time, but as the number of Epeople increases (from period T=26 onwards) the volume of letter communications tends to decline. The overall price of letters is assumed to trend downwards relatively slowly over time whereas the price of ecommunications is assumed to exhibit a more pronounced downward trend. A noticeable feature of the model is that the population is not assumed to fully comprise of Epeople over the simulation period. In fact, the model assumes that after 20 time periods from the start of the technology related substitution process the proportion of Epeople increases to almost three-quarters of the total population and thereafter increases slowly over time. This reflects a very long run s-shaped logistic profile (embodied in equation (4)) akin to Nikali's concept of a long term corrugated s-curve (see Nikali 2008) driven by the overlapping and interaction of different technology drivers and sender-recipient relationships.

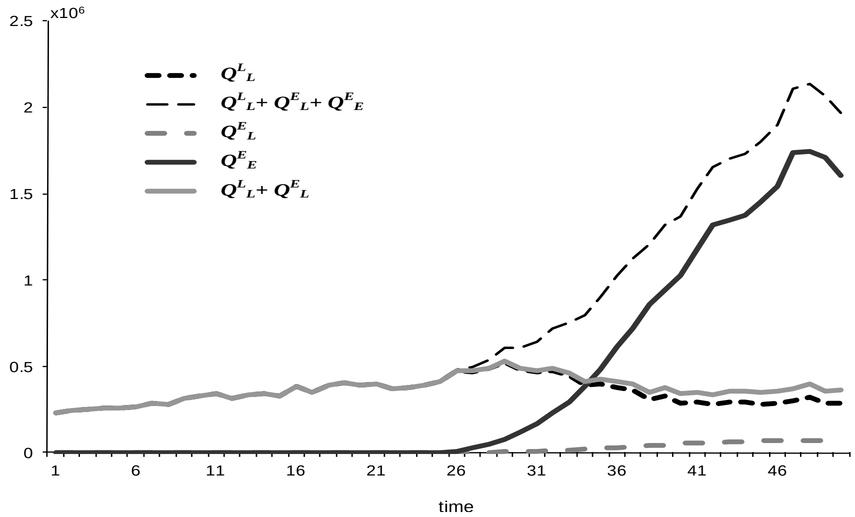
The simulation properties of the theoretical model yield a true price elasticity of demand for letters that increase over time (from around -0.4 to -0.5) and depend on the evolution of technology related esubstitution amongst the population (that is, the number of Epeople). A further point to note about the simulation properties of the theoretical model is that during the early phase of technology related substitution, say t=26 to 30, there is no noticeable impact on the demand for letter mail or the letter price elasticity. However, in the following 10 periods, t=31 to 40, as the number of Epeople increases and the price of ecommunications decline there is a considerable decline in the demand for letters and the price elasticity of letters increases over time. Over the last 10 years of the simulation period, t=41 to t=50, both the rate of decline in letter demand and the letter price elasticity of demand start to level out as the proportion of Epeople by then is increasing very slowly over time.

Table 1. Hypothetical true parameters for theoretical model					
Illustrative parameters for endogenous model variables					
Demand for letters to communicate with Lpeople (Q_L^L)					
$C_0 = 4$		$\beta = -0.4$	$\beta l = 1.0$		u~ N(0, $(0.03)^2$)
Demand for letters to communicate with Epeople (Q_L^E)					
$C_1 = 2$	$\beta 0 = -0.25$	$\beta 1 = -0.5$	$\beta 1e = 1.0$	$\gamma = 0.25$	v~ N(0, $(0.03)^2$)
Demand for ecommunications to communicate with Epeople (Q_E^E)					
$C_2 = 2$		$\delta = -0.4$	$\beta ee = 1.0$		w~ N(0, $(0.02)^2$)
Illustrative parameters for exogenous model variables					
Evolution of Epeople					
$C_3 = 0.78$		$C_4 = -0.025$	$\alpha = 0.34$		t = 1,2, …,50
Price of ecommunications (P_E)		Price of letters (P_L)		Economic activity (Y)	
$\phi = 0.97$	x~ N(0, 0.05)	Equal to UK total addressed inland letter traffic prices[1]		Equal to UK GDP[2]	

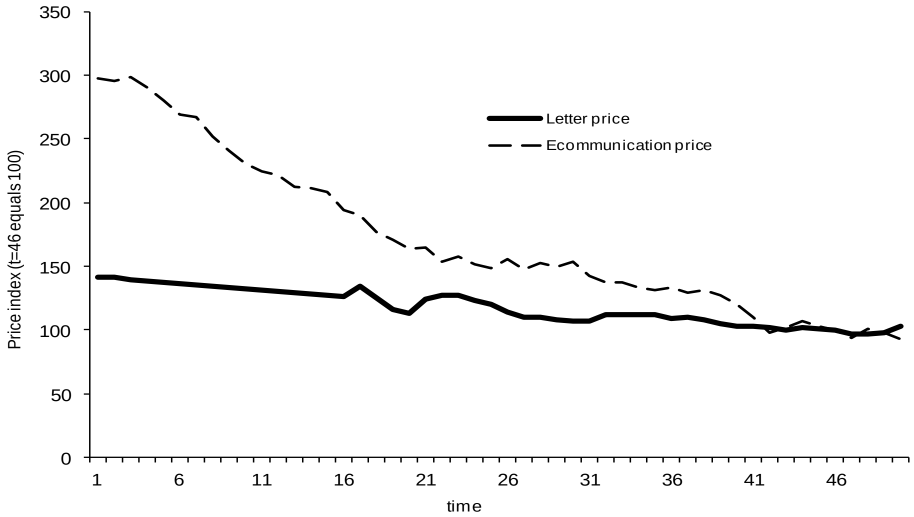
Note:
1. For periods t=16 to t=50 letter prices set equal to estimates for UK total addressed inland letter traffic prices for the financial year period 1976/77 to 2010/11. For periods t=1 to t=15 letter prices were assumed to follow a downward trend.
2. Equal to UK GDP for the period 1960 to 2009

Figure 2. Hypothetical stylized model true values

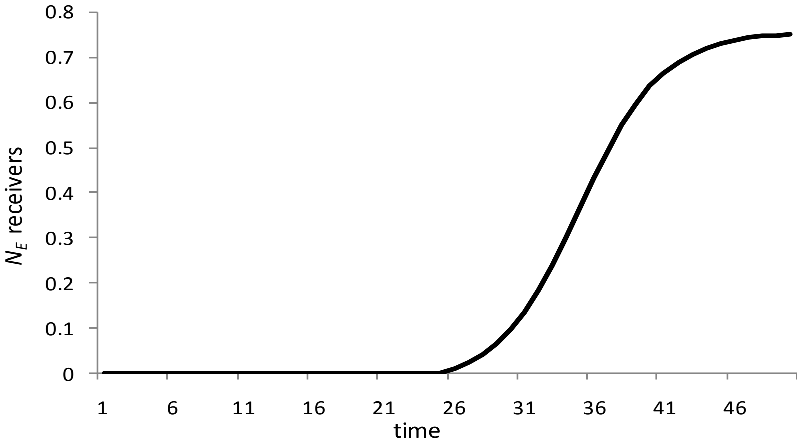
2a. The demand for letters and ecommunications



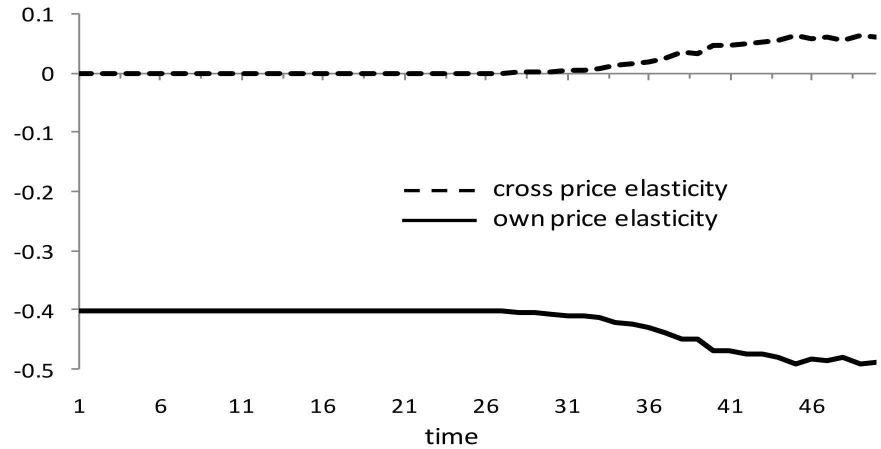
2b. Price of letters and ecommunications



2c. Evolution of Epeople



2d. Letter price elasticities



3. ESTIMATING THE MODEL OVER TIME: AN ILLUSTRATIVE EXAMPLE

This section explores the extent to which standard econometric time series techniques can provide useful estimates of the demand for mail in an environment where a hypothetical econometrician does not have full knowledge of the true model and cannot observe all the variables influencing the demand for mail. In particular, we provide estimates of a number of regression models that the econometrician could estimate at different points in time. Five key stages are examined. Stage 0, the pre-Epeople period (that is, t=1 to t=25); Stage 1, the early technology adopter phase which covers the period t=26 to t=30; Stage 2, the first half of the main Epeople transition phase which refers to the period t=31 to t=35; Stage 3, the second half of the main Epeople transition phase which relates to the period t=36 to t=40; and, finally, stage 4 which covers the last 10 periods of the simulation model, t=41 to t=50.

The hypothetical econometrician

The hypothetical econometrician in our paper is tasked with estimating the demand for letter traffic in the absence of knowing the *true* demand for mail model and not being able to observe all the variables underpinning it.

While the econometrician does not know the true model, it is assumed that they have a theoretical demand for mail model framework in mind that is broadly similar to that outlined in equations (1) to (3), but no knowledge of the true parameter values. In full awareness that they are unable to observe all variables, the hypothetical econometrician is assumed to proceed to estimate a demand for mail model using standard econometric techniques with observed data on the total volume of letter traffic (Q_L) the price of letters (P_L), economic activity (Y) and proxy variables to account for technology related letter substitution. Furthermore, in the knowledge that they are unable to observe Q_L^L, Q_L^E, Q_E^E, N_L , N_E and , P_E the econometrician is assumed to be aware that they do not know when the letter substitution process truly started nor the extent to which it has progressed to at any moment in time.

The challenges facing our hypothetical econometrician and the regression analysis they are assumed to adopt to estimate the demand for mail are very similar to a number of postal econometric time series studies. In fact, most demand for mail studies using time series regression techniques include economic activity variables, the price of letters, a variety of proxy variables to account for the impact of technology, as well as a number of other factors[4]. With regards to variables that are used to proxy the impact of technology related substitution no study that we are aware of includes an explicit estimate of the price of ecommunications (P_E). Instead, most studies incorporate variables that attempt to measure the spread of internet related technologies or simple time trends. For example, Boldron et al. (2010) and Meschi et al. (2011) use broadband penetration variables, Nikali (2008) includes email usage variables, Soteri et al. (2009) incorporates the percentage of advertising spending that goes to the Internet as does Veruete-Mckay et al. (2011) which also include time trend break variables to account for technology related substitution effects on the demand for letter traffic.

[4] For example, other factors such as quality of service, the price of non-internet related substitutes, seasonal dummies and other variables tend to be included in applied econometric studies. For simplicity, these additional factors were excluded from our theoretical model and illustrative econometrics analysis.

Illustrative econometric estimates in stage 0: the pre-Epeople phase (*t*=1 to *t*=25)
It is assumed that prior to technology potentially having a net negative impact on letter mail (that is up to period *t*=25, where $N^E = 0$) the econometrician estimates a demand for mail model with the following specification using ordinary least squares (OLS):

$$q_{Lt} = \hat{c}_0 + \hat{\beta} p_{Lt} + \hat{\beta} l y_t + \hat{u}_t \qquad (8)$$

where lower case variables refer to the natural logarithm of upper case variables contained in table 1 (for example $y_t = \ln(Y_t)$).

The estimated model using the simulation model variables yields the results reported in Table 2. The OLS estimated parameters for both the economic activity variable (Y) and the price of letters (P) are close to the true values reported in Table 1. One point to note from the estimated results in Table 2 is that the estimated GDP coefficient has a high level of statistical significance but the price term has a lower level of statistical significance. In addition, the model properties of the estimated model are reasonable. For example, the R2 is high and the Durbin Watson statistic is around 2 and close to its desired value. In practice the applied econometrician would apply a battery of diagnostic tests to evaluate the properties of the model and inform their model selection strategy, see for example, McAleer (1995) and Hoover and Perez (1999)[5]. However, this is not the focus of this paper and we refrain from undertaking an extensive econometric model selection investigation and examine a more limited set of indicators, namely estimated parameter t-statistics, R^2 values and the Durbin Watson statistic

The adoption of a strict 5% significance criterion could lead the hypothetical econometrician to erroneously reject the price variable from the model and conclude that the price elasticity of demand is not significantly different from zero. This dilemma is one that applied econometricians confront on a regular basis. In general, most econometricians at this point would explore ways of improving the model specification, consider adopting a different model specification, or accept a lower level of statistical significance on the grounds that the estimated elasticity is consistent with economic theory and, if external evidence exists, similar to the findings of other studies. Given we know the true model and our interest is in the estimation properties of the model from *t*=26 onwards, when the model begins to reflect the effect of esubstitution, we refrain from estimating alternative model specifications and including additional variables in this paper.

Table 2. Stage 0, the pre Epeople estimated econometric model

	Estimated coefficient	T statistic	Memo item: Table 1 simulation model true parameter value
y	0.88	7.9	1.00
p	-0.50	-1.5	-0.40
c	3.44	1.2	4.00
R^2 = 0.95	Durbin Watson =1.93		Sample period *t*= 1,2,...,25

[5] For example, applied econometricians using the LSE econometric modelling approach would adopt diagnostic tests that would include the Durbin Watson test, LM tests for serial correlation, heteroskedasticity tests, functional form tests and parameter stability tests amongst others, to test down from a general to specific model and assess model specification.

Stage 1 of the Epeople transition period: the early adopter phase (*t*=26 to *t*=30)

When a new technology comes into existence it can take many years to realise its potential and, as is often the case, may take forms that its early inventors may have not quite imagined. For example electronic and internet related communication technology has been around for some time, however, it was not until the early to mid 2000s that this technology started to emerge as a household medium and even then it was increasing from a very low base. Since the majority of letter traffic in advanced postal networks is sent by businesses to individuals, stage 1 of the substitution phase is assumed to refer to the early adopter Epeople phase, that is periods t=26 to t=30, where the number of people able to receive ecommunications from business senders is slowly starting to increase from a zero base.

An important question to examine is how our hypothetical econometrician would attempt to estimate or update their model of the demand for letters during stage 1 of the Epeople transition period. For simplicity we assume that the econometrician undertakes such a task soon after the end of period t=30. At this point their estimation strategy is likely to include the following two paths, amongst others: first, to update their previous model (8) with new information (that is, data for t=26 to t=30); second, to include new variables to account for changes in the external environment.

Table 3 reports five regression results that are consistent with such a strategy. In particular, R3.1 reports the estimated coefficients generated by updating the pre-Epeople model reported in table 2 with five additional data observations (t=26 to t=30). The regressions R3.2 to R3.4 include time trend break terms to account for potential negative effects of people using new technology to reduce letter volumes and R3.5 includes a proxy variable for the unobservable proportion of Epeople choosing to receive digital communications instead of letters. The latter is assumed to be an unbiased proxy for N_E. This may however be more difficult to achieve in reality, as variables that are typically used, such as households with internet or broadband access, may not necessarily be good indicators of the proportion of Epeople. Furthermore, good quality statistics on variables such as this are unlikely to exist in the early phase of the adoption of a new technology. The investment and time required to test and disseminate high quality statistics from official sources (such as the UK Office for National Statistics) is likely to be available only a number of years after such a technology has established its importance. Therefore during the early adopter phase, and possibly beyond, such proxy variables may be informed by survey information.

In the early adopter Epeople phase, there is very little to choose between the five sets of results reported in table 3 on statistical grounds. All five regressions yield correctly signed coefficients for the economic activity and price variables and they posses reasonably sized t-statistics, as well as comparable adjusted R^2 and Durbin Watson statistics. In fact, it is a moot point, but if the time trend term variables and Epeople proxy variable were assessed against a 5% statistical region criterion they would be dropped from the model to leave a model specification identical to that for the pre Epeople phase. However, if such a model were adopted, this would lead to large forecasting errors going forward, as noted by Cazals et al. (2008) in their analysis of forecasting errors using models containing an unknown break in the trend.

The letter price elasticity of demand estimates contained in Table 3, using data up period *t*=30, lie in the range -0.50 to -0.69 and are all higher than the true price elasticity of demand of -0.40 in period *t*=30. Estimates containing a lagged price term yielded insignificant estimates and are not reported. Furthermore, tests for a structural break in letter prices after *t*=25 also yielded insignificant results All the estimated economic activity elasticities were in a tight range of 0.85 to 0.88.

Table 3. Stage 1, the early adopter Epeople phase

	Memo item: Table 2 pre-Epeople l		R3.1		R3.2		R3.3		**R3.4**		R3.5		Simulation model true values
	Est. Coeff	T Stat	Est. Coeff	T Stat	Est. Coeff	T Stat	Est. Coeff	T Stat	**Est. Coeff**	**T Stat**	Est. Coeff	T Stat	
y	0.88	7.9	0.87	8.0	0.87	7.8	0.85	7.7	**0.86**	**8.1**	0.85	7.8	1.00
p	-0.50	-1.5	-0.52	-1.9	-0.62	-2.0	-0.66	-2.1	**-0.65**	**-2.3**	-0.69	-2.3	-0.40
Trend T23					-0.00	-0.7							
Trend T26							-0.01	-0.9					
Trend T29									**-0.04**	**-1.7**			
Epeople ratio proxy											-0.68	-1.3	
R2	0.9456		0.9628		0.9634		0.9640		**0.9664**		0.9650		
Adjusted R^2	0.9406		0.9600		0.9592		0.9598		**0.9625**		0.9610		
Durbin Watson	1.93		1.92		1.94		1.96		**2.04**		1.98		
Sample period	T=1,2,...25		T=1,2,...30		T=1,2,...30		T=1,2,...30		**T=1,2,...30**		T=1,2,...30		

Note: The term T23 refers to a trend break term starting in period t=23. Such that: prior to *t*=23, T23=0; in period *t*=23, T23=1; in period *t*=24, T23=2; in period *t*=25, T23=3; and so on. The other trend terms (T26 and T29) are generated in a similar way.

In terms of choosing a preferred model specification to inform business analysis the hypothetical econometrician is assumed to choose a model with: coefficients possessing signs consistent with economic theory and high t-statistics; acceptable diagnostic statistics; and some maximum statistical or information criterion[6]. In our illustrative econometric modelling of the demand for mail our hypothetical econometrician is assumed to proceed on this basis and uses the Durbin Watson as a test for model specification and the highest value of the adjusted R^2 to choose between statistically acceptable models.

The result of adopting such a model selection strategy would be to choose model R3.4. So that in the early adopter Epeople phase our hypothetical econometrician would estimate a demand for mail model that included a time trend break term starting in period *t*=29 (that is, the variable T29) which is 3 years after the true Epeople phase came into existence) and is estimated to be subtracting around 4% from letter volumes per period from *t*=29 onwards[7].The estimated letter price elasticity of demand of -0.65 in the preferred model would be somewhat higher than the true elasticity of -0.40 in stage 1 (*t*=26 to *t*=30) but not significantly different from this value,

[6] Such as the Schwarz Bayesian criterion, Akaike criterion or adjusted R^2 statistics.

[7] Table 3 contains a selection of time trend break terms in order to provide an illustrative example of alternative estimates in choosing a preferred model. It is not an exhaustive examination of all possibilities which, if this was undertaken, could result in different estimated parameters.

whereas the estimated economic activity elasticity of 0.86 would be a little lower than its true value of unity but again not significantly different from its true value. This suggests that OLS regression analysis is able to provide reasonable estimates of the impact of economic activity but price elasticity estimates may potentially be subject to a greater degree of uncertainty.

A notable feature of the chosen model (R3.4) is that the estimated price elasticity is somewhat higher than was estimated by the econometrician in stage 0 (reported in table 2), that is -0.65 versus -0.50, and could lead to the erroneous conclusion that the price elasticity of letter demand has increased due to competition from digital media. Instead, it is the case that both estimated elasticities are not significantly different to their true values and differences between them may be due to random noise. To properly examine this point would require Monte Carlo analysis and running the simulation model many hundreds of times using different parameter values for variables reported in Table 1 and repeating the regression analysis each time. Such a process is not the focus of this paper.

Stage 2. The first half of the main Epeople transition period (*t*=31 to *t*=35)
Stage 2 of the model transition phase represents the era in which observed letter traffic volume data begins to show a clear decline. It is likely that this stage of the technology transition period best characterises the position that many countries with advanced postal networks have been in recent years.

In this environment the hypothetical econometrician is assumed to estimate a demand for letter model having observed time period *t*=35 and therefore observed a clear downward trend in the demand for letters. The econometrician is therefore likely to be actively searching for new variables to include in their model specification to account for the increasing penetration of new technology amongst the population and, equally important, sender preferences to take advantage of new and lower cost communication media. A key issue here is that it is not clear which variables should be included in the econometric model. For it is not simply a matter of including estimates of the penetration rate of a particular technology, as esubstitution does not solely depend on the availability of technology, but also on how and to what extent it is being used to substitute for letter mail, which depends on the relative price of letter and digital communications in our model.

The time varying profile of the true price elasticity during *t*=31 to *t*=35 reported in Figure 2 raises the possibility that linear OLS regression analysis using time trends or technology trending variables could lead to an underestimation of the letter price elasticity of demand. However, while this is a possibility, it is also the case that any potential bias could work in the opposite direction. For example, it is also possible that the inclusion of poor technology penetration proxies, or poor time trend break terms, may not properly capture the loss in letter traffic in the regression analysis and this impact could be allocated elsewhere in the model. Any such unaccounted substitution impact would only have one of three places to go: first it could lead to downward bias in the economic activity term; second, it could generate upward bias in the estimated price elasticity; or third it could be go into the model residual category and yield poor model property evaluation criteria (such as a low R^2 or poor Durbin Watson statistic).

Table 4 contains the results of a number of regressions run by the hypothetical econometrician using three groups of variables to account for the rising level of esubstitution taking place in stage 2 of the evolution of Epeople. The first group contain a number of single trend break terms starting at different points in time; the second group undertakes a similar exercise with two break terms in each regression, as the econometrician is assumed to suspect multiple non-linearities are occurring in the true model; and the third group includes a proxy variable for the proportion of Epeople in the population.

Specifications including a lagged term in letter prices and structural break letter price elasticities were again tested. Similar to our previous results the estimated models containing these terms yielded poor results (in terms of wrongly signed variables, coefficients being very close to zero, and low t-statistcs) and are not reported. In fact, this seems to be a general finding across all the Epeople stages and we do not refer to model specifications of this type again in the paper.

Prior to examining the properties of the regression outputs reported in Table 4, it is worth noting that the exclusion of any terms to account for increasing substitution lead to poor estimated model properties. For example, as hypothesised above, the omission of variables to account for substitution has led to all three concerns manifesting themselves in regression R4.1. That is, the estimated magnitude of the economic activity variable is biased downwards substantially; the estimated price elasticity is biased upwards significantly; and the Durbin Watson statistic is considerably worse than all the other estimated models reported in table 4. The latter observation would lead our hypothetical econometrician to quickly reject this model and consider alternative model specifications.

A model evaluation exercise based on the same criteria as that described in stage 1 would lead the hypothetical econometrician to prefer model R4.9. This model estimates the economic activity elasticity to be 0.88 which is reasonably close to unity and not significantly different from it. The estimated price elasticity is -0.57 which is higher than the true price elasticity of around -0.42 in stage 2 (equal to the average letter price elasticity between t=30 and t=35) but again not significantly different from it.

It is noticeable that a number of the estimated models reported in table 4 could be acceptable to the hypothetical econometrician. For example, there is not too much to differentiate between the six models R4.4, R4.6, R4.7, R4.8, R4.9 and R4.10 on statistical grounds. The range of estimates for the economic activity variable from these models is in the range 0.84 to 0.88, which are a little lower than the true elasticity of unity, but nevertheless provide reasonable estimates. With respect to the estimated price elasticities these lie in the range -0.55 to -0.71 which is somewhat higher than the true elasticity of around -0.42.

A number of additional regressions were performed to test the presence of a break in the price elasticity coefficient but in all cases the estimated coefficients were either wrongly signed or possessed very low estimated coefficients and t-statistics. This suggests that OLS time series econometric techniques may be able to provide reasonable central estimates for income and price elasticities if model specifications incorporate time trend break terms, or non-linear proxy variables to account for esubstitution, and possess reasonable model properties and diagnostic statistics. But such techniques might not be able to identify slow moving changes in the true price elasticity resulting from structural changes in the composition of the population and incentives by senders to reduce the demand for letters.

Table 4. Stage 2, the first half of the rapid Epeople transition period

	Econometric model using single trend break term									
	R4.1		R4.2		R4.3		**R4.4**		R4.5	
	Est. Coeff	T Stat	Est. Coeff	T Stat	Est. Coeff	T Stat	**Est. Coeff**	T Stat	Est. Coeff	T Stat
Y	0.54	3.3	0.76	6.0	0.77	6.7	**0.88**	8.4	0.85	7.3
P	-0.99	-2.3	-1.18	-3.7	-1.04	-3.6	**-0.55**	-2.1	-0.45	-1.5
Trend T23			-0.02	-5.3						
Trend T26					-0.03	-6.4				
Trend T30							**-0.05**	7.9		
Trend T33									-0.10	-6.6
R^2	0.8966		0.9456		0.9552		**0.9657**		0.9572	
Adj. R^2	0.8901		0.9404		0.9508		**0.9624**		0.9530	
DW.	0.74		1.47		1.70		**1.98**		1.62	

Sample period for all regressions R4.1 to R4.5 is *t*=1,2,...35

	Econometric model using double trend break terms								Model using Epeople proxy		Simulation model true values
	R4.6		R4.7		R4.8		**R4.9**		R4.10		
	Est. Coeff	T Stat	Est. Coeff	T Stat	Est. Coeff	T Stat	**Est. Coeff**	T Stat	Est. Coeff	T Stat	
y	0.84	7.7	0.87	7.8	0.87	7.9	**0.88**	8.4	0.86	8.1	1.00
p	-0.66	-2.2	-0.59	-1.9	-0.64	-2.2	**-0.57**	-2.2	-0.71	-2.7	-0.40 to -0.42
Trend T23	0.01	0.9									
Trend T26			-0.01	-0.8							
Trend T27					-0.02	-2.0					
Trend T28	-0.04	-3.7									
Trend T29							**-0.03**	-4.4			
Trend T30											
Trend T31			-0.05	-2.6							
Trend T32					-0.04	-2.1					
Trend T33											
Trend T34							**-0.05**	-1.7			
Epeople ratio proxy									-0.85	-7.7	
R^2	0.9626		0.9633		0.9642		**0.9674**		0.9646		
Adjusted R^2	0.9576		0.9584		0.9594		**0.9631**		0.9611		
DW.	1.91		1.99		2.00		**2.10**		1.99		

Sample period for all regressions R4.6 to R4.10 is *t*=1,2,...35

Note:
The term T23 refers to a trend break term starting in period t=23. Such that: prior to *t*=23, T23=0; in period *t*=23, T23=1; in period *t*=24, T23=2; in period *t*=25, T23=3; and so on. All the other trend terms in the table are generated in a similar way.

Stage 3. The second half of the main Epeople transition period (*t*=36 to *t*=40)

Stage 3 of the Epeople transition phase potentially represents the most interesting time period of our simulated model, as it portrays a period of high access and usage of ecommunications with steadily declining observed letter volumes. It is likely that this stage of the technology transition period best characterises the position that many countries are likely to be in or will be entering in the near future.

In this environment the hypothetical econometrician is assumed to estimate a demand for letter model having observed time period *t*=40 and therefore a long period of declining letter volumes. The econometrician by this time is assumed to have settled on using an approach based on time trend break terms (such as Veruete-McKay et al., 2011) or incorporating proxy variables to account for letter substitution (such as Nikali, 2008 or Meschi et al., 2010) and update their preferred model, or preferred set of models, on a regular basis.

Proceeding on this basis, the hypothetical econometrician is assumed to have rejected model specifications with no trend break and a single trend break term in favour of a double trend break model or proxy variable for esubstitution. In effect the double trend break model is a flexible linear approximation to a general s-shaped technology curve prior to it starting to flatten out. A number of illustrative estimated outputs using both approaches are reported in table 5.

The OLS regression results reported in table 5 reach broadly similar conclusions to those reported in the lower half of table 4. In particular, the model specification R5.3 which incorporates trend breaks starting in the time periods towards the end of stage 1 and end of stage 2 are preferred in terms of overall model properties. That is, on the basis of correctly signed and statistically significant time trend parameters, the highest adjusted R^2 criterion and a reasonable Durbin Watson statistic. Interestingly, the other double trend specifications reported in table 5 are considerably less satisfactory, which suggests that while the double trend break specification may perform well for periods of time it can be sensitive to the time periods one chooses.

The estimated price elasticity in the double trend break model (-0.54) is close to the true non-linear price elasticity model (which lies in the range -0.43 to -0.47), as is the economic activity variable (0.88) which is not significantly different to its true value. Similar conclusions are also reached when using the Epeople proxy variable reported in equation R5.4. It therefore seems that static OLS can potentially yield reasonably good estimates for economic activity and price elasticity estimates when structural changes are impacting the demand for letter mail. However, the quality of these estimates can be affected by the choice of variables chosen, as shown by the deterioration in the price elasticity estimates reported in R5.2 and to a lesser extent R5.1.

Table 5. Stage 3, the second half of the rapid Epeople transition period

Econometric model using double trend break terms or Epeople proxy

	R5.1		R5.2		R5.3		R5.3		R5.4		Simulation model true values
	Est. Coeff	T Stat	Est. Coeff	T Stat	**Est. Coeff**	T Stat	Est. Coeff	T Stat	Est. Coeff	T Stat	
y	0.91	6.27	0.97	8.2	**0.88**	8.3	0.77	6.2	0.95	7.3	1.00
p	-0.30	-0.75	-0.23	-0.7	**-0.54**	-2.1	-0.56	-1.8	-0.59	-1.8	-0.43 to -0.47
Trend T23	0.03	2.7									
Trend T26			0.01	0.9							
Trend T28	-0.09	-7.4									
Trend T30					**-0.05**	6.7					
Trend T31			-0.08	-7.6							
TrendT33											
TrendT34							-0.12	-11.8			
TrendT35					**-0.04**	-3.1	0.12	1.70			
TrendT39											
Epeople ratio proxy									-1.18	-15.7	
R^2	0.9233		0.9501		**0.9609**		0.9421		0.9362		
Adjusted R^2	0.9146		0.9443		**0.9563**		0.9353		0.9309		
DW.	1.26		1.94		**2.20**		1.33		1.50		

Sample period for all regressions R5.1 to R5.4 is *t*=1,2,...40

Note:
The term T23 refers to a trend break term starting in period *t*=23. Such that: prior to *t*=23, T23=0; in period *t*=23, T23=1; in period *t*=24, T23=2; in period *t*=25, T23=3; and so on. The other trend terms in the table are generated in a similar way.

Stage 4. The final phase of the Epeople transition period (*t*=40 to *t*=50)

In the final phase of the Epeople transition period the increase in the number of people with access to technology that enables letter substitution starts to slow down considerably, such that in the final years of the simulation period the number of Epeople is hardly increasing at all (see Figure 2).

The change in the trend rate of increase in Epeople during stage 4 of the technology related transition phase leads to less satisfactory econometric model estimates using the double trend break model. For example, while a number of model specifications yield reasonable estimates for economic activity (around -0.9) and price elasticities (-0.4 to -0.6) the Durbin Watson statistic declines considerably to around 0.8 and the adjusted R^2 also declines somewhat.

It is quite possible that the introduction of a third break term to account for the slowdown in the diffusion of technology related letter substitution could improve the model but this is not explored in this paper. The key point to note here is that the use of time trends to account for factors changes are mainly backward looking and model specifications using specific time trends may not be reliable for long periods of time. It is therefore important to regularly update such models and undertake rigorous diagnostic tests to examine the properties of econometric models.

The estimated properties of the model specifications using the Epeople proxy variable performs a little better in terms of its estimated elasticities for economic activity (around 0.95) and letter prices (-0.6), but again it performs poorly in terms of its Durbin Watson statistic. Note that the inclusion of a lagged price variable in the model, as in the true model (see equation (2)), does not improve matters.

4. SUMMARY AND CONCLUSIONS

This paper has examined how an applied econometrician may attempt to estimate the demand for mail using standard econometric techniques in an environment where they observe partial information at different stages of the evolution of the communications market.

In order to explore this issue in a logical and consistent manner we constructed a dynamic model of the demand for communications in an evolving environment and simulated the outputs for this model. We then explored how a hypothetical econometrician might estimate a demand for letter model using only partial information from the true model during different stages of development in the communications market. In particular the analysis explored the hypothetical econometrician's choice of variables and model specification and the potential effect these can have on the estimated elasticities for price and economic activity and estimates of the impact of esubstitution on the demand for mail.

A particular point to note from the paper is that developments in the communications market are a dynamic diffusion of technology and demographic trends that interact with economic incentives (such as the relative price of letters versus digital communications) to generate non-linear and time varying price elasticities. The paper shows that where the combined impact of advances in technology and changes in demographic trends lead to relatively slow changes in letter price elasticities over time, then standard econometric time series techniques can provide reasonable estimates for price and economic activity elasticities.

However, because of the unobservable nature of how technology is actually impacting letter demand and the absence of variables that provide satisfactory proxies for this process it can be difficult to obtain reasonable estimates of the impact of substitution. It is worth noting, perhaps surprisingly, that time trend break terms perform a little better than esubstitution proxy variables in terms of both statistical criteria and producing price elasticity estimates that are nearer to the true price elasticity.

Furthermore, the incorporation of poor proxies may adversely impact price elasticity and economic activity estimates of the demand for letters. In order to avoid, or at least limit, the likelihood of this taking place this paper highlights the need to adopt an econometric model selection strategy that encompasses a number of criteria.

In particular, the paper showed that model selection criteria that simultaneously examined the following four factors helped to avoid the adoption of models with highly misleading estimates for letter price and economic activity elasticities of mail demand. First, estimated coefficients should have signs that are consistent with economic theory or strong business priors. Second, variables with t-statistics that do not automatically pass some standard pre-set level of significance (e.g. a 5% test-statistic region) should be included in the specification if there are valid economic or business reasons for including such variables. Third, econometric diagnostic tests should be used to assess model properties and inform model selection criteria. Fourth, statistical selection criteria can be used to help choose between competing model specifications.

When the impact of structural changes on letter demand is relatively small, such as in stage 1 of the evolution of technology in our simple model (that is, the early adopter Epeople phase), adopting the above four pronged model selection strategy generates little benefit in terms of providing more accurate estimates for price elasticity and economic activity. Clear benefits of adopting such a model selection strategy are highlighted in stages 2 and 3 of the technology evolution phase (that is during the rapid transition period to higher numbers of Epeople). In particular, the results reported in tables 4 and 5 show that the four pronged selection criteria led to an automatic rejection of a number of poorly specified models with misleading estimated price and economic activity price elasticity. While the benefits relative to model specifications with similar statistical diagnostic tests (e.g. Durbin Watson and adjusted R^2 statistics) were less clear.

In conclusion therefore, in an evolving communications market where technology and demographic trends are interacting and leading to slowly changing price elasticities over time, our theoretical model suggests that econometric time series techniques can provide reasonable estimates for price and economic activity elasticities if the econometrician avoids serious model specification problems and regularly reviews and updates their model, as recommended by Fève et al. (2010). However, as always, random error and differences in econometricians' model specification criteria will lead to variations in estimates and the econometrician is unlikely to be able to pinpoint the estimated elasticities accurately. The extent to which our results would hold under different scenarios and whether econometric estimates for price and income elasticities of demand for letters in an evolving communication market are unbiased remain topics for future potential research.

REFERENCES

Boldron, F., C. Cazals, J.P. Florens and S. Lécou (2010), 'Some dynamic models for mail demand: the French case', in M.A. Crew and P.R. Kleindorfer (eds), *Heightening Competition in the Postal and Delivery Sector*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 99-115.

Cazals , C. , J.P. Florens, F. Rodriguez and S. Soteri (2008), 'Forecast uncertainty in dynamic models: an application to the demand for mail', in M.A. Crew and P.R. Kleindorfer (eds), *Competition and Regulation in the Postal and Delivery Sector*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 63-73.

De Donder, P., H. Cremer, P. Dudley and F. Rodriguez (2012), 'Competition Between Mail and Electronic Substitutes in the Financial Sector : a Hotelling Approach'. To be presented at the Institut d' Economie Industrielle and La Poste 7th conference "The Economics of the Postal Sector in the Digital World", Toulouse, March 22-23, 2012.

Fève, F., J.P. Florens, F. Rodriguez and S. Soteri (2010), 'Forecasting mail volumes in an evolving market environment', in M.A. Crew and P.R. Kleindorfer (eds), *Heightening Competition in the Postal and Delivery Sector*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 116-34.

Fève, F., J.P. Florens, L. Veruete-McKay, S. Soteri and F. Rodriguez (2012), 'Uncertainty and projections of the demand for mail', in M.A. Crew and P.R. Kleindorfer (eds), *Multi-modal Competition and the Future of Mail*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 77-93.

Hoover, K.D.and S.J. Perez (1999), ‘Data mining reconsidered: encompassing and the general-to-specific approach to specification search’, *The Econometrics Journal, 1999*, volume 2, pp.167-191.

McAleer, M.(1995), ‘Sherlock Holmes and the search for truth : a diagnostic tale’, in *Surveys in Econometrics*, Cambridge, MA, USA, Basil Blackwell, pp 91-138.

Meschi, M., M. Cherry, C. Pace and M. Petrova (2011), ‘Understanding the impact of e-substitution on letter mail volumes: a multi-country panel study’, in M.A. Crew and P.R. Kleindorfer (eds), *Reinventing the Postal Sector in an Electronic Age*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 47-60.

Nikali, H. (2008), 'Substitution of letter mail for different sender-receiver segments', in M.A. Crew and P.R. Kleindorfer (eds), *Competition and Regulation in the Postal and Delivery Sector*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 89-106

Nikali, H. (2011), ‘Does the level of price elasticity change with the progression of substitution?’, in M.A. Crew and P.R. Kleindorfer (eds), *Reinventing the Postal Sector in an Electronic Age*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 36-46.

Soteri , S., F. Fève, J.P. Florens and F. Rodriguez (2009), 'Internet advertising and direct mail: trends and analysis for the UK', in M.A. Crew and P.R. Kleindorfer (eds), *Progress in the Competitive Agenda in the Postal and Delivery Sector*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 209-22.

Veruete-McKay, L., S. Soteri, J.C. Nankervis, and F. Rodriguez (2011) ‘Letter traffic demand in the UK: an analysis by product and envelope content type’ in *Review of Network Economics*, Vol.10: Issue 3.