The effect of user interaction on the demand for mobile text messages: evidence from cross-country data (2002-2007)*

Bruno Basalisco†

First version: June 2007; This version: March 2009

Abstract

This work measures mobile telephony demand estimating a point-to-point model of the text messaging traffic exchanged cross-network. The empirical analysis relies on a dataset spanning 15 countries for the period 2002-2007. Our econometric estimates of impact and total price elasticities confirm that incoming traffic is a key driver of SMS demand. This study thus confirms the findings of the existing point-to-point literature while extending it by disentangling the impact of incoming traffic from that of network size. The results hold implications for theories of network competition (the separability of demand debate), optimal pricing by telecommunications operators and regulatory analyses.

Keywords: Mobile telephony; SMS; Price elasticities; Network effects; Information exchange; Point-to-point demand; Complementarity; Separability of demand; Network competition.

JEL classification: L13, L43, L96.

---

*My thanks go to my Ph.D. supervisor Tommaso Valletti for his advice, support and encouragement and to Jonathan Sandbach for his help in gaining access to the data and for many valuable comments. I would also like to thank participants to the ITS 19th European Conference and to seminars at LBS and University of Pisa for helpful comments. The usual disclaimer applies.

†Imperial College Business School, London SW7 2AZ, UK. E-mail: b.basalisco@imperial.ac.uk

1 Introduction

> "SAMUEL MORSE would be proud. The form of communication he pioneered, the telegram, may have gone the way of the horse-drawn carriage, but it lives on through that most modern means of keeping in touch: the text message." (The Economist, 2003, April 5th)

Mobile text messaging (also known as SMS, or simply texting) is an extremely popular means of communication. While originally conceived as a simple network application for notifications, mobile texts have fastly grown into a broad business and are particularly popular among young consumers. For instance in the UK, the birthplace of the text message, users sent 30+ billion messages in 2005 and nearly 20% of country-wide mobile expenditure could be accounted for by SMS, "originally an unexpected by-product of mobile networks" (Office of Communications, 2006a, p.18). As a reflection of the dramatic growth in the use of text messaging across the world, the global mobile messaging market was worth approximately $130 billion in revenues in 2008 and is expected to climb to $224 billion by 2013, according to Portio Research (The Washington Post, 2008).[1]

Moreover, mobile operators, on top of charging users for their text messages usage, are ever more often conveying SMS-advertisements to users as an extra source of revenues. Indeed, operators are experimenting with altogether new business models centred on messaging. Blyk, a mobile virtual network operator (MVNO) which has recently entered the UK market (and is expanding into other European countries) has focused on text advertising as its only source of revenue, while offering free mobile service to its users.[2]

As the SMS market takes off, so does the interest of telecommunications regulators, several of

[1] While in North America texting had a slow diffusion compared to most of Europe (The Economist, 2003), recent US figures indicate that 75 billions texts were sent in June 2008 alone. Moreover, wireless data services (primarily texting) now account for 20 percent of US mobile operators' revenues (Reardon, 2008), which is broadly in line with the UK figures above. In France, once amongst the few European laggards in the diffusion of texting (The Economist, 2004), "the number of SMS exchanged multiplied by 8.6 in volume between 2000 and 2005. Exchanges of data (SMS and MMS) represent about 12% of mobile operators' sales. In 2005, close to 13 billion SMS were sent, generating sales of over €1 billion." (Autorité de Régulation des Communications Électroniques et des Postes, 2006a).

[2] Blyk offers a bundle of 217 text messages and 43 call minutes per month as part of its free user allowance. Users must opt-in to receive up to 6 advertising texts per day. Only those aged 16 to 24 can sign up.

which have conducted analyses of SMS markets. In a nutshell, regulators are debating whether each mobile operator holds significant market power associated to a competitive bottleneck in the delivery of SMS addressed to its customers – which could lead to the imposition of ex-ante remedies. Mobile operators face the prospect of a sharp reduction in SMS termination rates, along the lines of what occurred for voice telephony. For instance, in 2006 France's ARCEP regulated the wholesale SMS price, imposing a ceiling on termination charges. A prominent empirical issue for this regulatory debate is the measurement of SMS demand elasticities and whether there is any substitutability between mobile voice and text communications.[3] This has prompted a series of recent studies which have modelled and estimated intra-modal demand in mobile markets (Andersson et al., 2006; Grzybowski and Pereira, 2008; Kim et al., 2008).

This work measures the impact of user interaction on SMS demand. We do so by adapting models of point-to-point telecommunications demand – so far applied to estimate fixed telephony demand – to the context of mobile telephony. In fact, the key innovation in the point-to-point literature (reviewed in the next section) is the acknowledgement that users of telecommunications services do not consume these in isolation. For instance, it is plausible to think that a mobile user may have a greater wish to send an SMS following reception of an SMS. This may be because her utility depends on the information gathered by the SMS exchange – and her reply adds to that.[4] Intuitively, this on its turn may lead to a further exchange and so on. Alternatively, a user may feel less the wish to send an SMS, having received one – possibly because this would not yield any gains in information. In general, the consumption behaviour of a given user will be affected by an interaction of such kind. In fact, our reference user will send more SMS following a reduction of its network SMS price. Furthermore, her SMS correspondents – at the receiving end of a greater number of SMS – will now have a different propensity to send SMS, even when they face the same price of sending texts (in the case where they are on other networks).[5] This on its turn will affect our reference user's wish to send SMS. As a consequence, the overall impact of a network's price variation on demand may be compounded or mitigated by the dynamics

[3] Having found a low correlation (0.13) between quarterly growth in SMS and mobile-to-mobile voice traffic in France (2001-2005), ARCEP concluded that mobile texting and calls are not substitutes – also in the light of qualitative factors reflecting the cultural aspects of consumer preferences (Autorité de Régulation des Communications Électroniques et des Postes, 2006b, p.52).

[4] This scenario has been defined "reciprocal communication" (Larson et al., 1990; Taylor, 1994).

[5] The propensity to send SMS will be higher in the reciprocity scenario, lower otherwise.

of user interaction.[6] Moreover, a retail price variation, due to its return effect on the amount of incoming communication, will affect termination revenues. Elasticity estimates revised in the light of user interaction can thus inform network operators' optimal pricing policies – with respect to both retail and wholesale revenues.

In the light of this, we will investigate empirically whether telecommunications demand does not only respond to price but also to the level of received communication. We calculate panel estimates of text messaging demand using a novel cross-country monthly usage and pricing dataset provided by Vodafone Group Plc. In fixed telephony markets, the point-to-point literature shows that calls generate calls – to the tune of "one-half to two-thirds of a call in return" (Taylor, 2004, p.129). None of these works, though, fully disentangles the interaction of network size with the incoming traffic variable, which is likely to result in inflated estimates for the latter (Taylor, 1994; Taylor, 2004). Our study is the first in this field to provide demand estimates accounting for the impact of both the level of incoming traffic and the size of the network (via the community of interest proxy, described below). We conclude that SMS generate up to one-fourth of an SMS in return. The elasticities computed yield implications for mobile operators' optimal pricing choice.

Finally, the findings gathered are relevant to those seeking to model competition between networks. In fact, several demand models have assumed that the utility that users gain from outgoing communication is independent from incoming communication – the separability assumption. This tenet cannot be consistent with the evidence of complementarity between sent and received SMS, which supports instead a reciprocal communications scenario. Moreover, as discussed below in section 2.1, whether the reciprocity assumption holds has implications on network firms' incentives to interconnect and at what price. Specifically, if users' behaviour is consistent with a reciprocal communications scenario, then networks have fewer incentives to induce a "connectivity breakdown" (e.g. via off-net communications prices spiralling upward) – which could substantially mitigate regulatory concerns.[7]

[6] We will show in the analysis below that user interaction always compounds the impact of price changes (i.e. the interaction-driven change in consumption and the price variation have opposite signs), whether reciprocity holds or not. The price elasticity of demand therefore increases when user interaction is significant.

[7] There is a broad corpus of literature on pricing and regulatory issues in the communications industries – especially on the subject of access (Armstrong, 1998; Baumol et al., 1997; Cave, 2006; Carter and Wright, 2003; DeGraba, 2000; Noam, 1994; Wright, 2002). Armstrong (2002) and Vogelsang (2003) provide excellent surveys

The next section reviews the relevant literature (empirical first and then theoretical). Section 3 introduces a simple model of telecommunications demand where consumers benefit from both sent and received messages; while the data and empirical strategy are discussed in section 4. The choice of econometric specification and results are discussed in section 5 and section 6 concludes and outlays further avenues of research.

2 Related literature

Our empirical study of SMS demand needs to be set into the general context of the point-to-point telephony literature, which focuses on voice telephony. Several works, analysing both long-distance national and international fixed telephony, have modelled and measured reciprocal calling patterns – defined by the impact of received calls on demand for (placed) calls. These studies aimed to refine estimates of price (and to a lesser extent income) demand elasticities, with respect to different types of telephony. The introduction of point-to-point models allowed to define demand by focusing on specific traffic routes. Having set one traffic direction (i.e. from country A to Country B) as the regressand, these empirical models then included return traffic (from B to A) amongst the explanatory variables.

The seminal work by Larson et al. (1990) compares estimates of point-to-point long distance US demand obtained both including and omitting return traffic as an explanatory variable. They find that a standard model specification yields slightly inelastic demand while the inclusion of return traffic allows to isolate both impact (short-run) and total (long-run) price effects. These effects show that demand is more inelastic than the conventional specification would suggest (though the total elasticity is not markedly different). According to their model, the gap between impact and total elasticities depends on the flow-through effect: a price change directly affects outgoing traffic, which in its turn affects incoming traffic and then outgoing traffic and so on – accruing to a full multiplier effect. Similar dynamics apply as well to responses to non-price explanatory variables (such as income). Larson et al. (1990) go as far as calling "induced" demand the total impact of return traffic, which will go side by side with the "autonomous" component, i.e. the level of outgoing traffic not explained by incoming traffic. Moreover, they stress that while estimates based on aggregate data can only reflect dominant patterns, disaggregated telecommunications traffic will be composed of different individual patterns of communication

and discussions.

amongst pairs of users. They theorize two extreme calling patterns depending on how information is produced: the case where received and placed calls are complementary ("reciprocal calling") and the case where they are substitutes ("information content").

Appelbe et al. (1988) study call demand between Canadian provinces and US macro-regions in a point-to-point model. Their work identifies significant complementarities between sent and received calls, while controlling for seasonality and factors such as postal strikes, income and market size. As shown in a follow-up paper (Appelbe et al., 1992), their point-to-point estimates, while relatively similar to those obtained by a standard time-series model, have a markedly stronger predictive power when tested against a posterior dataset. Both studies provides elasticity estimates for one direction of traffic only (Canadian provinces to US regions), due to the lack of suitable US pricing data. They work around this by imposing a symmetry constraint on traffic across both directions. They therefore assume that the extent to which consumers interact / return calls is the same in US to Canada traffic as the other way round. This is a common feature of the point-to-point literature (including this work), since no other study apart from Larson et al. (1990) provides estimates for each directions of traffic, given data constraints (Einhorn, 2002).

Further studies analyse international telephony demand in a point-to-point framework, while controlling for the socio-economic ties between their reference country (i.e. US and Spain) and each terminus of the international traffic routes modelled.[8] Acton and Vogelsang (1992), who do not include reverse-direction traffic among the explanatory variables, focus on price elasticities. They find evidence of negative own-price effects, while cross-price effects are statistically insignificant. On the other hand, Garin-Munoz and Perez-Amaral (1998; 1999), find own-price elastic demand and complementarities between sent and received calls (i.e. a positive demand coefficient for the incoming traffic regressor), as well as a significant demand impact of touristic flows. A point-to-point definition of demand for international telephony is not only a feature of academic studies. In fact, Ofcom's latest market review of wholesale international services (Office of Communications, 2006b) identifies economic markets on a route by route basis. This methodology had been used already by its predecessor organization (Office of Telecommunications, 2003) and is currently adopted by the European Commission's market analysis guidelines (European Commission, 2002).

[8] The routes analysed are of the following kind: US-Europe, Spain-Africa/Asia and Spain-Europe.

Return traffic flows have also been analysed in the different context of the telecommunications economics literature on callback, which addresses two phenomena. First, user-driven callback follows customer efforts to internalise call externalities.[9] Second, operator-driven callback traffic is associated to the action of specialised callback carriers / resellers which provide international telecommunications by reversing / rerouting calls in order to benefit from arbitrage opportunities (historically lying in the international settlements regime).[10] The post-liberalisation increase of competition on international telephony reduced the scope for arbitrage on fixed telecommunications and the interest of telecommunications economists on operator-driven callback.[11] Where the point-to-point literature differs from the callback literature is in the rationale for alternate traffic. In callback, two correspondents reverse the direction of a call only for (joint) cost saving reasons. This is possible because of the two-way nature of voice communications. On the other hand, two users may want to alternate the origin of the information exchange in order to compose a better information picture – the focus of point-to-point analyses. The latter can happen even where there are sufficient cost saving incentives for user-driven callback. In any two-way means of communication, this poses a theoretical challenge to the identification of which effect

[9] As shown by Littlechild (1975), this relates to formal or informal agreements by consumers aimed at minimising joint costs of communications by making the party with the cheaper telephony service place most of the calls. This may involve some form of collateral compensation to the designated calling party.

[10] Throughout the Nineties, as liberalization policies reshaped industry structures across many national markets, the international settlement rates historically agreed by the monopolistic PTOs became the subject of an intense debate. As a consequence, many studies of fixed international telephony demand analysed the impact on traffic flows of retail and (international) access pricing (Acton and Vogelsang, 1992; Hackl and Westlund, 1995; Garin-Munoz and Perez-Amaral, 1996; Sandbach, 1996). Einhorn (2002) provides an extensive review of this literature. Callback patterns were identified and estimated across different international routes. The entry of callback carriers eventually led to the erosion of market power by PTOs on international telephony, contributing to a fall in prices and thus a reduction in the magnitude of callback opportunities on many competitive routes. A discussion of these trends in the instructive case of Hong Kong is in Lam (1997).

[11] However, popular brands – such as RebTel and SkypeMobile – are currently introducing mobile-based voice offers relying on callback – still an attractive business model given the large price-cost differential in international mobile telephony. The current arbitrage driver is the transition from PSTN to next-generation IP networks and their interconnection models. "[Calling-party-pays] assumes that the party that placed the call is the sole beneficiary of the call, that he or she should bear the total retail costs of the call, and that his or her network should consequently bear the total wholesale cost of the call. The rationality of this assumption is dubious in the current network, and even more so in an IP-based world where the direction of the call can trivially be reversed" (WIK-Consult, 2008, p.XI).

prevails – as acknowledged by Acton and Vogelsang (1992, p. 322). In text messaging, a one-way form of communication, the nature of the information exchange will drive what traffic patterns are established (e.g. depending on whether the reciprocity scenario holds or not), since callback phenomena cannot play a significant role by definition.

Finally, recent empirical studies of mobile telecommunications demand address issues of intra-modal competition.[12] Andersson et al. (2006) present a theoretical model of consumer choice between text and voice which accounts for the impact of incoming traffic. Their reduced form model estimation of demand by the customers of a Norwegian mobile operator (1998–2004) finds that calls are a substitute for text messages for small network sizes, and a complement otherwise; thus concluding that network size fundamentally shapes SMS demand. On the other hand, Grzybowski and Pereira (2008) develop and test a structural model (which does not consider incoming communications), finding mobile calls and texts to be complements. They estimate a Tobit model for panel data with individual random effects, using individual bill-level data from a sample of customers of Portuguese operators (2003–2004). Kim et al. (2008) provide an alternative structural approach by means of a two-stage discrete/continuous mixture choice model. This reflects the empirical context modelled, since the study analyses consumer-level data of subscribers to a single network which offers a set of price plans with voice-only allowances (and a constant sms price).[13] They find that accounting for heterogeneity in users' voice and SMS satiation levels – which they stress is a critical source of endogeneity – texts and calls are substitutes. Both the intra-modal substitutability result and the extremely inelastic demand estimates provided reflect the market elasticity of the set of consumers in the dataset, taking their subscription choice as given.

[12] The publicly available literature in this field is somewhat limited by data availability constraints. Taylor (2004, p.130) points out that mobile telecommunications firms are keen to consider as a proprietary asset any information on the extent to which mobile communications may be a substitute or complement to other means of communication, due to the commercial sensitivity of such data.

[13] This study analyses 9 monthly observations of a subset (<10,000) of the consumers of an undisclosed mobile service provider, described as the third operator in an Asian country (in 2002), then totalling around 2 million customers (Kim et al., 2008, p. 5).

2.1 The demand modelling debate in the theory of networks competition

The industrial organisation theoretical literature on network competition has recently produced conflicting views on demand modelling. Seminal works such as Armstrong (1998) and Laffont et al. (1998) introduce the simplifying assumption that users derive utility from communicating only in proportion to the length of calls placed. They further assume that those who are called will always accept and receive an indefinite amount of calls at no cost, although they do not explicitly benefit from it.

Following works have extended this framework by considering that users benefit from received communications too (and may be charged for it), which allow them to refine the analysis of call externalities (Kim and Lim, 2001; DeGraba, 2003; Jeon et al., 2004). This raises the question of whether users benefit in the same way from calls placed or received. And, if that is not the case, of which trade-off between sent and received calls underlies the consumers' utility function – and ultimately the demand for communications services. In particular, the above studies assume that users' demands for placing and receiving calls are independent from each other – the separability assumption. For instance, according to the model by Kim and Lim (2001), a user placing n calls and receiving m calls would gain utility equal to $u(n) + v(m)$, where $v(.) = \theta u(.)$. The empirical implication of this are the following. If demand for communications services follows the separability assumption, then regressing the amount of calls (or texts) placed by a consumer over the number of incoming calls (texts) will yield a coefficient estimate of zero. This would imply a nil estimate for the impact of incoming traffic when measuring aggregate patterns, which will be tested in this study. If this is not the case on aggregate, then separable utility cannot be an appropriate way to model individual consumers' demand in general. This is suggested by the evidence of a positive impact of incoming traffic on demand in several disparate markets, as found in the point-to-point demand literature discussed above.

Hermalin and Katz (2004) analyse information exchange and call externalities in a single-network scenario. They model utility as a function of each consumer's benefit from the information exchanged (in a message sent or received) and analyse the conditions for the exchange to take place. Users exchange messages sequentially and thus they may strategically delay communications in the expectation of leaving the costs to their counterparts. The equilibrium outcome will depend on the distribution of message exchange values. This leads Hermalin and Katz (2004, p. 444) to stress that the restrictive assumptions made in previous models of network competition about the values of exchanging calls could be determinant in shaping the equilibrium outcome.

Cambini and Valletti (2008) model the information exchange in a network competition framework, while abstracting from the specific sequence of interaction. They focus instead on the aggregate relation between the amount of outgoing and incoming information messages (e.g. calls, texts, etc.) exchanged by users. The cumulative impact on received messages of a user placing one extra message is defined as "propagation" factor: this could be either negative, nil (if the separability holds) or positive – if users of telecommunications services perceive sent and received messages as complements (in particular equal to 1 in the case of perfect complementarity). Cambini and Valletti (2008) conclude that if sent and received messages are complements (substitutes), networks' pricing incentives will be such that the interconnectivity breakdown result (Jeon et al., 2004) is substantially mitigated (exacerbated). Moreover, where networks can set the level of reception tariffs and negotiate over a reciprocal access charge, complementarity implies that the interconnectivity breakdown will never take place.[14] Competing operators will then have an incentive to deliver outgoing off-net calls in the expectation of receiving calls from other networks – thus to some extent internalizing the externality between callers subscribed to different networks.

3 The model

Our strategy is to trade off a structural analysis in favour of a sharper focus on the impact of received communication – which we capture by selecting exclusively one means by which users exchange information, i.e. SMS. This trade-off is evident in structural analyses of intra-modal demand such as Andersson et al. (2006), which include incoming communications in their analytical model although excluding this factor from their empirical analysis (which uses network size as the only proxy for user interaction). Thus, we choose to set aside the issue of voice/text intra-modal usage, which is best analysed using a structural approach – as undertaken by works such as Grzybowski and Pereira (2008). Instead, similarly to Kim et al. (2008), we opt to restrict our simple model by assuming that a consumer's SMS usage decision is independent from the

[14] Provided that the value that users give to a received call is in a range close to the value of placed calls. This range will be greater the higher the value of the propagation factor and will tend to infinity in the case of perfect complementarity.

consumption of mobile calls and the choice of operator – which is consistent with the empirical aim of our work.[15]

Two mobile networks – denoted as j, k – have given sets of subscribers, totalling n_j, n_k users.[16] A subscriber to network j derives utility either from the amount of information obtained (I) or from a composite good (X). The former on its turn depends on the amount of messages she sends or receives. Specifically, we will focus exclusively on messages exchanged cross-network, i.e. with subscribers to the other operator. The consumer's optimisation problem is the following:

$$\begin{aligned} & \max U_j\left(I_j, X_j\right) \\ \text{s.t.} \quad I_j &= f\left(q_j, q_k\right) \\ M_j &= X_j + p_j q_j, \end{aligned}$$

where:

U: the consumer utility

X: the quantity of outside good, the price of which is normalised to 1

M: the income

I_j: the amount of information exchanged – constructed according to the function f – which yields benefit to user j

q_j, q_k: the amount of cross-network text messages sent by consumer j (or k)

p_j, p_k: the text message price faced by consumer j (or k)

Optimality requires that the MRS between the composite good and outgoing text messages is equal to the ratio of their prices:

$$\frac{\partial U_j}{\partial X_j} / \left(\frac{\partial U_j}{\partial I_j}\frac{\partial f}{\partial q_j}\right) = \frac{1}{p_j}$$

Anecdotal evidence suggests that consumers spend a relatively small portion of their income to mobile telephony and only a minor part of that on SMS. Thus, it is reasonable to set aside income

[15] A more detailed discussion of our empirical strategy is in the following section.

[16] For the sake of simplicity we will assume that no consumers subscribe to both networks. Anecdotal evidence suggests that in practice some consumers may own and switch between multiple SIM cards or have several handsets simultaneously on. A most noticeable consequence of this is that in some countries (such as Italy) mobile subscription figures exceed the total population count (which includes all age groups).

effects by assuming that consumers have utility functions quasi-linear in the outside good, which implies that user i's utility will be $I + X$. Therefore the optimality condition becomes:

$$\frac{\partial f}{\partial q_j} = p_j$$

The above implies that the equilibrium amount of text messaging demanded, q_j, will be a function of the outgoing SMS price, quantity (and price) of the composite good, income, as well as the amount of cross-network text messages received. Therefore:

$$q_j^* = q_j^* \left(p_j, q_k^*\right)$$

Similarly so for a subscriber to the alternative network (k), whose demand for off-net SMS will be:

$$q_k^* = q_k^* \left(p_k, q_j^*\right)$$

Let us assume an isotropic or balanced calling (or texting, in our case) pattern – which is in line with the literature on networks interconnection (Armstrong, 1998; Laffont et al., 1998a; Laffont et al., 1998b).[17] If each user on network j is sending $q_j^*(.)$ text messages to each user on network k, then the total amount of SMS traffic flowing from network j to network k will be:

$$Q_{jk}^* = n_j n_k q_j^*$$

This leads to the equations below, which we will use to inform our empirical estimation:

$$\log Q_{jk}^* = \log\left(n_j n_k q_j^*(.)\right) = \log q_j^*(p_j, q_k^*) + \log\left(n_j n_k\right) \quad \text{(1a)}$$

$$\log Q_{kj}^* = \log\left(n_j n_k q_k^*(.)\right) = \log q_k^*(p_k, q_j^*) + \log\left(n_j n_k\right) \quad \text{(1b)}$$

4 Empirical strategy and data

Many commentators in the mobile industry have stressed that SMS services emerged as an unexpected add-on to the provision of calls.[18] In fact, customer mobile expenditure, which in

[17] Dessein (2004) presents an interesting model where heterogeneity in customer type (i.e. business vs. residential) affects how users perceive the substitutability of competing networks and thus influences operators retail and access pricing. In mobile telephony, though, it is unlikely that a stark distinction between work and leisure use of subscriber line occurs. Anecdotal evidence suggests that frequently business and private use can coexist on the same account / line, eased by the nature of mobile communications. Whether this phenomenon takes place or not is not a concern for our empirical analysis, due to the aggregated nature of our data, which is discussed in the next section.

[18] For instance, Ofcom (2006)(Office of Communications, 2006a, p.18) – as mentioned above at p.2.

the early years of mobile was entirely due to call traffic, is still prevalently associated to voice (around or above 80%, the remainder being messaging and other data services). Since voice is the leading factor in the take up of mobile telephony, we can assume that consumers' decision to subscribe to a network is exogenous to the amount of SMS exchanged. Moreover, the operators' drive to increase mobile penetration and then to compete for custom in saturated markets has led to an extreme proliferation of tariffs for voice usage. These are amongst the key parametres in users' subscription decision, together with coverage/call quality and choice of handset (where the latter is subsidised by the operator). Users can face call prices which vary as a function of time-of day and whether the recipient belongs to the same network. This adds complexity to any demand estimation considering voice telephony, which has led some studies to opt for simplification by focusing only on users facing constant call prices. On the contrary, SMS pricing is sensibly less exposed to tariff proliferation. This benefits our analysis of mobile usage, which we have chosen to focus exclusively on SMS traffic, while setting aside the issue of intra-modal competition.

This study uses a novel pooled time series / cross country dataset relative to 15 countries, made available to us by Vodafone Group. This set of countries, where Vodafone operates a wholly-owned subsidiary, is listed in Table 1 below. Notice that all countries present a CPP regime.[19]

[Insert Table 1 here]

For each of these countries we observe aggregated data for the period April 2002 – February 2007 (59 months, totalling 777 observations after dropping some incomplete data). These consist of the monthly number of sent and received SMS, the price of sent messages, firms' market shares and total country subscription figures.

The messaging usage data we examine is of the following type. For each mobile user, the network operator collated all messages sent (received) in a month, irrespective of recipient (sender)

[19] The Caller-Pays-Principle (or Calling-Party-Pays) applies for instance in European markets, as subscribers receiving a text message do not face a charge (apart while roaming). In countries such as the US instead, received SMS are billed. For instance a Verizon standard price plan includes a $0.20 charge per sent or received text. Similarly to incoming call minutes, these may be included as part of bundles / allowances. See: http://support.vzw.com/terms/txt_messaging.html (last accessed: November 2008). For an introduction to the rationales for Caller-Pays-Principle vs. Receiver-Pays-Principle see for instance Vogelsang (2003, pp. 850–852) .

– hence summing communications between different user pairs.[20] For each country, we observe network-wide aggregates (both for outgoing and incoming traffic).[21]

The pricing data made available to us reflects the average price of a text message charged to Vodafone subscribers. This includes both pay-as-you-go (i.e. pre-pay) and pay-monthly (post-pay) consumers.[22] While the pricing indicator for users on bundle deals may suffer from some endogeneity, it allows us to include in our analysis not only consumers facing a constant SMS price – as in Grzybowski and Pereira (2008) – but also those on non-linear pricing plans, which are increasingly popular.[23]

Furthermore, in order to control for the size of the network, we construct a variable which reflects the size of each point-to-point route (i.e. the cross-network messaging traffic exchanged within one country). Following Larson et al. (1990), we define a variable ($COM = n_j n_k$) as the product of the number of Vodafone subscribers times the subscribers not on the Vodafone

[20] Thus our data does not allow us to differentiate between business and private usage.

[21] Up to March 2005, messaging counts include multimedia messages (MMS) traffic, where this service was available (anecdotal evidence suggests very low MMS usage pre-2005). The dummy variable $DMMS$ takes the value of one before April 2005. We use $DMMS$ to control for unobserved MMS traffic, as discussed in greater detail in the next section.

[22] The latter often face non-linear tariffs. Where customers are on a price plan including bundles of call minutes and text messages then Vodafone applies the following weighted average procedure to compute a linear equivalent for prices. For example, let us consider a consumer spending a flat fee of 30€ per month on a bundle offer, which includes up to 500 minutes and 200 text messages monthly and which charges any consumption beyond the inclusive allowance at the following rates: 0.20€/call and 0.10€/SMS. Let us say that this user places 300 minutes of calls and 100 SMS in a month. The algorithm computes the expenditure at current usage levels as if subject to the marginal rates for extra-allowance calls and texts (in our case, respectively: 60€ + 10€ = 70€, which implies a 86% / 14% expenditure split). The flat fee is then allocated according to this split (in our case: 86% of 30€ = 25.71€ for calls and 14% of 30€ = 4.28€ for SMS). This is then divided by the actual call minutes and SMS sent, in order to obtain the imputed price (0.08€ per minute and 0.04€ per SMS). Therefore, for consumers subject to non-linear tariffs, the pricing data that we observe is a weighted average, built so to allocate the flat fee according to text/voice usage and the prices of consuming extra text/calls beyond the allowance thresholds (which proxy the marginal tariffs).

[23] Moreover, subscribers to price plans with bundles are likely to differ in their usage patterns from those who do not. Since most bundle deals aim to exert second degree price discrimination, those taking up offers of such kind are most likely to be those who wish to consume more. Any analysis which excludes customers on certain types of tariffs is thus exposed to selection bias, while this is not an issue in this study, given the nature of the data used.

network. This indicator will thus reflect the total number of potential interaction between the two groups at both ends of each traffic route considered.[24] We have also tried including alternative measures of network size, such as the number of subscribers, which yield comparable results. Nonetheless, as equations 1a and 1b show, our simple analytic model predicts that cross-network traffic will depend on the product of subscribers across networks, which is the network size variable which we choose to include.[25]

Moreover, while we have a direct measure of SMS received by all Vodafone customers from users on other networks, we derive the number of SMS sent by Vodafone subscribers to all other networks by assuming an isotropic texting pattern. Descriptive statistics of the data used are presented in Table 2 below.

[Insert Table 2 here]

The nature of the data enables us to focus on aggregate relationships, by taking a point-to-point approach (Rea and Lage, 1978). Point-to-point models have been applied to estimate demand on a route by route basis. Specifically, past studies used traffic data aggregated at different levels, i.e. for: cities/suburbs, to measure intercity traffic (Larson et al., 1990); macro-regions (e.g. dividing Canada in 6 regions, the US in 4), measuring national (long-distance) and international[26] telephony (Appelbe et al., 1988); countries, focusing on international telephony (Acton and Vogelsang, 1992; Garin-Munoz and Perez-Amaral, 1998; Garin-Munoz and Perez-Amaral, 1999). In this study, the unit of analysis is the set of subscribers to a mobile network. We will analyse cross-network text message exchanges and we will apply the concept of route by aggregating SMS usage along the subscription dimension. Since for each of the countries considered we only observe the aggregate outgoing and incoming data generated by Vodafone

[24] Many commentators have highlighted the role of close user groups: consumers who communicate with each other frequently and that strategically subscribe to the same network. This can be the case where networks apply network-based price discrimination (Laffont et al., 1998b) as well as where close user groups can benefit from "family and friends"–type tariffs (the first of which was introduced by MCI). As the COM variable reflects only potential interactions amongst users on different networks (Vodafone vs. non-Vodafone, alike all variables considered in this study), text usage associated to close user group behaviour does not affect our analysis.

[25] A common shortcoming of the use of subscription data is that figures are in practice likely to be inflated because of multiple SIM ownership. See footnote 16 at p.11.

[26] This could count as a quasi-national long distance communications setting as US and Canada share a legacy of an integrated Bell system. For instance they have the same country code – which means that users do not have to compose it to call from US to Canada and conversely – unlike standard international telecommunications.

customers, we can measure total traffic between the Vodafone network and all other operators, where the latter are taken as a whole. Hence our estimates reflect demand patterns established on each of the 15 cross-network routes thus defined, one per country. Thus, for each route, we conceive of, on one end, all of Vodafone's subscribers and, on the other, all other operators' subscribers.

The level of aggregation of the data observed allows us to study the dominant patterns of information exchange, assuming no consumer heterogeneity. In practice, every mobile user can exchange information in a different way, although a mobile operator's pricing strategy is informed to a good extent by the average user behaviour, which will be our focus.[27] Moreover, the panel nature of the dataset presents the obvious advantage of screening for country-specific effects, which is an improvement both over studies based on single country time series data (Andersson et al., 2006). Notwithstanding that, we are aware that the cross-sectional dimension of the dataset is limited (compared for instance to household survey panels), which we will take into account when evaluating different model specifications – discussed in the next section.[28]

5 Econometric specification and results

Our empirical model descends from the theoretical framework presented in the previous section. Based on eqs. 1a and 1b, we build the following system of simultaneous double-logarithmic equations displayed below. Notice that now we replace the subscripts j, k with "v" (Vodafone users) and "$v-$" (non-Vodafone users) for the two networks which define – for each country i in month t – the traffic route analysed. While we represent a fixed effect specification, we have tested the robustness of our results across a range of different specifications, as discussed below

[27] Heterogeneity in how consumers exchange information can nonetheless be pivotal to some telecommunications business models. Over time many companies have deployed strategies clearly aimed at attracting users that are skewed towards receiving communications – in order to exploit above cost termination rates. For instance, in the Nineties, several new entrants to US markets (CLECs) have targeted dial-up users. More recently, many EU mobile operators have offered rebates to consumers as a function of received traffic (such as Three in the UK). Because of this dynamics, high reciprocal access charges set by incumbents have worked against them and ultimately been reduced. An example of this intyerplay was Tele2's strategy against Telia in Sweden (Hultkrantz, 2002, p. 147).

[28] Our prior is to expect the width of the dataset to be potentially affecting the asymptotic properties of panel regression estimators.

(all results are included in Table 3).

$$\log Q_v^{it} = \alpha_i + \beta_1 \log p_v^{it} + \beta_2 \log q_{v-}^{it} + \beta_3 \log COM^{it} + u_v^{it} \tag{2a}$$

$$\log Q_{v-}^{it} = \delta_i + \gamma_1 \log p_{v-}^{it} + \gamma_2 \log q_v^{it} + \gamma_3 \log COM^{it} + u_{v-}^{it} \tag{2b}$$

The variables above are:

Q_v^{it} The number of text messages sent from Vodafone customers to other networks

α_i, δ_i The unobservable country fixed effects

p_v^{it} The price charged by Vodafone for sending a text

q_{v-}^{it} The number of text messages per user received by Vodafone customers from other networks

$COM^{it} = n_v^{it} n_{v-}^{it}$ The total potential cross-network connections between pairs of users (Vodafone vs. non-Vodafone)

p_{v-}^{it} The price charged by all other networks for sending a text

u_v^{it}, u_{v-}^{it} The error terms, assumed to have the standard statistical properties, i.e. u^{it} is i.i.d. and $u^{it} \sim (0, \sigma)$

Since the information on the price of SMS originated on non-Vodafone networks (p_{v-}) is unavailable to us, we will focus on the estimation of the first equation. This method to overcome data constraints is in line with the point-to-point demand literature, as discussed at p. 6. Following Garin-Munoz and Perez-Amaral (1998; 1999), we will address the potential endogeneity of incoming traffic by using as an instrument the one-month lagged value of incoming messages.[29] Moreover, we will perform specification robustness checks by also including the value of the Herfindahl-Hirschman index of the mobile network operators active in each country (HHi) as an instrument for the text price.[30]

We can assess the total effect of a variation in the regressors by using the reduced form

[29] Performing a Hausman test on the specification with and without the instrument yields the following statistics: chi2(3) = 34.37; Prob>chi2 = 0.0000, supporting the validity of this instrument.

[30] We have also checked whether to use the number of mobile operators per country ($NOps$) to instrument text price. Since this variable is highly correlated with HHi (corr = -0.94), both are not jointly significant instruments. Although $NOps$ is a valid instrument on its own, we settle for HHi since F-test statistics indicate that, when taken on their own, the latter is the strongest instrument of the two.

model, which after substituting the simultaneous equations 2a and 2b is:

$$
\begin{aligned}
\log Q_v^{it} &= \alpha_i + \beta_1 \log p_v^{it} + \beta_2 \log\left(\frac{Q_{v-}^{it}}{n_{v-}^{it}}\right) + \beta_3 \log COM^{it} + u_v^{it}, \text{ i.e.} \\
\log Q_v^{it} &= \alpha_i + \beta_1 \log p_v^{it} + \beta_2 \log\left(\delta_i + \gamma_1 \log p_{v-}^{it} + \gamma_2 \log q_v^{it} + \gamma_3 \log COM^{it} + u_{v-}^{it}\right) + \\
&\quad -\beta_2 \log n_{v-}^{it} + \beta_3 \log COM^{it} + u_v^{it}
\end{aligned}
$$

The total (long-run) price elasticity will thus be:

$$\frac{\partial \log Q_v^{it}}{\partial \log p_v^{it}} = \frac{\beta_1}{1-\beta_2\gamma_2}$$

The lack of a consistent price indicator for other networks' text messages prevents us from testing whether the reciprocal texting effect works differently for Vodafone customers and other operators' ones. If we assume symmetry ($\beta_2 = \gamma_2$), the long-run price elasticity will be equal to:

$$\frac{\beta_1}{1-(\beta_2)^2} \tag{3}$$

Because of the log-log nature of the model, β_1 and β_2 represent impact elasticities and only reflect the first effect on the amount of sent messages of variations in their price and in the level of incoming messages. Notice that if incoming traffic affects demand significantly ($\beta_2 \neq 0$), whether positively or negatively, it will increase its long-run ("multiplier") elasticity. In fact, if sent and received messages are complements ($\beta_2 > 0$) then a price reduction by Vodafone will make its users send more off-net SMS, which will generate more incoming texts, on its turn further raising demand for text messages. If instead sent and received messages are substitutes ($\beta_2 < 0$), the very same price cut will still make its users send more off-net SMS, which will this time generate less incoming texts, on its turn again further raising demand for text messages.

Table 3 below presents the results of our estimation. All of the following are instrumental variable specifications including the lagged value of incoming messages, as discussed here.[31] Specification I presents the standard two-stages least-squares estimates, which pool all observations and thus do not exploit the panel nature of the data. A between estimator follows in II, which yields estimates based on the cross-sectional component of the data. As the number of countries is relatively limited, this estimator may not be sufficiently robust. We then move to consider the trade-off between random and fixed effects (specifications III and IV, respectively).

[31] An inspection of first stage results for our chosen specification confirms the validity of the instrument, which yields a statistically significant coefficient of 0.81 (t value = 45.92; P > |t| = 0.000). The first stage F-statistic is F(19,728) = 150.12; Prob > F = 0.0000.

First, a random effects econometric model would require country effects to be uncorrelated with the regressors, which is unwarranted in our case. Second, since our panel has a bigger temporal than cross-sectional dimension, our prior is that a fixed effects specification may be more appropriate than RE. As common practice, we have conducted a Hausman test, which confirms our expectations and supports adopting a specification including fixed effects.[32] These will capture any heterogeneity which may be country specific, as each route is associated to mobile traffic in a different country. Moreover, specification V includes year and month dummies (Y_y, M_m). The introduction of the time dummies controls for the seasonal nature of text demand and any time related unobservables.[33] When added to this specification, $DMMS$ is not a statistically significant control and leaves other coefficients unaltered.[34] In fact, the year and month dummies already in place may be in part controlling also for any residual factor from the inclusion of MMS data in messaging counts before April 2005. As a further check, we test for the effect of $DMMS$ on individual coefficients and find a difference in the price coefficient. We therefore include in our chosen specification VI a simple difference-in-difference estimate, testing the impact on demand of a dummy interaction term, defined as $DMMS \log p_v$. This is retained because of the statistically significant (though minor) impact of the interaction term on demand, although this leaves all other coefficients virtually unchanged when compared to V. The equation we choose to estimate is thus the following:

$$\log Q_v^{it} = \alpha_i + \beta_1 \log p_v^{it} + \beta_2 \log q_{v-}^{it} + \beta_3 \log COM^{it} + \beta_4 DMMS \log p_v^{it} + \sum_{y=2003}^{2007} Y_y + \sum_{m=1}^{11} M_m + u_v^{it} \tag{4}$$

Furthermore, as a robustness check, specifications VII to XII replicate results for specifications

[32]The Hausman test rejects the hypothesis that both RE and FE estimators are consistent – yielding the following statistics: chi2(4) = 15.11; Prob>chi2 = 0.0045.

[33]As a further robustness check, we have also tried the least restrictive specification which includes a complete set of time dummies (58 – one per month, instead of the chosen 18 year and month dummies). Since estimates remain unchanged and the significance of our results is unaffected (given the large number of observations), we choose to keep the year and month dummies specification as it allows the inspection of seasonal patterns of demand. The estimation results relative to the impact of the time controls are presented and briefly commented in the Appendix.

[34]Furthermore, the hypothesis that regressors' coefficients jointly differ before and after the change in reporting is rejected by a Chow test – with statistics: F(4, 724) = 11.42; Prob > F = 0.0000. This reassures us that any impact of the underlying MMS traffic on our analysis is minor.

I to VI respectively, by instrumenting not only q_{v-} (with its lag) but also p_v with HHi.

According to our model, the nature of the information exchange (i.e. whether a user wishes to send more or less texts when receiving SMS) will depend on the sign of:

$$\frac{\partial f}{\partial q_j \partial q_k} \tag{5}$$

If for two users engaged in communications, incoming communications stimulate further outgoing communications (reciprocity scenario), we expect eq. 5 to be positive. If this is a dominant pattern across users (across all information exchanges) then on aggregate we expect to observe a positive sign for β_2. If instead the separability assumption holds (i.e. demand for text messages is independent from received SMS), then on aggregate we expect to observe β_2 to be nil. Alternatively, if receiving SMS reduces – all else equal – a user's wish to send text messages then β_2 should result negative.

5.1 Discussion of findings

[Insert Table 3 here]

We measure a positive effect on SMS demand of received traffic, which holds across all specifications. A 10% increase in the amount of incoming SMS is associated to 1.2% more text messages sent. This is consistent with the hypothesis that sent and received text messages are complements in the information exchange. The impact of our variable reflecting the size of the network / community of interest is such that a 10% increase in the amount of total cross-network links is associated to 3.5% more text messages sent.

In the light of the literature discussed in section 2, a key contribution of our work is the joint identification of the demand impact of both the network size (i.e. community of interest) and incoming traffic. In fact, Larson et al. (1990) find the impact of network size not to be significantly different from zero, leading Taylor (2004) to suggest that the incoming-traffic variable may be interacting with network size and picking up the impact on demand of the size of the community of interest. On the other hand, Andersson et al. (2006) include both network size and incoming communication flows in their analytic model but do not include the latter amongst their explanatory variables in their empirical specification. They find that the size of the network has a strong impact on text usage (affecting whether mobile calls and texts are substitutes or complements). Because incoming traffic is not accounted for, this result may be in part picking up the effect of user interaction. Our study thus shows that both user interaction and network size play a role in shaping demand and that the magnitude of the latter appears to be dominant.

We find demand for text messages to be slightly inelastic as the coefficient on price (which we can interpret as impact elasticity) equals -0.95.[35] Moreover, by correcting the impact elasticity figures for the "multiplier" effect of return traffic flows, we can compute the total (or long run) price elasticity. As shown in eq. 3, for any $\beta_2 \neq 0$, the overall impact of price variations on demand will increase, due to the positive-positive or negative-negative cross-user feedback loop underlying the information exchange. Therefore, we expect the total elasticity to be higher than the impact elasticity. This (small) correction yields a figure for total price elasticity figure of -0.96.

Let us now compare our results with the point-to-point literature. Appelbe et al. (1988) estimate reciprocal calling coefficients of around 0.5 (ranging from 0.24 to 0.72 across route types); Larson et al. (1990) find values of 0.67 / 0.75 (from metropolitan areas to their suburbs and conversely);[36] and for Spanish originated international calls the estimates in Garin-Munoz and Perez-Amaral (1998; 1999) are 0.69 (inter-continental calls) and 0.78 (calls to Europe). Our analysis of SMS demand yields an incoming traffic coefficient of 0.12, which is sensibly below what previously established in the literature. This may be explained by the fact that in our study the effects of both incoming traffic and network size are identified as statistically significant. Instead, previous estimates of the demand impact of incoming traffic incorporate the effect of network size too (Taylor, 1994; Taylor, 2004).

Furthermore, in order to validate the above intutition, we have analysed the implications of our modelling assumptions in the context of the literature. The theoretical framework presented in section 3 is reflected in our empirical specification, which sets total traffic as the dependent variable and per-user traffic amongst the regressors. Previous point-to-point studies instead estimate either per-user outgoing traffic as a function of per-user incoming traffic (Garin-Munoz

[35] Alternatively, if we include HHi as a further instrument, the price coefficient is estimated at -0.76, while all other coefficients remain virtually unchanged.

[36] Larson et al (1990) – the only work estimating both directions of routes – finds relatively similar coefficients on incoming traffic on both directions. Notice that metropolitan users return calls to a slightly higher tune than suburban users. This may reflect a different composition of business / leisure users in metropolitan and suburban areas.

and Perez-Amaral, 1998; Garin-Munoz and Perez-Amaral, 1999) or alternatively use total levels of both outgoing and incoming traffic (Appelbe et al., 1988; Appelbe et al., 1992; Larson et al., 1990). The specifications chosen may explain why previous studies are unable to disentangle user interaction and network size effects. As a thought experiment, let us modify our econometric specification (see eq. 2a) so to use total traffic values for both its usage variables. We estimate the following:

$$\log Q_v^{it} = \psi_i + \theta_1 \log p_v^{it} + \theta_2 \log Q_{v-}^{it} + \theta_3 \log COM^{it} + u_v^{it}, \text{ i.e.} \qquad (6)$$
$$\log Q_v^{it} = \psi_i + \theta_1 \log p_v^{it} + \theta_2 \log q_{v-}^{it} + (\theta_2 + \theta_3) \log COM^{it} - \theta_2 \log n_{v-}^{it} + u_v^{it}$$

We therefore expect θ_1 and θ_2 to be the same as the β_1 and β_2 obtained from the estimation of eq. 2a, while the effect of the network size variable to be confounded. In fact, the coefficient on COM should be higher under our original specification than in estimates obtained where total traffic is a regressor (as in eq. 6 and in the early point-to-point literature). Were it not for other differences in the two specifications, our β_3 should equal $\theta_2 + \theta_3$.[37]

If instead we transform our specification by substituting per-user traffic as the dependent variable we will obtain the following:

$$\log q_v^{it} = \xi_i + \lambda_1 \log p_v^{it} + \lambda_2 \log q_{v-}^{it} + \lambda_3 \log COM^{it} + u_v^{it}, \text{ i.e.} \qquad (7)$$
$$\log Q_v^{it} = \xi_i + \lambda_1 \log p_v^{it} + \lambda_2 \log q_{v-}^{it} + (1 + \lambda_3) \log COM^{it} - \log n_{v-}^{it} + u_v^{it}$$

Thus, λ_1 and λ_2 should be the same as β_1 and β_2, while the coefficient on COM should be higher in our model than where on the left hand side there is per-user traffic (as in eq. 7 and in the later point-to-point literature). Here, specification differences apart, our β_3 should be equal to $1 + \lambda_3$.[38] This analysis therefore suggests that previous point-to-point studies, by the nature of their specification, underestimate the magnitude of the impact of network size. As a further check, we have tested these two alternative specifications on our data (results are included in the Appendix in tables A2 and A3). As predicted, we find both θ_3 and λ_3 to be smaller than

[37] Notice that the specifications in eqs. 6 and 7 differ from ours also due to the inclusion of the n_{v-} term. Where this is part of the dependent variable, the estimation of the impact of any regressor requires the restriction that the elasticity of total traffic to the number of subscribers is one (Garin-Munoz and Perez-Amaral, 1998, p.245). Moreover, under both alternative models, the estimation results may suffer from the collinearity between n_{v-} and COM.

[38] See the previous footnote.

β_3 (note that λ_3 is even negative) and that the coefficients on price and incoming traffic remain virtually unchanged, which we interpret as a sign of the robustness of our results.

We can also suggest a further interpretation of why we find a lower demand impact of incoming traffic on SMS demand than what established in the literature. In fact, the different nature of the telecommunications technology (mobile vs. fixed) may be associated to different patterns of information exchange. Moreover, SMS present specificities when compared to any voice communications, leading to different externality dynamics. Call (or message) externalities occur where users on the receiving end do not face usage-based charges for the communication, while deriving benefit from it. This is the case in CPP regimes (established in all the countries in our dataset). In fact, unlike voice communications, which require two users to be simultaneously on the phone, SMS are exchanged sequentially and are not only initiated but also delivered to the recipient upon the decision of the sender only. Therefore, while somebody at the receiving end of a phone call can always hang up if the call yields her no (or negative) benefit – thus constraining caller demand – this cannot be the case in the context of SMS exchanges.[39] This factor, while making the identification of text messaging demand less problematic than voice demand, could pose a limit in a straightforward comparison of estimate results between any one-way and two-way means of communication.[40]

6 Conclusions

This work is to the best of our knowledge the first in estimating a point-to-point demand model of mobile telecommunications demand, specifically demand for text messages. We leverage the cross-country time series nature of our dataset, which allows us to identify aggregate patterns

[39] Provided that networks message conveyance and storage function well, delivery will occur as soon as the recipient is on the network. Of course, the recipient may wish to discard the message before reading it; although depending on the mobile phone software she will have to open the message or at least be exposed to the notification / sender ID – which can be seen as a limit case of information exchange.

[40] Notice that in theory, in some cases there may be negative text message externalities. On the other hand, the opportunity cost of time dedicated to receiving communications – which can be modelled as an exogenous source of noise and identify receivers "demand" (Jeon et al., 2004; Cambini and Valletti, 2008) – is in the case of SMS most likely determined by the receiver herself, in her choice of a convenient moment to read a message. If that is the case then in practice, an "inconvenient truth" will always be read at the most convenient time; thus message externalities will always be non-negative, alike call externalities.

of SMS consumption and information exchange, controlling for country fixed effects. We have found evidence of a statistically significant, although modest, impact of incoming traffic on mobile text messaging demand. If we consider the estimates obtained with both random and fixed effects specifications, this is equivalent to demand elasticity values in the range 0.10 – 0.15. This is a smaller effect than what established in previous empirical demand analyses based on point-to-point models, so far applied to study international and long distance fixed telephony. On the other hand, we find that the joint demand impact of incoming SMS and network size (i.e. $\beta_2 + \beta_3$) is around 0.5 – 0.6. This is within the range of the magnitude of the demand impact which the previous point-to-point literature – which did not disentangle those two factors – associates to incoming communications only ("one-half to two-thirds of a call in return" (Taylor, 2004, p.129)).

The empirical evidence that incoming communications positively affect demand suggests a reappraisal of contrasting takes on demand modelling in the IO network competition literature. Our findings provide an empirical rejection of the separability assumption (Kim and Lim, 2001; DeGraba, 2003; Jeon et al., 2004): the analysis of aggregate traffic shows that the level of incoming messages significantly affects demand for sent SMS. The evidence presented – in line with previous point-to-point studies – is consistent with models that account for the interdependency of sent and received communications in demand of telecommunications services (Hermalin and Katz, 2004; Cambini and Valletti, 2008). If we think of consumers sending and receiving messages as part of an exchange aimed at creating a stock of beneficial information – i.e. as part of an information "production function" – our evidence implies that mobile users perceive sent and received messages as complements. The exchanges of information by individual mobile users, which underly the aggregate trends observed in our data, are therefore likely to be along the lines of the reciprocity scenario (Larson et al., 1990).

Moreover, our finding of a positive coefficient associated to incoming SMS traffic yields clear implications for mobile operators' pricing policies. An operator unaware of return messaging flows would consider a lower elasticity value and price too highly accordingly. In fact, the point-to-point demand model estimation yields a larger total price elasticity, which should lead to a lower optimal price. This is because of the complementarity between outgoing and incoming communications. A price cut will stimulate outgoing cross-network messages, on its turn generating return messages, which will then stimulate further outgoing messages via a flow-through effect and so on. Accounting for the demand impact of received text messages, we obtain a long run price elasticity of -0.96. Note that a firm trying to maximize revenues will price its services so to obtain a market outcome denoted by unit–elastic demand, if possible. A stylised description of the mobile industry includes high sunk costs and a marginal cost of sending an SMS almost close to zero.[41] In this scenario, the finding of near unit-price elasticity of demand can be interpreted as evidence of profit maximizing firm behaviour.[42]

Further empirical research may address two issues related to this work but beyond its scope. First, our estimates are bound to the monthly nature of the dataset used. Different intervals of data aggregation (weekly, daily, hourly, if not shorter – given the nature of text messaging exchanges) are associated to different definitions of the information exchange process, which could lead to different point-to-point demand estimates. The empirical challenge associated to the point-to-point method is to estimate a continuous process (the exchange of information between users via SMS), based on data aggregated on a discrete interval. Further research could assess to which extent the time dimension of aggregation in the dataset affects the demand patterns observed. Second, our focus on the exchange of information via SMS informs our finding that the demand for sent messages depends on received messages. Anecdotal evidence suggests that one user may send a text message and this may elicit a return call, for instance. Further studies could extend a point-to-point model by taking into account that consumers exchange information simultaneously via several means of communications (mobile voice, fixed telephony, web-based instant messaging, email). Modelling intra-modal or inter-modal communications options, while accounting for the role of received communication, could extend our understanding of the dynamics of information exchange, were sufficient data to be available.

---

[41] In mature mobile markets, we can reasonably assume that investments in extra infrastructural capacity for the purpose of SMS conveyance are not needed. In practice, network maintenance and the pursuit of quality improvements (e.g. speed of SMS delivery at congested times / locations) may require some investment.

[42] In practice, networks pay termination charges to send text messages across networks. The marginal cost effectively faced by the network thus needs to include this charge, which will be reflected in the pricing. While cross-sectional data on termination fees has been made available to us, its lack of variability over time means that it can not be a useful input supporting a more sophisticated identification strategy. An example of an empirical strategy leveraging termination charges data is in Genakos and Valletti (2008).

A Appendix

Table A1 below presents the estimation results for the time dummies – in our chosen specification VI.

[Insert Table A1 here]

Almost all controls are statistically significant. The most relevant exceptions are those for the Year 2007 – where there are only two months worth of observations – and the month of January. Since no monthly dummies have positive impact, the month of December (our baseline) yields the highest impact on the total number of texts sent. Higher SMS usage is commonly associated to Christmas and New Year. Since the period between December 31st and January 1st itself is the busiest for SMS traffic, this may explain why the January dummy is not statistically different from the December baseline. For instance, "O2 has announced that 166 million messages were sent over its network in a 24-hour period ending at 7.30 am on New Year's Day 2009".[43]

Tables A2 and A3 below present the results from the estimation of eqs. 6 and 7, as discussed at p. 22. In order to ensure the comparability of the results, the regressions use a specification identical in all but the traffic variables to our chosen specification VI.

[Insert Table A2 here]

[Insert Table A3 here]

References

Acton, J. P. and Vogelsang, I. (1992). Telephone demand over the atlantic: Evidence from country-pair data. *Journal of Industrial Economics*, 40(3):305. 0022-1821.

Andersson, K., Foros, O., and Steen, F. (2006). Text and voice: complements, substitutes or both? CEPR Discussion Papers - Industrial Organization, No. 5780.

Appelbe, T. W., Dineen, C. R., Solvason, D. L., and Hsiao, C. (1992). Econometric modelling of canadian long distance calling: A comparison of aggregate time series versus point-to-point panel data approaches. *Empirical Economics*, 17(1):125–140.

[43] See http://www.pocket-lint.co.uk/news/news.phtml/20119/21143/o2-record-number-texts-new-year.phtml.

Appelbe, T. W., Snihur, N. A., Dineen, C., Farnes, D., and Giordano, R. (1988). Point-to-point modelling: An application to canada-canada and canada-united states long distance calling. *Information Economics and Policy*, 3(4):311–331.

Armstrong, M. (1998). Network interconnection in telecommunications. *Economic Journal*, 108(448):545. 0013-0133.

Armstrong, M. (2002). The theory of access pricing and interconnection.

Autorité de Régulation des Communications Électroniques et des Postes (2006a). Arcep notifies the european commission and submits its draft decision on sms call termination for public consultation. france becomes the first country in europe to consider regulation on this wholesale market.

Autorité de Régulation des Communications Électroniques et des Postes (2006b). Décision n. 06-0593 de l'autorité de régulation des communications Électroniques et des postes en date du 27 juillet 2006 portant sur la définition des marchés pertinents de gros de la terminaison d'appel sms sur les réseaux mobiles en métropole, la désignation d'opérateur disposant d'influence significative sur ces marchés et les obligations imposées à ce titre.

Baumol, W., Ordover, J., and Willig, R. (1997). Parity Pricing and Its Critics: A Necessary Condition tor Efficiency in the Provision of Bottleeneck Services to Competitors. *Yale Journal on Regulation*, 16(1):145–164.

Cambini, C. and Valletti, T. M. (2008). Information exchange and competition in communication networks. *Journal of Industrial Economics*, 56(4):707–728.

Carter, M. and Wright, J. (2003). Asymmetric network interconnection. *Review of Industrial Organization*, 22(1):27.

Cave, M. (2006). Encouraging infrastructure competition via the ladder of investment. *Telecommunications Policy*, 30(3-4):223–237.

DeGraba, P. (2000). Bill and keep at central office. Technical report, FCC Office of Plans and Policy working paper.

DeGraba, P. (2003). Efficient intercarrier compensation for competing networks when customers share the value of a call. *Journal of Economics & Management Strategy*, 12(2):207. 1058-6407.

Dessein, W. (2004). Network competition with heterogeneous customers and calling patterns. *Information Economics and Policy*, 16(3):323–345.

Einhorn, M. A. (2002). International telephony: a review of the literature. *Information Economics and Policy*, 14(1):51–73.

European Commission (2002). Commission guidelines on market analysis and the assessment of significant market power under the community regulatory framework for electronic communications networks and services. 2002/C 165/03.

Garin-Munoz, T. and Perez-Amaral, T. (1996). Demand for international telephone traffic in spain: An econometric study using provincial panel data. *Information Economics and Policy*, 8(4):289–315.

Garin-Munoz, T. and Perez-Amaral, T. (1998). Econometric modelling of spanish very long distance international calling. *Information Economics and Policy*, 10(2):237–252.

Garin-Munoz, T. and Perez-Amaral, T. (1999). A model of spain-europe telecommunications. *Applied Economics*, 31(8):989–997.

Genakos, C. and Valletti, T. M. (2008). Testing the 'Waterbed' Effect in Mobile Telephony. *SSRN eLibrary*.

Grzybowski, L. and Pereira, P. (2008). The complementarity between calls and messages in mobile telephony. *Information Economics and Policy*, 20(3):279–287.

Hackl, P. and Westlund, A. H. (1995). On price elasticities of international telecommunication demand. *Information Economics and Policy*, 7(1):27–36.

Hermalin, B. and Katz, M. (2004). Sender or receiver: who should pay to exchange an electronic message? *RAND Journal of Economics*, 35(3):423. 0741-6261.

Hultkrantz, L. (2002). Telecommunications liberalisation in Sweden: Is "intermediate" regulation viable? *Swedish Economic Policy Review*, 9:133–161.

Jeon, D. S., Laffont, J.-J., and Tirole, J. (2004). On the receiver pays principle. *RAND Journal of Economics*, 35(1):85. 0741-6261.

Kim, J.-Y. and Lim, Y. (2001). An economic analysis of the receiver pays principle. *Information Economics and Policy*, 31:231–260. mimeo, Dongkuk University.

Kim, Y., Telang, R., Vogt, W. B., and Krishnan, R. (2008). Empirical Analysis of Mobile Voice and SMS Service: A Structural Model. *SSRN eLibrary*.

Laffont, J., Rey, P., and Tirole, J. (1998a). Network competition: I. overview and nondiscriminatory pricing. *RAND Journal of Economics*, 29:1–37.

Laffont, J., Rey, P., and Tirole, J. (1998b). Network competition: Ii. price discrimination. *RAND Journal of Economics*, 29:38–56.

Lam, P.-L. (1997). Erosion of monopoly power by call-back. lessons from hong kong. *Telecommunications Policy*, 21(8):693–695.

Larson, A., Lehman, D., and Weisman, D. (1990). A general theory of point-to-point long distance demand. In De Fontenay, A., Shugard, M., and Sibley, D., editors, *Telecommunications Demand Modelling*. North-Holland, Amsterdam.

Littlechild, S. C. (1975). Two-part tariffs and consumption externalities. *Bell Journal of Economics*, 6(2):661. 0361-915X.

Noam, E. (1994). Beyond liberalization II: The impending doom of common carriage. *Telecommunications Policy*, 18(6):435–452.

Office of Communications (2006a). Determination to resolve a dispute between Hay Systems Ltd and T-Mobile UK about SS7-based network access for SMS termination.

Office of Communications (2006b). Review of wholesale international services markets. explanatory statement and notification on market identification and market power determinations.

Office of Telecommunications (2003). Wholesale international services markets. identification and analysis of market and determination on market power. final explanatory statement and notification. Kim.etal2008Se.

Rea, J. D. and Lage, G. M. (1978). Estimates of demand elasticities for international telecommunication services. *Journal of Industrial Economics*, 26(4):363–381.

Reardon, M. (2008). U.S. text usage hits record despite price increases. *CNET News*. Available at: http://news.cnet.com/8301-1035_3-10038634-94.html.

Sandbach, J. (1996). International telephone traffic, callback and policy implications. *Telecommunications Policy*, 20(7):507–515.

Taylor, L. D. (1994). *Telecommunications Demand in Theory and Practice*. Kluwer Academic Publishers, Dordrecht; Boston.

Taylor, L. D. (2004). Customer demand analysis. In Cave, M., Majumdar, S., and Vogelsang, I., editors, *Handbook of Telecommunications Economics*, volume I. North-Holland, Amsterdam.

The Economist (2003). No text please, we're american. *The economist*, 366(8318):58. 0013-0613.

The Economist (2004). Je ne texte rien. *The economist*, 372(8383):57. 0013-0613.

The Washington Post (2008). Report: Text messaging will continue to be the darling of mobile data revenues. *The Washington Post*. Syndicated online from mocoNews.net. Available at: http://www.washingtonpost.com/wp-dyn/content/article/2008/12/03/AR2008120302106.html.

Vogelsang, I. (2003). Price regulation of access to telecommunications networks. *Journal of Economic Literature*, 16:830–862.

WIK-Consult (2008). The future of IP interconnection: Technical, economic and public policy aspects. Technical report, Study for the European Commission.

Wright, J. (2002). Access pricing under competition: An application to cellular networks. *Journal of Industrial Economics*, 50(3):289–315.

Table 1. Countries included in the dataset

Albania
Australia
Czech Republic
Egypt
Germany
Greece
Hungary
Ireland
Italy
Malta
Netherlands
New Zealand
Portugal
Spain
UK

Table 2. Descriptive statistics

The first set are the variables observed; the second are transformations of the above.
The key variables used in the regression are in bold.

Variable	Description	Unit	Obs	Mean	Std. Dev.	Min	Max
HHi	Herfindahl-Hirschman index		885	0.39	0.09	0.23	0.54
PriceT	**Price of sent SMS**	**£**	**843**	**0.06**	**0.02**	**0.00**	**0.12**
Subs	Subscribers by country	mln	885	21.40	23.70	0.24	86.60
TOut	SMS sent (total)	mln	844	216.23	292.65	1.67	2032.29
TIn	Off-net SMS received (total)	mln	778	81.95	93.85	0.09	548.94
VShare	Vodafone company share by country		885	35%	13%	10%	63%
COM	**Network size proxy (community of interest)**	**bn**	**885**	**217000**	**385000**	**14**	**1720000**
TInPU	**Off-net SMS received (per user)**		**778**	**14.25**	**10.20**	**0.55**	**45.50**
TOff	**SMS sent off-net (total)**	**mln**	**844**	**142.92**	**195.90**	**1.03**	**1359.95**
TOffPU	SMS sent off-net (per user)		844	23.26	17.92	3.21	108.76
TOn	SMS sent on-net (total)	mln	844	73.31	99.95	0.64	673.91
TOnPU	SMS sent on-net (per user)		844	15.55	17.81	1.26	99.91
TOutPU	SMS sent (per user)		844	38.80	33.65	5.71	184.05
VSub	Vodafone company subscribers by country	mln	885	6.62	7.92	0.15	30.80

Table 3. Estimation results

Dependent variable:	I	II	III	IV	V	**VI**	VII	VIII	IX	X	XI	XII
Total texts sent (log)	2SLS	2S BE	2S GLS	2S FE	2S FE	**2S FE**	2SLS	2S BE	2S GLS	2S FE	2S FE	2S FE
Text price (log)	-0.76	-0.63	-0.91	-0.92	-0.96	**-0.95**	-1.64	-2.08	-0.26	-0.25	-0.75	-0.76
	[0.03]**	[0.25]*	[0.02]**	[0.02]**	[0.02]**	**[0.02]****	[0.20]**	[1.89]	[0.32]	[0.34]	[0.08]**	[0.08]**
Texts received per user (log)	0.63	0.65	0.15	0.15	0.12	**0.12**	0.63	0.58	0.25	0.24	0.11	0.11
	[0.02]**	[0.14]**	[0.02]**	[0.02]**	[0.02]**	**[0.02]****	[0.03]**	[0.30]	[0.06]**	[0.06]**	[0.02]**	[0.02]**
Community of interest (log)	0.48	0.47	0.42	0.41	0.36	**0.35**	0.52	0.54	0.65	0.69	0.33	0.33
	[0.01]**	[0.04]**	[0.02]**	[0.02]**	[0.03]**	**[0.03]****	[0.01]**	[0.12]**	[0.12]**	[0.14]**	[0.03]**	[0.03]**
Interaction of price (log) and MMS dummy						**0.03**						0.03
						[0.01]**						[0.01]**
Constant	-14.27	-13.70	-11.79	-11.47	-10.06	**-9.80**	-18.21	-19.97	-17.32	-18.36	-8.69	-8.45
	[0.21]**	[1.63]**	[0.46]**	[0.49]**	[0.79]**	**[0.79]****	[0.92]**	[8.54]*	[2.86]**	[3.60]**	[1.03]**	[1.03]**
Year and month dummies?	No	No	No	No	Yes	**Yes**	No	No	No	No	Yes	Yes
Total price elasticity	-1.26	-1.09	-0.93	-0.94	-0.97	**-0.96**	-2.72	-3.13	-0.28	-0.27	-0.76	-0.77
R-squared	0.92	0.94	0.84	0.90	0.92	**0.92**	0.79	0.73	0.80	0.68	0.91	0.91
		(between)	(overall)	(within)	(within)	**(within)**		(between)	(overall)	(within)	(within)	**(within)**
Observations	762	762	762	762	762	**762**	762	762	762	762	762	762
Number of ID		15	15	15	15	**15**		15	15	15	15	15
Standard errors in brackets												
* significant at 5%; ** significant at 1%												

Instruments are: Lagged texts received per user (log) and Herfindahl index (the latter only for specifications 7-12).
Note that our chosen specification and the key variables are in bold.

Table A1. Time dummies results

Dependent variable:	(6)	(12)
Total Texts sent (log)	**IV FE**	**2S FE**
year==2003	**0.13**	**0.14**
	[0.02]**	**[0.02]****
year==2004	**0.19**	**0.22**
	[0.02]**	**[0.03]****
year==2005	**0.14**	**0.21**
	[0.03]**	**[0.05]****
year==2006	**0.05**	**0.16**
	[0.04]	**[0.07]***
year==2007	**-0.06**	**0.11**
	[0.06]	**[0.10]**
month==1	**0.00**	**-0.04**
	[0.03]	**[0.03]**
month==2	**-0.13**	**-0.17**
	[0.03]**	**[0.03]****
month==3	**-0.05**	**-0.09**
	[0.03]	**[0.03]****
month==4	**-0.13**	**-0.17**
	[0.03]**	**[0.03]****
month==5	**-0.12**	**-0.15**
	[0.02]**	**[0.03]****
month==6	**-0.15**	**-0.17**
	[0.02]**	**[0.03]****
month==7	**-0.11**	**-0.11**
	[0.02]**	**[0.03]****
month==8	**-0.10**	**-0.11**
	[0.02]**	**[0.03]****
month==9	**-0.13**	**-0.15**
	[0.02]**	**[0.03]****
month==10	**-0.12**	**-0.13**
	[0.02]**	**[0.03]****
month==11	**-0.13**	**-0.14**
	[0.02]**	**[0.03]****
Standard errors in brackets		
* significant at 5%; ** significant at 1%		

Note that the omitted year is 2002 and the omitted month is December

Table A2. Estimation results (specifying both usage variables as total traffic)

Dependent variable:	
Total Texts sent (log)	2S FE
Text price (log)	-0.95
	[0.02]**
Total texts received (log)	0.15
	[0.02]**
Community of interest (log)	0.28
	[0.02]**
Interaction of price (log) and MMS dummy	0.03
	[0.01]**
Constant	-7.88
	[0.74]**
Year and month dummies?	Yes
Total price elasticity	-0.97
R-squared	0.93
	(within)
Observations	762
Number of ID	15
Standard errors in brackets	
* significant at 5%; ** significant at 1%	

The instrument used is: Lagged texts received per user (log).
Note that the key variables are in bold.

Table A3. Estimation results (specifying both usage variables as per-user traffic)

Dependent variable:	
Texts sent per user (log)	**2S FE**
Text price (log)	**-0.94**
	[0.02]**
Texts received per user (log)	**0.16**
	[0.02]**
Community of interest (log)	**-0.18**
	[0.02]**
Interaction of price (log) and MMS dummy	**0.03**
	[0.01]**
Constant	**5.23**
	[0.74]**
Year and month dummies?	**Yes**
Total price elasticity	**-0.96**
R-squared	**0.89**
	(within)
Observations	**762**
Number of ID	**15**
Standard errors in brackets	
* significant at 5%; ** significant at 1%	

The instrument used is: Lagged texts received per user (log).
Note that the key variables are in bold.