

# DEMAND FOR AND PRICING OF MOBILE INTERNET: EVIDENCE FROM A REAL-WORLD PRICING EXPERIMENT\*

**Maija Gao**

Helsinki School of Economics, HECER and Nokia Corporation

**Ari Hyytinen**

Bank of Finland and

The Research Institute of the Finnish Economy (ETLA)

**Otto Toivanen**

HECER, University of Helsinki, and

The Research Institute of the Finnish Economy (ETLA)

## Abstract

Commercialization of innovations frequently stumbles. A prominent recent example are the early (i.e. *pre-3G*) mobile phone-enabled Internet services, whose European take-up was slower than expected. To determine why, we build a structural model of demand for such services and estimate it using consumer-level panel data from a real world pricing experiment. The experiment allows for a decomposition of the number of wireless connections into the number of needs - instances where a consumer would establish a connection if the price were zero - and the conditional probability of establishing a connection. We find that needs were plenty and potential consumer surplus several magnitudes higher than that attained. Marginal costs implied by the model are higher than those extracted from a structured survey of industry experts, indicating that prices were sub-optimally high. We find that pricing reduced usage substantially.

JEL: L11, D12

Keywords: wireless Internet, demand, new goods, pricing, services, surplus, two-part-tariff, welfare.

---

\* We would like to thank Colin Cameron, Helen Weeds, Glenn A. Woroch and seminar participants at Department of Economics at University of California, Berkeley, Haas School of Business, the Nordic ICT workshop in Helsinki, EARIE 2004 Berlin, the CEPR conference on economics of electronic communication markets at IDEI, Toulouse, London School of Economics, University of Bern, and KU Leuven for comments. We also thank the individuals who generously gave their time and shared their views either in interviews and/or in the survey we carried out. This research is part of the wireless communication research program of ETLA, the Research Institute of the Finnish Economy. Financial support from Nokia and Tekes is gratefully acknowledged. Hyytinen and Toivanen also gratefully acknowledge financial support by the Yrjö Jahnesson Foundation. Much of this research was completed while Hyytinen and Toivanen were visiting University of California, Berkeley, whose hospitality is gratefully acknowledged. The views expressed are those of the authors and do not necessarily reflect the views of the organizations the authors are affiliated with.

# 1 Introduction

Most of the existing empirical literature on new goods analyzes products that can be claimed successful (see, e.g., Trajtenberg 1989, Hausman 1997 and Petrin 2002), at least in some respects. A stylized fact is, however, that the commercialization of product and service innovations is difficult and that the launches of new goods frequently stumble and often fail (e.g. Scherer and Harhoff 2000). Quantitative analyses of such failures are rare, but potentially important in furthering our understanding of what can go wrong. The objective of this paper is to provide such an analysis by studying the demand for and pricing of the first wireless Internet services that were introduced in Europe at the end of the 1990s using an early wireless Internet technology, the Wireless Application Protocol (WAP).<sup>1</sup>

Wireless Internet generated high expectations that materialized in the hundreds of million euro of fees paid by mobile phone operators for the third-generation (3G) European mobile phone licenses (Klemperer, 2002). The take-up of the first (i.e., *pre-3G*) wireless services enabled by WAP was however not as rapid as expected.<sup>2</sup> Soon after its launch, some commentators announced WAP a

---

<sup>1</sup> The wireless technologies currently in use are more advanced than WAP, which was introduced in Europe in 1999. Shortly put, WAP was a first-generation solution to provide wireless services.

<sup>2</sup> The European take-up of the WAP enabled services was slow for example compared to wireless Internet in Japan, where i-mode, a service brand of NTT DoCoMo, took off rapidly after its introduction in February 1999. At that time, the emphasis in the U.S. was perhaps more on laptops and PDAs, but wireless internet services using mobile phones (e.g. AT&T's mMode), on which Europe and Asia initially concentrated, have recently been introduced also in the U.S..

failure.<sup>3</sup> While the current view and numbers clearly challenge the most critical accounts,<sup>4</sup> the question of why the take-up stumbled still remains. This question has more than historical relevance as any new product launch, and the currently ongoing launch of the third-generation mobile phone networks in particular, faces these same problems.

The first step in addressing the puzzle of why the launch of a particular new good or service fails to take off is to understand its demand: Is there *no latent demand* (i.e., no demand at a *zero-price*) for it, or is the quantity demanded low because of prices? Conditional on the latent demand being there, the second step in addressing the puzzle is to understand the impact of prices: Can sub-optimally high prices explain the slow take-off? We follow these steps in our search for an explanation for the slow take-up of the early European wireless services.

To understand whether there was insufficient demand for the new wireless services at any price (no latent demand), we make use of consumer-level data from a real world pricing experiment implemented in Fall 2001, and a structural model of demand. Owing to the experiment, we observe the prices charged and

---

<sup>3</sup> For example, the Nielsen Norman Group published a “WAP Usability Report” in December 2000. The report was based on a field study of WAP users in London and had a section titled “WAP Doesn’t Work” in the executive summary. It concluded that “When users were asked whether they were likely to use a WAP phone within one year, a resounding 70% answered no. WAP is not ready for prime time yet, nor do users expect it to be usable any time soon”. See also an analyst report titled “WAP in Europe: Has It Missed the Boat?” (by Lonergan, D. from the Yankee Group, published in 2000). Ph.D. Jacob Nilsen, cited by the Business Week to be “one of the world’s foremost experts in Web usability” and by Stuttgarter Zeitung, Germany “the world’s leading expert on user-friendly design”, called WAP the “Wrong Approach to Portability” in his October 1999 Alertbox -article (October 31, 1999, <http://www.useit.com/alertbox/991031.html>, accessed 30 May 2004).

<sup>4</sup> Views differ, but many see WAP as a constantly developing technology that has recently provided a bridge to newer generations of wireless technologies. A concrete example comes from Digital Airways, who introduced in October 2003 a new WAP-compatible version of Wapaka Web, its Java-based WAP simulator for the Web. The motivation was the take-off of WAP. The company states: “We had to respond to a growing demand for a WAP 2.0 version of Wapaka” and “Indeed, WAP is back.” (see <http://www.3g.co.uk/PR/Sept2003/5891.htm>, accessed 30 May 2004). Other estimates also speak for a kind of comeback of WAP: According to figures from the Mobile Data Association (the U.K.), the number of WAP page impressions viewed in the UK more than doubled during the nine months prior to May 2003. In Finland, usage has grown similarly.

the quantity demanded (i.e., the number and average length of connections) for a panel of consumers both during four non-experimental and three experimental two-week periods. Prior to and after the experiment, prices were at their normal (equilibrium) levels. During the experiment, both the per-minute price and the fixed connection fee were set to zero. Our structural model of demand decomposes the number of wireless connections into the number of needs that arise during a two-week period, and the conditional probability of establishing a connection, given a need. We think of a need as being e.g. a need to check the weather forecast before embarking on a long drive, and define a need as an instance where a consumer would establish a connection of strictly positive length if the price of doing so was zero. The experiment and decomposition allow us to identify the magnitude of latent demand for wireless services and to study separately the effect of pricing on the conditional probability of establishing a connection, given a need, and on the length of the connection.

Our data come from a Finnish mobile phone operator.<sup>5</sup> A strength of the data is that it combines individual characteristics and multiple observations per individual over a short time period, allowing us to estimate the parameters of latent demand and the conditional connection probability as functions of those characteristics.<sup>6</sup> We identify the parameters of our (econometric) model using variation that the experiment induces and estimate them using a flexible two-step *m*-

---

<sup>5</sup> Finnish operators have some track record in pioneering new services: The first digital mobile phone (GSM) call in the world was made in 1991 in Finland. Finland is one of the leaders in adoption of mobile telephony in general, and of wireless services in particular (see, e.g., Hausman 2002, Rouvinen and Ylä-Anttila 2003). Together with Japan, Finland was among the first countries where customers gained access to more advanced wireless services. At the time of the experiment, already some 16% of mobile phones were equipped with the necessary technology to utilize the kind of services we study. One might therefore expect that if anywhere, WAP enabled wireless services should have taken off early in Finland.

<sup>6</sup> Demand estimates for new services and goods are often wrought with empirical difficulties. Lack of detailed data is a primary reason. Bajari and Benkard (2003) and Berry, Linton and Pakes (2004) spell out some methodological difficulties.

estimator (e.g. Newey 1994a). The demand estimations show that consumers had demand for the first-generation wireless services, but it remained latent. The annual average latent (satiation) demand was 300 connections per consumer. If anything, this number appears consistent with the pre-launch hype and demand projections for the early wireless Internet. However, the average probability of establishing a connection, given a need, was only 10%. This low connection probability is explained by pricing and surprisingly elastic demand.

As a second step, we examine whether it was optimal for the operator to set prices that resulted in such a low connection probability. To this end, we back out from our structural model the marginal costs implied by the actual prices – which, given our assumption about the market power of the operator, give us an estimate of the *lower* bound of marginal costs implied by the model. We compare these estimates to the results of a structured survey of industry experts that provided us alternative estimates of marginal costs. We find that even after making a series of conservative assumptions, the marginal costs implied by the model are *higher* than the marginal costs estimates of the industry experts, implying that the prices were too high. The story that emerges is thus that sub-optimal prices reduced usage considerably.

We have also performed a number of counterfactual experiments. First, our upper-bound estimate (using actual prices) for the average welfare gain per consumer is 16 euros per year with a corresponding annual profit of 17 euros/customer, suggesting total welfare of 33 euros. This realized gain is lower than the *potential* total welfare, estimated to be over 60 euros. Second, the optimal two-part tariff, using the survey-based marginal costs estimates, implies that both the connection fee and the per-minute price should have been lower. The increase in profits (consumer welfare) from moving from actual to optimal prices is sensi-

tive to the choice of demand specification, but could have been as much as 10.8 (12.5) euros per customer.

The remainder of the paper is divided into five sections. In Section 2 we discuss the technology, its relationship to the previous and forthcoming technologies, and the services available. In Section 3 we present our structural model, derive our estimation equations, and discuss our identification strategy. In Section 4 we detail the data and its characteristics, and discuss both the demand and supply side implications of our estimation results. We also report the industry expert estimates of marginal costs and discuss their implications for the optimality of the actual prices. We provide an interpretation of our econometric findings and explore their robustness in Section 5. There we explore for example whether consumer experimentation, advertising, shifting demand in time or specification choices could explain the results. We close with a summary in Section 6.

## 2 The market, the technology and the experiment

The market we study is the service market enabled by the wireless Internet. The services were accessed using a mobile phone with a “first-generation” micro browser. The operator whose data we use launched its WAP-based wireless Internet service in late 1999.<sup>7</sup> At the time of the experiment in 2001, the operator’s own service portfolio consisted of 67 services, ranging from news, sports results

---

<sup>7</sup> WAP is a set of protocols that underlie one strand of the first technologies for the wireless Internet. The WAP architecture is similar to that of the WWW-browsing architecture, with the exception that WAP requires an intermediate layer (“the WAP gateway”), which determines how the wireless terminal and the Internet-architecture communicate. The devices and services available during the experiment were based on WAP version 1.1. The wireless technologies in use today, including later versions of WAP, are bridging the gap to the third and later generation technologies.

and weather services to games, betting results and TV-listings.<sup>8</sup> These services were tailored for a wireless user. The users also had access to external Internet sites.<sup>9</sup>

Pricing of the early wireless services was simple. A customer paid a fixed monthly fee for her wireless plan. The monthly fee depended on the plan she subscribed to, and no plan offered “free minutes”. Nor did the plans involve any leasing of handsets, because a Finnish law prohibited the practice.<sup>10</sup> There was no additional monthly fee for using wireless Internet, but an additional data call fee applied to all WAP connections. This fixed connection fee was 0.09 euros per connection. The per-minute online charge was either 0.12 or 0.17 cents, depending on the wireless plan. Additional content charges also applied, but these depended on the service provider.<sup>11</sup>

A crude but illustrative counterpart of the services we study comes from wireline Internet:<sup>12</sup> Using the early WAP was like having a dial up Internet ac-

---

<sup>8</sup> Kakkori (2001) provides a complete account of the types of services that were available in May 2001 via the operator’s own wireless portal: They include search services, ticket order, travel information, weather forecast, certain financial services, health-related services, news, communications, cinema, humor, dating, phone personalization, music, games, radio listings, tips where to eat and drink, TV listings, sports results, horoscope, betting results, and various operator services.

<sup>9</sup> Because of the small screen sizes and limited input capabilities of the early mobile browsers, the range and quality of services that the consumers in our sample were able to access are more limited than what are available today.

<sup>10</sup> The prices of WAP handsets available in 2001 were, approximately, from 250 EUR (Ericsson R320s) to 435 EUR (Nokia 6210). According to industry estimates, there were about 655000 WAP compatible mobile phones in Finland at the time of the experiment. They accounted for about 16% of the stock of digital mobile phones. The proportion of WAP-enabled phones was growing rapidly, however. It has been estimated that in 2001, nearly half of the new mobile phones were able to utilize WAP.

<sup>11</sup> It has been estimated that during 2001, about 60 % of the operator sponsored wireless services had an additional contents charge (Kakkori, 2001). See Kakkori (2001, p. 24) for examples of contents charges. Data on the distribution of the usage between the services with and without additional content charges is not available to us. We know, however, that during the experiment, a large majority (about 95%) of the per-two week charges for WAP-usage were zero (despite the dramatically increased usage). This means that customers accessed mostly services with no additional charges.

<sup>12</sup> Both WAP and the other leading wireless Internet technology of the time, i-mode in Japan, transmitted the data at 9600 bits per second, which is quite slow. The display sizes were also quite small. For example, one of the most often used terminals had a display with 96x60 pixels.

cess. This meant that it took a while to download data intensive services and/or applications. The wireless technologies currently in use are based on intermediate technologies, often called “2.5G”. For the end-user, they are like a dedicated Internet access<sup>13</sup> that is often used for data transmission only and that sometimes comes with enhanced quality (speed). Accessing wireless services enabled by the next generation mobile networks (“3G”) is a bit like accessing the Internet using a high-speed connection such as DSL or a cable modem.

The Finnish mobile phone operator whose data we use is one of the major firms in the market. The experiment was conducted as an advertising campaign for the new WAP enabled services, and consisted of lowering the per-minute-price and the fixed connection fee to zero for six weeks in Fall 2001 for all customers of the operator.<sup>14</sup> The operator advertised the campaign in TV, radio, and the major national and local newspapers both before and during the experiment. We believe that all customers with a WAP enabled phone were aware of the campaign taking place. After the campaign, the prices returned to their previous levels.

### 3 The model and estimation strategy

The first step in our search for an explanation for the slow take-up of the early WAP enabled wireless services is to infer whether there was latent demand for them. The second step is to understand the supply side.

---

<sup>13</sup> This is like using Integrated Services Digital Network, known as ISDN or ISDL, which is an early version of Digital Subscriber Line, DSL.

<sup>14</sup> One could argue that in order to get reasonable welfare estimates our data should include variation also up, not only down, from the actual prices. We agree with this in principle. It turns out, however, that for reasonable values of marginal costs (see below), the actual prices in our data were above (short-run) monopoly prices. We therefore think that our data contain information on how customers behave with “high” and zero prices.

### 3.1 Step 1: Demand side analysis

#### **The theoretical model of demand**

We use of a discrete-continuous demand model that allows us to identify the magnitude of the latent demand, accounts for consumer heterogeneity and has finite satiation levels of consumption both with respect to the number and length of connections.<sup>15</sup> These features are important, for the number and (average) length of established connections vary between periods for a given consumer, and between consumers within a given period, and are bounded even during the experiment periods when the prices are zero.

The model is built on the analysis of a single wireless service connection: whether or not to make it, and if, at what length. We assume that the utility function from consuming a connection of length (in minutes)  $q$  when the connection fee is  $K$  and the per-minute price is  $p$  is given by the commonly used additively separable form  $V(q, \alpha) - (K + pq)$ , where  $V(\cdot)$  is concave and twice continuously differentiable in  $q$ , and  $\alpha$  is a demand parameter explained below.<sup>16</sup>

We introduce two types of demand shocks that determine observed behavior. The timing is as follows: First, a consumer faces consumption (connection) needs that are derived from a stochastic process. We define a need as an instance where a consumer would establish a connection of strictly positive length if the prices of doing so were zero. The need distribution will below be made a function of consumer characteristics: We imagine for example that there can be cross-

---

<sup>15</sup> Dubin and McFadden (1984) are an early example of a discrete-continuous demand model to which class our model also belongs.

<sup>16</sup> This is a commonly used transformation both in the Industrial Organization literature. For its derivation, see e.g. Tirole (1988, pp. 143). It is also commonly used in telecommunications analysis (e.g. Mitchell, 1978).

sectional variation in the frequency at which consumers have a need to check the weather forecast (through wireless internet) before embarking on a long drive.

Second, once a connection need arrives, the consumer receives a connection-specific demand shock,  $\alpha$ . We assume that  $V(q, \alpha)$  is increasing in  $\alpha$  and that it determines the strength of the need, i.e., the satiation length of a given connection. The satiation length will also be made a function of consumer characteristics. For example, one might presume that young people have higher satiation demands for games. We follow the long literature on discrete choice by making *Assumption 1*: The connection-specific demand shock,  $\alpha$ , is an independent and identically distributed random shock with a known probability density function  $f(\alpha)$ , with support  $[0, \infty)$ .

The cumulative density function (c.d.f.) of  $\alpha$  is denoted  $F(\alpha)$ , which we take to be continuous, monotonically increasing, and twice continuously differentiable. Because the consumer establishes a wireless connection only if her consumer surplus from the connection is non-negative, the optimal length of a service connection is

$$(1) \quad q(p, K, \alpha) = \arg \max [V(q, \alpha) - (K + pq)]$$

if  $CS(p) \equiv \int_p^\infty g(\tau; \alpha) d\tau \geq K$  and zero otherwise. This decision rule determines

whether or not to make a single wireless service connection, and if, at what length. This means that *conditional* on a connection need, a consumer will establish a wireless service connection with probability  $\pi = [1 - F(\bar{\alpha}(p, K))]$ , where  $\bar{\alpha}(p, K)$  is the unique value of the shock that is implicitly defined by

$CS(p; \bar{\alpha}(p, K)) \equiv \int_p^\infty q(\tau; \bar{\alpha}(p, K)) d\tau \equiv K$ . It follows that for a sequence of y con-

nection needs, the number of connections established follows a binomial distribution with parameters  $y$  and  $\pi$ .

To introduce the stochastic process generating the need shocks, we lean on a large telecommunications engineering literature and assume that the needs are draws from a Poisson distribution:<sup>17</sup>

*Assumption 2:* The number of connection needs during a given time interval,  $y$ , is distributed Poisson with mean  $\mu$ .

Assumption 2 means that the expected number of consumption needs for wireless services during a time interval is  $\mu$ . Our model, coupled with Assumptions 1 and 2, yields the following result:

*RESULT:* The number of connections made during a given time interval,  $Y$ , is given by a Poisson-stopped Binomial:  $Y \sim \text{Poisson}$  with  $E[Y|X] = \mu\pi$ .<sup>18</sup>

Summing up, there are two types of demand shocks that determine observed behavior: First, consumers face connection needs that are derived from a Poisson process. Once a need arrives, the consumers receive a connection-specific draw of  $\alpha$  from the distribution  $f(\alpha)$ , which determines how strong the need is. Given the connection need and armed with knowledge of prices  $p$  and  $K$  and the value of  $\alpha$ , the consumer calculates whether her utility is maximized by establishing a connection of the optimal length, or by not connecting at all. A connection is established with probability  $\pi = 1 - F(\bar{\alpha}(p, K))$ , and the number of connections is thus the result of consumer optimization.

---

<sup>17</sup> See e.g. <http://www.jdstelecom.com/telecom.php> (“Use to find the number of trunks required for an offered traffic to have a specified probability of blocking. All assume random (Poisson) arrivals and exponential call holding times.”), where tools based on our type of modeling of the number and length of calls is promoted (accessed 30 May 2004).

<sup>18</sup> The result follows directly from the properties of the two stochastic processes (see, e.g., Lemma 1.1.4 of Cameron and Trivedi 1998, pp. 8) and Assumptions 1 and 2.

## Operationalization of the model

To make the model operational, we have to make further assumptions about the expected number of consumption opportunities for wireless services during the time interval  $[\mu]$ , the density of the demand shift shock  $[f(\alpha)]$ , and the form of demand determining the optimal connection length  $[q(p)]$ .

The number of consumption needs that consumer  $i$  faces during time period  $t$  ( $= 1, 2, \dots, 7$ ),  $y_{it}$ , corresponds in our model to the consumer's satiation demand during the period. Satiation demand is typically unobservable. We can however measure it, because during the experiment periods  $p_t = K_t = 0$ , implying that  $\bar{\alpha}(p_t, K_t) = 0$  and  $\pi = 1 - F(0) = 1$  for  $t = 3, 4, 5$ . The number of connections made during the experiment period therefore reflects latent demand and a complete fulfillment of needs. Thus  $y_{it} = Y_{it}$ , and by Assumption (2),  $y_{it}$  is distributed Poisson with mean  $\mu$ . We allow for heterogeneity and model the mean satiation demand of consumer  $i$  flexibly as a function of demographics. We assume, specifically, that  $\mu_i = \exp(g(\underline{z}_i; \underline{\mu}))$ , where  $g(\underline{z}_i, \underline{\mu})$  is an initially unknown function of the vector of consumer demographics  $\underline{z}_i$  and the associated vector of parameters  $\underline{\mu}$ .

The theory is silent about the distribution of the demand shift parameter  $\alpha$ , which assumes a different value for each need that consumer  $i$  faces during period  $t$ . Following the telecommunications engineering literature, we specify that  $\alpha$  has a (stationary) exponential distribution. However, we allow for heterogeneity as follows:  $\alpha_{ij} \sim \exp(\lambda_i)$ , with  $E[\alpha_{ij}] = 1/\lambda_i$ , where  $j$  indexes connection needs ( $j = 0, \dots, y_{it}$ ) of consumer  $i$  during period  $t$ . The mean depends on consumer demo-

graphics, i.e.,  $\lambda_i = k(\underline{z}_i, \underline{\lambda})$ , where  $k(\underline{z}_i, \underline{\lambda})$  is an initially unknown function of the vector of consumer demographics  $\underline{z}_i$  and the associated vector of parameters  $\underline{\lambda}$ .

Absent an established practice, we consider two different demand specifications for  $q(p)$ . We assume that the demand for the service (and thus the length of the connection) is either a linear ( $q_{it,j} = \alpha_{ij} - \gamma_i p_t$ ) or a log-linear ( $q_{it,j} = \alpha_{ij} \exp(-\gamma_i p_t)$ ) function of the per-minute price  $p_t$ . Besides simplicity, the strength of these demand functions is that they allow us to parameterize  $\gamma_i$  as a function of consumer demographics ( $\gamma_i = \underline{\gamma}' \underline{z}_i$ ). They make our analysis comparable with recent analyses of telecommunications demand,<sup>19</sup> provide a robustness check to each other and, consistent with our data, allow for a point of satiation.

### **Estimation and identification of demand parameters**

Together with the above auxiliary assumptions, the theoretical model implies that the number of connections made by consumer  $i$  during period  $t$ ,  $Y_{it}$ , follows a Poisson-stopped Binomial process with mean  $E[Y_{it} | p_t, K_t, z_i] = \mu_i \pi_{it}$ . However, when estimating the model parameters we only make use of the moment condition given by the conditional mean.<sup>20</sup> The conditional mean can be re-written as

$$(2) \quad E[Y_{it} | p_t, K_t, z_i] = \mu_i [1 - F(\bar{\alpha}(p_t, K_t))] = \exp(g(\underline{z}_i, \underline{\mu}) - \lambda_i \bar{\alpha}(p_t, K_t)),$$

where  $\bar{\alpha}(p_t, K_t) = (\underline{\gamma}' \underline{z}_i) p_t + \sqrt{2(\underline{\gamma}' \underline{z}_i) K_t}$  in the case of linear demand and

$\bar{\alpha}(p_t, K_t) = (\underline{\gamma}' \underline{z}_i) K_t \exp((\underline{\gamma}' \underline{z}_i) p_t)$  in the case of log-linear demand. It is our theo-

---

<sup>19</sup> Our analysis differs from the previous studies because we use the two functional forms side-by-side and data on wireless services, not voice calls. In addition, our data is from an experiment, and we use flexible estimation methods. For studies using the linear demand function, see for example Miravete (2002) and Miravete and Röller (2003). The log-linear demand is known as “Perldemand” in the telecommunications literature; see Taylor (2002) for a recent review of this literature.

<sup>20</sup> That is, we do not impose the potentially very restrictive Poisson density or variance assumption; see, for example, Gourieroux, Montfort and Trognon (1984a,b), and Wooldridge (1997).

retical model that imposes the restriction that the coefficients of the per-minute price and the connection fee are not allowed to differ. Lack of sufficient (inter-temporal or cross-sectional) price variation means that it is not feasible to relax this restriction in our econometric specification.

In identifying the parameters and unknown functions of the conditional mean, we utilize the unique feature of our data that we have three periods where both prices ( $p_t$  and  $K_t$ ) are set to zero. Although in general we have censoring in the model, i.e., a consumer establishes a connection only if she faces a need *and* if the associated demand shock is large enough, during the experiment periods, there is *no* censoring. The experiment ( $p_t = K_t = 0$ ) implies that whenever there was a need, and whatever the realization of the demand shock, a connection was established. To make use of this identifying information we adopt a two-step  $m$ -estimator where the first step is a non-parametric (series) estimator (see Newey 1994a for general consistency results and also Newey 1994b and Pakes and Olley 1995, who consider related semiparametric  $m$ -estimators). In the first step, we a) use data on the number of connections during the experiment periods to estimate flexibly the initially unknown function determining the satiation number of connections (i.e., the expected number of needs of consumer  $i$ ) and b) use the average connection length during the experiment periods to estimate flexibly the initially unknown function determining the mean of the demand shift shock. In the second step, we use the predicted values from the first-step and identify the price effects using the price variation induced by the experiment.

The first step identifies and estimates the latent demand as follows: First, because  $y_{it} = Y_{it}$  for  $t = 3, 4, 5$ , variation in the number of connections equals variation in the number of needs. This variation identifies  $\mu_i = \exp(g(\underline{z}_i, \underline{\mu}))$ . We run regressions of the form

$$(3) \quad E[Y_{it} | p_t = 0, K_t = 0, z_i] = \mu_i = \exp(g(z_i, \underline{\mu}))$$

for  $t = 3, 4$  and  $5$ , and estimate  $g(z_i, \underline{\mu})$  flexibly using a power series estimator and cross-validation. Cross-validation leads to a mean-square error minimizing choice of the number of terms and allows for a choice of data-dependent number of terms (see Li 1987, Hausman and Newey 1995, Newey 1997). We let  $z_i$  consist of consumers' age, gender and place of residence, and their powers and interactions.

The other parameter of latent demand is  $\lambda_i = k(z_i, \underline{\lambda})$ , which determines the mean of the demand shift shocks. Variation in connection lengths across consumers during the experiment identifies  $\lambda_i = k(z_i, \underline{\lambda})$ , because every time a consumer faced a consumption need during the experiment, she utilized it with probability one and made a wireless service connection of the length that gave her satiation. In terms of our model, this implies that

$$(4) \quad E(q_{it} | p_t = 0, K_t = 0, z_i) = 1/\lambda_i = 1/k(z_i, \underline{\lambda}),$$

for  $t = 3, 4$  and  $5$ . To estimate  $k(z_i, \underline{\lambda})$  flexibly we use a power series estimator and cross-validation. For this estimation we use data on the *average* duration of connections of consumer  $i$  during period  $t$  ( $t = 3, 4$  and  $5$ ), because that is what we observe.

In the second step we estimate the price effects, parameterized by  $\gamma' z_i$ , using the conditional mean equation to which the predicted values from the first step have been plugged in. That is, we insert  $\hat{\lambda}_i$  and  $\hat{\mu}_i$  into (2) to obtain

$$(5) \quad E[Y_{it} | p_t, K_t, z_i] = \hat{\mu}_i \left[ 1 - \hat{F}(\bar{\alpha}(p_t, K_t)) \right] = \hat{\mu}_i \exp(-\hat{\lambda}_i \bar{\alpha}(p_t, K_t)).$$

where  $\bar{\alpha}(p_t, K_t)$  is a function of  $\gamma' z_i$ . The price variation induced by the experiment identifies the price effects, and we allow them to vary with the age, gender and place of residence of the consumer. To estimate  $\gamma$ , we make use of the mo-

ment condition (5) and the method of pseudo-maximum likelihood, which belongs to the class of  $m$ -estimators (as required; see Newey 1994a, p. 3 and Cameron and Trivedi 2005, p. 200-202). Following the recommendation of Cameron and Trivedi (2005, p. 202) and the example of Pakes and Olley (1995), we bootstrap the second step standard errors.

### 3.2 Step 2: Supply side analysis

One of the central benefits from estimating a structural model is the ability to recover supply-side unobservables, which in our case are marginal costs. A challenge that we face is that one of the explanations for the slow take-up of the early WAP services is that, conditional on the latent demand being there, the pricing of these services was not necessarily optimal.

To explore whether the prices were set (sub)optimally and, by implication, whether a pricing mistake could explain the slow take-up, we resort to a three part approach. We first use the supply side moments (first order conditions) of our model, assuming monopoly pricing, to estimate the implied marginal costs of both a new connection, and a one minute lengthening of an existing connection. The monopoly assumption seems feasible on the following four grounds (that we detail further in part A of the Appendix): First, mobile internet services were additional (add-on) services within an already chosen wireless plan; second, the expenditures on voice calls dominated over expenditures on mobile internet; third, the lack of number portability (see e.g. Viard 2006) strongly restrained customers from switching operators; and fourth, assuming monopoly pricing will give us a lower bound estimate of marginal costs. If one's prior is that the prices were too high, the static monopoly assumption is conservative. Given that the operator may have behaved as a monopolist in relation to WAP services, we can solve the first order conditions for (static) profit maximization, and use these to estimate the

marginal costs implied by our model. We provide the FOCs in part A of the Appendix.

The second part of our approach makes use of the results of a structured survey that we administered to acknowledged industry experts. The data generated by this survey is a substitute for the (hard) marginal cost data which we lack. In the survey the respondents were first asked point estimates of the two marginal costs. We then asked a couple of auxiliary questions and finally elicited information on the *distribution* of marginal costs, following the work of Dominitz and Manski (1996, 1997). While straightforward in principle, conducting a structured survey for this paper's purposes was not trivial for two reasons: First, the respondents were typically not familiar with the economic concept of marginal cost.<sup>21</sup> Second, it took a considerable amount of effort to be able to identify people who have in depth knowledge about the costs and who thus *a priori* are in a position to be able to give informative answers. We were able to elicit 18 responses with a 50% response rate (for details, see part B of the Appendix).

The third part of our approach is to compare the (mean and median) marginal cost estimates of the industry experts with the estimates from our model.

## 4 Empirical analysis

### 4.1 Data

The sample was constructed by identifying all the customers of the operator who had a private phone and who during the two middle weeks of the six-week experiment (i.e., during period 4) established at least one wireless Internet connection. While the campaign was well advertised and had been running for two

---

<sup>21</sup> Indeed, of the 36 experts that we approached, half declined to answer our questionnaire after having read the questions, citing as reason that they felt they did not have the necessary expertise or information to provide answers.

weeks, it may still be possible that not all (potential) users are included in our data.<sup>22</sup> Because the expected number of needs during a two-week period varies over customers, one group of the excluded could be those with a very low number of needs: given that prices were zero the maintained assumption is that everybody hit with a need establishes a connection. Missing these individuals from our sample may induce an upward bias in our estimates of the number of needs. A simple calculation indicates that this type of selection is not likely to be an issue.<sup>23</sup> Second, if one drops the maintained assumption of establishing a connection at zero prices given a need, we might miss customers with high opportunity costs. Such selection would bias our results towards too low price elasticities.<sup>24</sup>

Table 1 displays the descriptive statistics for the sample, separately for the non-experiment and experiment periods, as well as the basic consumer demographics. We have a balanced panel with 14882 consumer-period observations ( $N = 2126$  and  $T = 7$ ). During the non-experiment periods, the average number of wireless connections (WAP\_COUNT) per a two-week period is 1.13. The average connection length (CALL\_DUR) is 2.66 minutes. During the experiment periods, usage grows dramatically: The average number of needs (=number of connections during the experiment periods) is 11.94 per a two-week period. The average satia-

---

<sup>22</sup> In Finland, mobile phone customers buy their phones from private vendors, not from operators. The operator in our case thus does not know how many of its customers had a WAP enabled phone during the observation period. Consequently, we don't know how many of the potential users we have in our data.

<sup>23</sup> Making the Poisson assumption and using our mean latent demand estimate, the probability of having zero needs during the 2<sup>nd</sup> experiment period is  $\Pr[y_{it} = 0] = e^{-11.54} \approx 0.00001$ . It appears that the bias in the mean estimate would have to be very large for this type of selection to be an issue.

<sup>24</sup> As the time period for the experiment was short, the price experiment was limited to wireless Internet services, there was no phone-number portability in Finland yet,<sup>24</sup> and the vast majority of mobile phone usage was created from voice calls and SMS messages, we do not believe there was any selection into the sample. It is, however, still conceivable that some individuals might have changed operators to take advantage of the experiment. These would presumably be people with a high number of needs but high price elasticity (as their own operators were offering the same services for positive prices). The existence of such selection would bias our estimates of the number of needs and our price elasticity estimates upwards.

tion connection length (= length of connection during the experiment period) is 5.54 minutes. There is, however, a lot of variation.

The table also shows that we have two kinds of wireless plans in the data: Plan “A” is clearly more popular, as 86% of the consumers subscribe to it. In this plan, the per-minute charge (WAP\_PMIN) is 0.12 euros and the fixed connection fee (WAP\_K) is 0.09 euros per connection. In the empirics that follow, we use data from this larger plan only and use the other for a robustness analysis.<sup>25</sup>

[Insert Table 1 here]

Table 1 shows that the experiment increases usage dramatically, and that any difference in behavior before and after the experiment are small. Below we demonstrate that the latter point holds closer scrutiny and therefore proceed treating pre- and post-experiment data identically.

## 4.2 Estimation results and their implications

For brevity, we turn directly to the demand and supply side implications of the model and present the first step cross-validation of  $\exp(g(z_i, \underline{\mu}))$  and  $k(z_i, \underline{\lambda})$  and estimation results ( $\hat{\lambda}_i$  and  $\hat{\mu}_i$ ) and the second step estimation results ( $\hat{\gamma}_i$ ) in part C of the Appendix.

### **Step 1: Demand side implications**

The demand side implications of the model are summarized in Table 2. It shows the decomposition of the demand for wireless services into the number of needs (i.e., the latent demand) and the conditional probability of establishing a connection, given a need. The expected number of connections established (i.e., the quantity demanded) as well as a number of different price elasticities are also re-

ported. We have computed these estimates for each individual as a function of her demographics and evaluated them, where appropriate, at the market price. The means and the standard deviations that are reported in the Table are computed over the cross-section of consumers.

[Insert Table 2 here]

The first demand side implication is the magnitude of latent demand: The average satiation demand is 11.5 needs per consumer over a two-week period, or about 300 connections a year.<sup>26</sup> Consumption needs arose thus almost daily. These numbers suggest that lack of latent demand for the new wireless Internet services cannot explain their slow take-up.

The second demand implication is that given a need, the conditional probability of opening a connection is on average only 10%, i.e., only every tenth need translates into a connection. This probability translates into a low expected number of connections (30 p.a.) at the then charged market prices. Thus, the quantity demanded appears to have remained low because of (high) prices. The estimated price elasticities, evaluated at the market price, provide further support for this conclusion. The average per-minute price ( $p_t$ ) elasticity of *connection length* is high, ranging from -1.87 (log-linear demand) to -2.28 (linear demand). The per-minute price ( $p_t$ ) and the fixed fee ( $K_t$ ) elasticities of the *number* of wireless connections are high, too. For the linear (log-linear) demand, the average per-minute price elasticity of the number of wireless connections is -1.59 (-4.47). The average fixed fee elasticity of the number of wireless connections is -0.39 (-2.37) for the

<sup>25</sup> We did not resort to using the data from the two plans together even if that would have given us cross-sectional variation in prices. The reason is that we wanted to avoid any selection bias that this might have induced (as selection into the plans is endogenous).

<sup>26</sup> The magnitude of latent demand is independent from the choice of the functional form for  $q(p)$ , because for  $p_t = 0$ , they are identical.

linear (log-linear) demand. Note that these are short-run elasticities, and the long-run elasticities, which our data do not allow us to estimate, may be even higher.

Figure 1 shows the age profile of latent demand and the conditional probability of opening a connection, given a need.<sup>27</sup> It suggests that consumer heterogeneity is important, too. It seems that needs decrease with age, but that the probability of actually satisfying the needs using the services increases with age. Thus a 20 year old had almost twice as many connection needs as a 70 year old, but his probability of connection was only half of that of the 70 year old.

[Insert Figure 1 here]

To sum up, the demand side results support two conclusions: First, lack of latent demand for WAP enabled wireless Internet services cannot explain their initially slow take-up. Second, demand remained latent because of high prices, and the early demand for the wireless services appears to have been highly elastic. These results raise the question of whether prices were optimally set.

## **Step 2: Supply side implications**

We report the marginal costs implied by the model and the (mean and median) marginal cost estimates of the industry experts in Table 3.<sup>28</sup> As can be seen from the Table, the linear and log-linear models imply quite different marginal costs: the linear demand model yields an estimated marginal connection cost ( $C$ ) of 0.03 euros, while that given by the log-linear model is 0.07 euros. As to the marginal cost of connection length ( $c$ ), the linear (log-linear) model yields an estimate of 0.07 (0.08) euros. Because of the maintained assumption of static monopoly pricing, these numbers should be thought as lower bound estimates of the marginal

---

<sup>27</sup> The plot is drawn for the linear demand specification using data for males that live in the Helsinki capital area. The pattern for log-linear demand is similar.

<sup>28</sup> No standard errors for the estimated marginal costs can be provided due to lack of variation in prices during the non-experiment periods.

costs: Most, if not all, models of dynamic monopoly or oligopolistic competition suggest that the observed prices should have been closer to marginal costs.

The average (median) expert opinion point-estimate of  $C$  is 0.015 (0.010) euros and that of  $c$  0.011 (0.010) euros.<sup>29</sup> The expected values of  $C$  and  $c$ , as elicited from the subjective marginal cost distributions, provide a similar picture. The average (median) of these over the survey respondents is 0.022 (0.013) and 0.020 (0.008) euros. The *maximum* point estimates from the survey are  $C = 0.050$  and  $c = 0.040$  euros. In the distribution part of the survey, we first asked for the lower and upper bound of the relevant marginal cost's support, and then the probability of the true value falling into each of the quintiles. The medians of the *upper* bounds were  $C = 0.040$  and  $c = 0.030$  euros.

We can make use of the survey data and the fact that our monopoly assumption gives us a lower bound estimate of the marginal costs to calculate the probability that the two marginal costs ( $C$  and  $c$ ) are at least as high as those backed out from the model. Assuming a normal distribution and taking the mean and standard error estimates from Table 3 as its first two moments, we find that these probabilities are exceedingly small for all but the marginal cost of connection from the linear model.

Comparing these two sets of marginal cost estimates supports the conclusion that the marginal costs implied by both demand models, especially the one on connection length, are too high.<sup>30</sup> These results support the conclusion that sub-optimal prices may explain the commercial failure of the early WAP services, and suggest a number of counterfactual experiments.

---

<sup>29</sup> Using answers to our technical check-question as a screen, we discarded 5 of the 18 respondents from the final sample (see section B of the Appendix). Using all 18 respondents' answers, the medians are the same as those reported in Table 3. The means are:  $C = 0.030$ ,  $c = 0.032$  which, while higher, are still below the model-based estimates.

[Insert Table 3 here]

### 4.3 Counterfactual experiments and surplus calculations

We perform three counterfactual experiments. As the first experiment, we infer optimal prices using the estimated demand model and the marginal costs provided by the expert survey. As the second experiment, we disallow the use of a two-part tariff. Third, we impose marginal cost pricing. We focus on usage and welfare implications. We use the average marginal costs calculated from the respondents' marginal cost distributions as these are the highest and thus for our purposes the most conservative estimates.<sup>31</sup>

[Insert Table 4 here]

There are four rows in Table 4 for each demand specification. The first one reports the figures for the actual prices; the second, for optimal prices; the third for the case with no connection fee (but with optimal  $p$ ); and the fourth for the case of marginal cost pricing. We report i) the probability of connecting given a need; ii) the average length of the connection; iii) annual producer surplus per customer; and iv) average annual consumer surplus.

We start with the results obtained using linear demand. With actual prices, the estimated annual producer (consumer) surplus per customer is 17.05 (16.40) euros. Marginal cost pricing would have yielded an average annual total welfare of 61.33 euros. The dead-weight loss is thus 46% of the welfare attainable at marginal cost pricing. Consumer behavior would have changed a lot had the prices

<sup>30</sup> It also seems that the relative magnitudes given by the log-linear model are more in line with the survey evidence than those obtained from the linear model.

<sup>31</sup> Yet an additional experiment of interest is to abolish the 22% value added tax (VAT) that applied to the early wireless services. This counterfactual suggests that keeping prices constant but abolishing the tax would have resulted in a 416% (140%) increase in the number of connections with log-linear (linear) demand. An important implication of this finding is that when the demand for a new good or service is elastic, taxation may reduce its usage and slow down its take off considerably (see also Goolsbee 2000).

been optimal: The average connection probability increases by 58% and the average connection length decreases by about 18% with optimal prices. The corresponding change in consumer (producer) surplus is 28% (5.5%). The dead-weight loss with optimal prices (conditional on the two-part tariff, i.e., actual pricing structure) is still 36%. Abolishing the possibility of using two-part tariffs increases the connection probability, shortens the expected duration of a connection, and increases both producer (3%) and consumer surplus (26%).

Turning to the log-linear specification (lower part of the Table) we find the changes to be larger than those obtained using the linear demand model. With actual prices, the estimated profit per customer (average consumer surplus) is 4.12 (1.09) euros. The dead-weight loss is almost 90%, as marginal cost pricing yields an average annual total welfare of 43.74 euros per customer. Going from actual to optimal prices increases the connection probability seven-fold and the connection length by 5%. Producer surplus more than triples, consumer surplus increases nearly 13-fold and the dead-weight loss decreases to 35%. Imposing  $K = 0$  has a bigger effect in relative terms than with the linear demand: The producer (consumer surplus) is 54% (95%) of that with an optimal two-part tariff.

## 5 Interpretation and robustness

### 5.1 Interpretation of the empirical evidence

#### **Understanding the demand for the early WAP services**

Our analysis shows that demand existed for the early wireless services at the time of the experiment in 2001. It however appears to have been surprisingly elastic. Our interpretation of the high own-price elasticity is that the pre-3G wireless services had close (quality-adjusted) substitutes. Early WAP compatible mobile phones had a relatively small and difficult to use screen and keypad. They did not

support (many) colors either, reducing the quality of graphics. The early wireless Internet technologies were also quite slow, partly competed with the existing short message services, and essentially repackaged a set of existing digital services that were relatively easily available through other channels of distribution. The availability of wireline Internet access was relatively widespread in Western Europe by 2001 (and broadband was on the rise, OECD 2001), and we conjecture that the (quality-adjusted) cost of access was competitive relative to the wireless services.

A comparison of the success of the early WAP with the early diffusion of i-mode, a service brand of NTT DoCoMo which took off quite rapidly after its introduction in February 1999 in Japan, provides a reality check of the above interpretation.<sup>32</sup> First, i-mode's overall usability appears to have been somewhat better. It was, for example, a bit faster, and supported colors. Second, when i-mode was introduced, the take up of wireline Internet was relatively modest in Japan compared to Western Europe, making its content relatively novel. Finally, it seems that a number of the services made available by i-mode were *not* widely available in Japan before, suggesting less repackaging.<sup>33</sup> All this suggests that i-mode had fewer substitutes and better quality than WAP. A back-of-the-envelope price comparison (see part D of the Appendix) nevertheless indicates that the prices of the early WAP enabled wireless services were, at least in Finland, in all likelihood higher than the prices of the comparable services in Japan: The conditional connection probability, given a need, would have been higher than it actually was, had the prices of the WAP services been lower and thus closer to the price level of i-mode.

---

<sup>32</sup> Most of the descriptive details in this paragraph are from Kakkori (2001).

<sup>33</sup> Examples of new services are messages across operators' networks, and picture downloads. Especially entertainment related services, such as "What's new" -information services and music sites, became popular early on (Marketing Interactive Network, 2000).

## **Understanding the pricing of the early WAP services**

Usually in the IO literature firms are assumed to set their prices optimally. Our results suggest that prices were sub-optimally high but leave open whether the operator systematically failed to maximize its profits when setting its prices or whether we have identified an isolated pricing mistake.

We are inclined to prefer the latter for the following two reasons. First, the operator has a long and successful history of providing (fixed and mobile) voice services which still form its main revenue source. It thus seems that it knew how to price voice services. Second, mobile internet services were a new product not only to the customers, but also to the firm, and quite different from the “old” voice services the firm was used to producing. It therefore seems possible that the firm made a mistake in pricing something it was unfamiliar with. That firms may make mistakes or not optimize in such a situation is not entirely new: Levitt (2006) provides evidence against optimal pricing in absence of sufficient market feedback using a case study, and Ellison (2005) discusses the need to apply behavioral economics in IO to the supply side of the market.

Our conjecture on why the operator made a pricing mistake is that the operator failed to appreciate how elastic the demand for the early WAP services was. As our results show, pinning down the elasticity/elasticities exactly (and understanding its drivers) was not a trivial task even when usage data and econometric models are employed. For what it is worth, our auxiliary survey evidence (see section 5.3) also support this conjecture: Even with hindsight, industry experts have widely differing opinions on how elastic the demand for the early WAP services was.

## 5.2 Robustness of the demand side analysis

### **Alternative explanations for the existence of latent demand and high estimated elasticities**

There are at least three competing explanations from outside our model that might explain either the magnitude of the latent demand or the high estimated elasticities, or both: *First*, experimentation by consumers during the experiment period in order to learn quality may have increased usage during the experiment period. This could bias our parameter estimates towards too strong latent demand or high elasticities. A testable implication of this type of learning is that if the consumers learned that they liked (disliked) the services, the demand (number of connections) after the experiment should have been higher (lower) than before the experiment. We should also observe declining usage during the experiment period. *Second*, some of the services (e.g. paying bills) are such that one might be able to shift their consumption in time. Knowledge of the start and end of the experiment period might have induced consumers to shift their consumption to the zero-price experiment periods, again biasing our results. The testable implication of this is that the demand in the non-experimental periods adjacent to the experiment periods would be systematically lower than in other non-experimental periods. *Third*, the advertising related to the price experiment may have shifted demand, implying again an upward bias both in the estimated magnitude of latent demand and price elasticities. If advertising's effects last a few weeks, the number of connections should be higher and their average length longer during the post-experiment periods than in the pre-experiment periods.

To test these implications we ran a number of reduced-form regressions explaining the number of connections and their length by consumer and period indicators. These regressions show that there is only a little, if any, evidence for learn-

ing, consumption shifting in time or an advertising effect (see part E of the Appendix).

### **Data and specification choices**

We have data from two different wireless plans (plans “A” and “B”), but only used data on plan “A” customers to generate our main findings above. We can therefore address the question of whether our demand-side results are an artifact of the estimating sample by re-estimating our model using data on customers from the other pricing plan. We have data on 306 consumers who subscribed to this (less popular) plan. The results (see part F of the Appendix) are in line with those obtained above with plan “A”.

One way to check the robustness of our results with respect to our specification is to consider simpler specifications. We imposed consumer homogeneity, i.e.,  $\mu_i = \mu$ ,  $\lambda_i = \lambda$  and  $\gamma_i = \gamma$  for all  $i$ . The results echo our previous findings. In the linear demand case the annual average consumer surplus is 15.53 euros per consumer. We have repeated our analysis also for other combinations of the explanatory variables, but found no significant differences.

A deeper criticism challenges our entire demand model. To provide an alternative estimate, we generate a standard reduced form estimate of demand elasticity and consumer surplus in spirit of Hausman (1997, 1999) and Brynjolfsson, Hu and Smith (2003). Hausman shows that a good measure of the total effect on consumer welfare can be based on the compensating variation (CV). The results from these estimations show that the own-price elasticity is -10.84 and the annual per-consumer surplus 0.56 euros (see part G of the Appendix for details). Estimating these elasticities and consumer surpluses in several alternative ways reinforces this result.

The final question we ask is which of the two demand specifications better fits the data. To answer the question, we performed an out-of-sample test: We calculated the predicted call lengths from both models at actual prices, and compared them to the observed call lengths during the non-experiment periods as these were *not* exploited in the estimation. The difference between predicted and actual is 4.13 (-.74) minutes for the linear (log-linear) model, both significant at the 1% level. The difference in the (mean) squared prediction error (between the two models) is 17.01 with p-value .000, i.e., the linear model's prediction error is systematically larger. Thus, while the log-linear model slightly but systematically underestimates the call length, it produces better out-of-sample estimates than the linear model. This finding suggests that we should emphasize the results obtained from the log-linear model over those from the linear model. Doing so would strengthen our conclusions.

### **Comparability with existing demand side evidence**

Many studies suggest that the demand for new goods and services is relatively inelastic, particularly in mobile telecommunications (Hausman 1997, 2002). In this light our results seem puzzling. However, the results from the emerging literature on the demand for Internet (access) are close to ours. Goolsbee (2006) finds that the demand elasticity for broadband Internet access ranged from -2.2 to -3.7 in the U.S around 1999. Varian (2000) portrays a similar picture. Using data from the Berkeley INDEX experiments, he finds that the own-price elasticities for bandwidth were between -2 and -3. Finally, Ellison and Ellison (2004) study the effects of price search on the Internet on demand elasticity, and document Internet price elasticities that are very high, sometimes of magnitude -50. None of these studies is directly comparable to ours, but they put the elasticity of the demand for the early WAP into a perspective.

Our welfare results and particularly estimates of the realized consumer surplus may also seem puzzling in light of the recent research on new goods and services, as a number of papers document large consumer gains from various new goods and services (Trajtenberg 1989, Breshanan and Gordon 1997, Petrin 2002, Brynjolfsson, Hu and Smith 2003) and particularly from (tele)communications product/service innovations (e.g., Hausman 1997, 1999, Goolsbee and Petrin 2004). For example, if we take one of our largest consumer surplus estimates, the realized per-consumer welfare from the early wireless services is less than 15% of the gains that satellite TV channel buyers experienced in the U.S. (Goolsbee and Petrin 2004).<sup>34</sup> The puzzle is, however, more apparent than real, for all the earlier papers analyze *successful* launches of new goods. As our counterfactual experiments show, the launch of the early WAP services could have been more successful and consumer surpluses substantially higher than those actually obtained.

### 5.3 Robustness of the supply side analysis

#### **Quality of marginal cost estimates**

Our supply side analysis relies to a large extent on the plausibility of the mean and median marginal cost estimates of the industry experts. We have considered their plausibility in two ways.

First, we analyzed the cost structure of WAP services via a series of in depth interviews with an industry expert to identify the drivers of the marginal costs of the operator supplying the WAP services. We learnt that the operator provided the service mostly in-house and owned (most of) the infrastructure needed for its provision. It also turned out that the amount of electricity that a WAP connection

---

<sup>34</sup> All these numbers are naturally based on short-run elasticities, but given the equipment costs, the estimated consumer welfare figures do suggest that the long-run figures (through adoption) wouldn't be much greater.

uses is negligible; that apart from one area in central (capital) Helsinki, congestion was not an issue; and that all the equipment owned by the operator and used to run the network constituted towards fixed, not variable cost (as the equipment did not depreciate with usage). Besides electricity, the only additional outlays from establishing or lengthening a WAP connection appears to have come from two sources. First, the operator paid (on the basis of an undisclosed contract) a small usage-based charge to firms operating and owning some but not all of the antennas and links (relays) of the network. Second, the operator had revenue sharing agreements with content providers which seemingly were of minor importance. All this information, in addition to the survey results, suggests to us that the relevant marginal costs were very low in absolute terms.<sup>35</sup>

As a second check of the quality of the marginal cost estimates, we asked the survey respondents whether a WAP connection was technically identical to a (GSM) voice connection as well as their point estimates of the two marginal costs of a GSM connection. This is a true statement, and we used it as a check on their understanding of the technology used for WAP services. Most respondents (13 out of 18) answered the question correctly. We only used the answers of those respondents that at least partly agreed with the claim.

### **Auxiliary survey evidence**

Expanding the supply side analysis to e.g. include other firms and alternative market structures is not possible due to lack of data. We can, however, benchmark some of our supply side findings against the views of the industry experts. As part of the survey we asked whether the charged prices were too high; what the optimal prices would have been; and what the price elasticity of WAP services was in

---

<sup>35</sup> The marginal cost estimates of the expert we interviewed in depth echo this view, as they were close to the lower bound of the estimates obtained via the survey.

2001. The survey gives some support for the statement that the charged prices were too high (average 2.7, mode = 3 and median = 3 on a 4-point Likert scale with 1 = strongly disagree, ..., and 4 = strongly agree). Four out of five respondents suggest lower prices than those actually charged.<sup>36</sup> The suggested prices are in fact close to the optimal prices implied by the log-linear model, as the mean (median) suggested per-minute price is 0.040 (0.050) euros and the mean suggested connection fee 0.039 (0.050) euros. Consistent with our interpretation, the respondents found it difficult to provide an estimate for the elasticity of the demand and the estimates varied considerably. While we do not want to put too much weight on these subjective, ex post views, they appear to be in line with our main conclusions.

## 6 Conclusions

New goods and services play a fundamental role in how markets improve our living standards. A stylized fact is that most product launches fail, implying that an improved understanding of what might go wrong is of great importance. The mobile Internet is an example of a recently launched new service that, at least initially, stumbled. In Europe, the early take-up of the new class of wireless services enabled by mobile phones was initially slower than expected. We studied why.

A key part of answering the question of how the launch of a particular new good or service succeeds is to understand its demand. We find that needs to use mobile Internet were plenty, and thus that lack of latent demand cannot explain their sluggish initial adoption. Contrasting the marginal cost estimates derived from our model with those obtained from a structured survey of industry experts

---

<sup>36</sup> There is only one price suggestion that is strictly *higher* than the actual prices: One respondent

we conclude that the marginal costs implied by the model are too high. Because of the conservative assumptions that we have made, our preferred interpretation of this – backed with auxiliary survey evidence – is that sub-optimally high prices explain the commercial failure of the early WAP services. Indeed, our counterfactual pricing experiments suggest that pricing (and taxation) of the services had a strong effect, curtailing usage. This finding provides support for the hypothesis of Levitt (2006) that deviations from profit maximization are more likely when firms operate in a new or unfamiliar environment, e.g., when they have had insufficient time or market feedback to learn to set prices optimally.

Our analysis also suggests that with the linear demand model, *both* producer *and* consumer surplus would have increased if the operator had priced optimally as a monopolist. With the log-linear demand (which fits the data somewhat better than the linear model), the increases would have been even larger. The realized producer surpluses per consumer and welfare gains to an (early) adopter of these services were on average moderate, 4-17 euros and 1-16 euros a year, respectively. These contrast with an estimated 40-60 euro potential total surplus. These findings are consistent with the early adoption experiences and critical accounts that were aired at the time our experiment was run.

Our results suggest that the very modest welfare created by what clearly was a run-of-the-mill new service could have been greatly increased by different pricing. The larger issue is that the risks attached to welfare creation through new goods do not stop at the invention stage, but continue well into the innovation and diffusion stage, and that a better understanding of how to handle those commercial risks, especially those arising from pricing, could significantly raise the social and private returns to innovation.

suggested that the optimal  $p$  is 0.10 euros.

## References

- Bajari, Patrick and Benkard, C. Lanier, 2003, Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches, mimeo, Stanford University.
- Berry, Steve, Linton, Oliver and Pakes, Ariel, 2004, Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *Review of Economic Studies*, forthcoming.
- Bresnahan, Timothy F. and Robert J. Gordon, eds., 1997, *The Economics of New Goods*, Chicago, IL: University of Chicago Press.
- Brynjolfsson, Erik, Hu, Yu (Jeffrey), and Smith, Michael D., 2003, Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers, *Management Science*, Vol. 49, No. 11, pp. 1580-1596.
- Cameron, Colin A. and Trivedi, Pravin K., 1998, *Regression Analysis of Count Data*, Cambridge University Press, Melbourne, Australia.
- Cameron, Colin A. and Trivedi, Pravin K., 2005, *Microeconometrics: Methods and Applications*, Cambridge, UK: Cambridge University Press.
- Dominitz, Jeff and Manski, Charles F., 1996, Eliciting Student Expectations of the Returns to Schooling, *Journal of Human Resources*, Vol. 31, pp. 1-26.
- Dominitz, Jeff and Manski, Charles F., 1997, Using Expectations Data to Study Subjective Income Expectations, *Journal of the American Statistical Association*, Vol. 92, pp. 855-867.
- Dubin, J. and McFadden D., 1984, An econometric analysis of residential electric appliance holdings and consumption, *Econometrica*, Vol. 52, No. 2, pp. 345-362.
- Ellison, Glenn and Ellison, Sara F., 2004, Search, Obfuscation, and Price Elasticities on the Internet, mimeo, Massachusetts Institute of Technology.
- Ellison, Glenn, 2005, Bounded Rationality in Industrial Organization, in: T. Persson and R. Blundell (Eds.), *Advances in Economics and Econometrics: theory and Applications*, Cambridge, UK: Cambridge University Press, forthcoming.
- Goolsbee, Austan, 2000, In a World without Borders: The Impact of Taxes on Internet Commerce, *Quarterly Journal of Economics*, Vol. 115, Issue 2, pp. 561-576.
- Goolsbee, Austan, 2006, The Value of Broadband and the Deadweight Loss of Taxing New Technology, *Contributions to Economic Analysis & Policy*, (B.E. Press Journals), Vol. 5, No. 1.
- Goolsbee, Austan and Petrin, Amil, 2004, The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable Television, *Econometrica*, Vol. 72, No. 2, pp. 351-381.
- Gourieroux, C., Montfort, A. and Trognon, Alain., 1984a, Pseudo Maximum Likelihood Methods: Theory, *Econometrica*, Vol. 52, No. 3, 681-700.
- Gourieroux, C., Montfort, A. and Trognon, Alain., 1984b, Pseudo Maximum Likelihood Methods: Applications to Poisson Models, *Econometrica*, Vol. 52, No. 3, 701-720.
- Hausman, Jerry A., 1997, Valuing the Effect of Regulation on New Services in Telecommunications, *Brookings Papers on Economic Activity. Microeconomics*, Vol. 1997, pp. 1-38.
- Hausman, Jerry A., 1999, Cellular Telephone, New Products and the CPI, *Journal of Business and Economic Statistics*, Vol. 17, No. 2 (April), pp. 188-194.
- Hausman, Jerry A., 2002, Mobile Telephone, in Martin E. Cave, Sumit K. Majumdar and Ingo Vogelsang (Eds.), *Handbook of Telecommunications*, Volume 1, The Netherlands: North-Holland, pp. 564-603.
- Hausman, Jerry A., and Whitney K. Newey, 1995, Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss, *Econometrica*, Vol. 63, No. 6 (November), pp. 1445-1476.
- Hausman, Jerry A., Hall, Bronwyn H., and Griliches, Zvi, 1984, Econometric Models for Count Data with an Application to the Patents-R & D Relationship, *Econometrica*, Vol. 52, Issue 4 (July), pp. 909-938.
- Kakkori, Matti, 2001, Comparing wireless Internet services: i-mode and WAP, Helsinki University of Technology, Department of Electrical and Communications Engineering, Master's thesis.
- Klemperer, Paul, 2002, How (Not) to Run Auctions: the European 3G Telecom Auctions, *European Economic Review*, Vol. 46, Issues 4-5 (May), pp. 829-845.
- Levitt, Stephen D., 2006, An Economist Sells bagels: A case study in profit Maximization, NBER Working Papers, No. 12152.
- Li, K. C., 1987, Asymptotic Optimality for  $C_p$ ,  $C_L$ , Cross-validation, and Generalized Cross-validation: Discrete Index Set, *Annals of Statistics*, Vol. 15, pp. 958-975.
- Marketing Interactive Network, 2000, MIN 1st i-mode User Survey, InfoCom Research, Inc
- Miravete, Eugenio J., 2002, Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans, *Review of Economic Studies*, Vol. 69, pp. 943-971.

- Miravete, Eugenio J. and Röller, Lars-Hendrik, 2003, Competitive Nonlinear Pricing in Duopoly Equilibrium: The Early U.S. Cellular Telephone Industry, CEPR Discussion Paper No. 4069.
- Mitchell, Bridger M., 1978, Optimal Pricing of Local Telephone Service, *American Economic Review*, Vol. 68, No. 4, pp. 517-536.
- Newey, Whitney K., 1994a, Series Estimation of Regression Functionals, *Econometric Theory*, Vol. 10, pp. 1-28.
- Newey, Whitney K., 1994b, The Asymptotic Variance of Semiparametric Estimators, *Econometrica*, Vol. 62, No. 6, pp. 1349-1382.
- Newey, Whitney K., 1997, Convergence Rates and Asymptotic Normality for Series Estimators, *Journal of Econometrics*, Vol. 79, pp. 147-168.
- OECD, 2001, The Development of Broadband access in OECD Countries, unclassified document of Working Party on Telecommunication and Information Services Policies, Directorate for Science, Technology and Industry, Organization for Economic Co-operation and Development.
- Pakes, Ariel and Olley, Stephen, 1995, A Limited Theorem for a Smooth Class of Semiparametric Estimators, *Journal of Econometrics*, Vol. 65, pp. 295-332.
- Petrin, Amil, 2002, Quantifying the Benefits of New Products: The Case of the Minivan, *Journal of Political Economy*, Vol. 110, No. 4, pp. 705- 729.
- Reiss, Peter C. and Wolak, Frank A., 2004, Structural Econometric Modeling: Rationales and Examples from Industrial Organization, In: *Handbook of Econometrics*, Volume 5, forthcoming.
- Rouvinen, Petri and Ylä-Anttila, Pekka, 2003, Case Study: Little Finland's Transformation to a Wireless Giant, in S. Dutta, B. Lanvin and F. Puaa (Eds.), *The Global Information Technology Report 2003-2004*, New York: Oxford University Press, pp. 87-108.
- Scherer, Frederic M. and Harhoff, Dietmar, 2000, Technology policy for a world of skew-distributed outcomes, *Research Policy*, Vol. 29, pp. 559-566.
- Taylor, Lester D., 2002, Customer Demand Analysis, in Martin E. Cave, Sumit K. Majumdar and Ingo Vogelsang (Eds.), *Handbook of Telecommunications*, Volume 1, The Netherlands: North-Holland, pp. 98-138.
- Tirole, Jean, 1988, *The Theory of Industrial Organization*, Cambridge, MA: The MIT-Press.
- Trajtenberg, Manuel, 1989, The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners, *Journal of Political Economy*, Vol. 97, Issue 2 (April), pp. 444-479.
- Wooldridge, Jeffrey, M., 1997, Quasi-likelihood methods for count data, in: Pesaran, M. H. and Schmidt, P. (eds.), *Handbook of Applied Econometrics*, Volume II: Microeconomics, Oxford Blackwell Publishers, pp. 353-406.
- Varian, Hal, 2000, Estimating the Demand for Bandwidth, mimeo, University of California at Berkeley.
- Viard, Brian, 2006, Do Switching Costs Make Markets More or Less Competitive?: The Case of 800-Number Portability, *RAND Journal of Economics*, forthcoming.

Table 1. Descriptive statistics

Non-experiment periods 1-2 and 6-7		Obs: N*4	Mean	Std. Dev.	Min	Max
Connection duration (CALL_DUR)						
- Prop. of customers with # of connections > 0		8504	0.239	0.427	0	1
- Connection length, conditional on > 0		2034	2.661	2.992	0.047	77
Number of connections (WAP_COUNT)		8504	1.131	4.568	0	118
Price per minute (WAP_PMIN, euro / min)						
- Wireless plan "A"		7280	0.120	0	0.12	0.12
- Wireless plan "B"		1224	0.170	0	0.17	0.17
Fixed per-connection fee (WAP_K, euros)		8504	0.090	0	0.09	0.09
Experiment-periods 3-5		Obs: N*3	Mean	Std. Dev.	Min	Max
Connection duration (CALL_DUR)						
- Prop. of customers with # of connections > 0		6378	0.797	0.402	0	1
- Connection length, conditional on > 0		5086	5.542	4.729	0.067	64
Number of connections (WAP_COUNT)		6378	11.942	22.213	0	325
Price per minute (WAP_PMIN, euro / min)						
- Wireless plan "A"		5460	0	0	0	0
- Wireless plan "B"		918	0	0	0	0
Fixed per-connection fee (WAP_K, euros)		6378	0	0	0	0
Consumer characteristics		Obs: N	Mean	Std. Dev.	Min	Max
Age (AGE, in years)		2126	36.45	12.41	18	86
Gender (GENDER, Male = 1)		2126	0.72	0.45	0	1
Location (CITY, Helsinki area = 1)		2126	0.23	0.42	0	1
Subscription plan (SUBTYPE, "A" = 1)		2126	0.86	0.31	0	1

Table 2. Structural parameters and economic implications

	Linear demand		Log-linear demand	
	Mean	Std. Dev.	Mean	Std. Dev.
Expected number of needs ( $\mu$ )				
- per two-week period	11.539	1.485	11.539	1.485
- per year	300.015	38.614	300.015	38.614
Prob(connect / need) ( $\pi$ )	0.100	0.042	0.100	0.039
Expected # of connections established ( $\mu\pi$ )				
- per two-week period	1.128	0.415	1.128	0.371
- per year	29.325	10.784	29.321	9.650
The price elasticity of:				
- wireless connection length				
w.r.t. per-minute price (p)	2.280	0.534	1.869	0.145
w.r.t. fixed connection fee (K)	-	-	-	-
- # of wireless connections				
w.r.t. per-minute price (p)	1.590	0.285	4.469	0.922
w.r.t. fixed connection fee (K)	0.390	0.032	2.367	0.339

Table 3. Marginal cost estimates

Marginal cost of	Model		Expert opinion (survey data)				Std.err.*
	Linear	Log-linear	Aver. 1*	Aver. 2**	Med. 1*	Med. 2**	
- Establishing connection, C	0.029	0.074	0.015	0.022	0.010	0.013	0.019
Prob[C > Model C]***	0.236	0.001	-	-	-	-	-
- Lengthening connection by 1 min., c	0.067	0.080	0.011	0.020	0.010	0.008	0.012
Prob[c > Model c]***	<0.001	<0.001	-	-	-	-	-

\* based on point estimates (survey data)

\*\* based on distribution estimates (survey data)

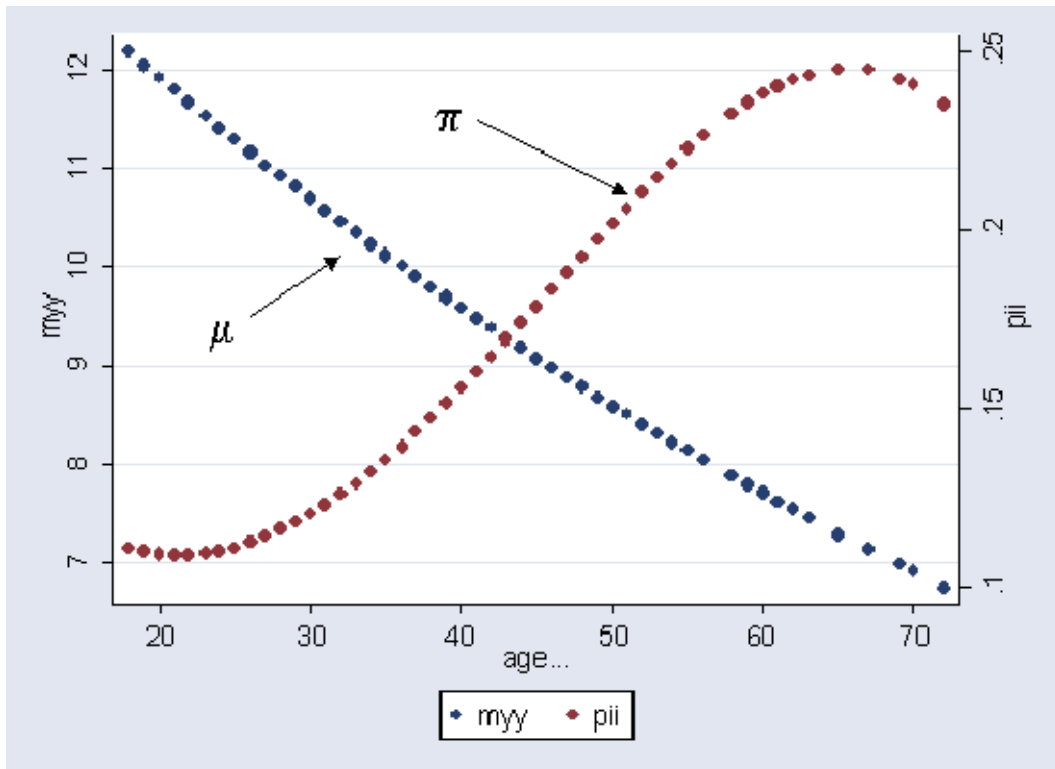
\*\*\* assuming normality and using the standard error reported in the last column

Table 4. Counterfactual experiments

	Connection probability ( $\pi$ )	Average connection length (min)	Producer surplus (euro)	Consumer surplus (euro)
Linear model				
Actual prices (p = 0.12, K = 0.09)	0.10	6.92	17.05	16.40
Optimal prices (p = 0.09, K = 0.02)	0.16	5.62	17.99	21.02
No connection fee (p = 0.096, K = 0)	0.22	3.89	17.64	20.63
Marginal cost pricing (p = c, K = C)	0.48	5.53	0.00	61.33
Log-linear model				
Actual prices (p = 0.12, K = 0.09)	0.06	2.31	4.12	1.09
Optimal prices (p = 0.055, K = 0.046)	0.45	2.43	14.96	13.58
No connection fee (p = 0.085, K = 0)	1.00	0.67	8.02	12.92
Marginal cost pricing (p = c, K = C)	0.83	3.15	0.00	43.74

Note: c = 0.020, C = 0.022

Figure 1. Consumption needs and the conditional connection probability



## APPENDIX

### A. Monopoly assumption and FOCs

*Monopoly assumption:* The operator whose data we use had a considerable degree of monopoly power over the first (actual and potential) users of the mobile internet services.<sup>37</sup> The primary reason for this is that most of the potential users were captive. There are four pieces of evidence supporting this claim: First, the consumers had subscribed to a wireless-voice plan before the new mobile services were available. This means that once available, the new service was a kind of add-on and consumed within the plan, i.e., conditional on being a subscriber to the particular voice-dominated service bundle. As we will argue, the pricing of the add-on (i.e., the new service within the bundle) was at the discretion of the operator. Second, over the period our data is from, the voice and earlier non-voice (e.g., text messaging) parts of the bill were clearly larger than that emanating from the use of mobile internet services. We therefore think it unlikely that a different price for the new mobile Internet services would have lead customers to switch operators and by implication, that there was any significant price competition in the dimension of the operator's offering we focus on. Third, even if a consumer wanted to switch, there was a considerable non-pecuniary hurdle: The cellular phone numbers were not portable at the time. According to survey evidence reported by the Finnish Ministry of Transports and Communications, portability was the most important obstacle to switching (see also Viard 2006). The statistics on the number of consumers who have switched after portability was implemented in June 2003 unambiguously support this survey evidence. Finally, assuming monopoly pricing will give us a lower bound estimate of marginal costs. Our prior is that prices were too high, implying that monopoly assumption works against our prior. The results confirm ex post that the monopoly assumption is reasonable in the sense of us finding that the actual prices were higher than static optimal monopoly prices. That is, the estimated optimal *monopoly* prices from our preferred model are *lower* than the actual prices that were charged. This revealed pricing behavior strongly suggests that the operator whose data

---

<sup>37</sup> To be sure, it was not strictly speaking a formal monopoly. The operator was over the time period we study one of the two dominant players in the Finnish mobile telecommunications market.

we have did *not* behave as if it faced significant price competition over the users of the early mobile services.<sup>38</sup>

*Optimal two-part tariff and first-order conditions (FOCs):* Ignoring consumer heterogeneity for brevity, the optimal two-part tariff solves

$$(A.1) \quad \max_{p,K} \Omega = [\mu(1 - F(\bar{\alpha}(\tau p, \tau K)))] [q_{\alpha > \bar{\alpha}}(\tau p, \tau K)(p - c) + K - C]$$

where  $c$  = marginal cost per minute of connection,  $C$  = marginal cost of opening/terminating a connection,  $q_{\alpha > \bar{\alpha}}$  = conditional connection length, and  $\tau$  = the valued added tax (VAT) parameter (22%), implying that per-minute price (connection fee) charged from the customer is  $\tau p$  ( $\tau K$ ). The above objective function also explicitly allows for the dependence of  $\bar{\alpha}$  and  $q_{\alpha > \bar{\alpha}}$  on the parameters of the tariff.

Let

$$A(p, K) \equiv \mu(1 - F(\bar{\alpha}(\tau p, \tau K))); \text{ and}$$

$$B(p, K) \equiv q_{\alpha > \bar{\alpha}}(\tau p, \tau K)(p - c) + K - C.$$

The general form of the FOCs is

$$(A.2) \quad \frac{\partial \Omega}{\partial i} = \frac{\partial A(p, K)}{\partial i} B(p, K) + A(p, K) \frac{\partial B(p, K)}{\partial i} = 0$$

for  $i = p, K$ . For the linear demand function, the parts of the FOCs can be written as:

$$(A.3) \quad A(p, K) = \mu \exp(-\lambda(\gamma \tau p + \sqrt{2\gamma \tau K}))$$

$$(A.4) \quad B(p, K) = \sqrt{2\gamma \tau K} + 1/\lambda](p - c) + K - C$$

$$(A.5) \quad \frac{\partial A(p, K)}{\partial p} = -\lambda \gamma \tau \mu (1 - F(\bar{\alpha}(\tau p, \tau K)))$$

---

<sup>38</sup> If and when rivals offer similar products at different prices, changing (raising) prices might lead some customers to change their operator. The estimated optimal monopoly prices mean, however, that should our operator have a lower degree of monopoly power than we have assumed, a change towards the optimal prices should have induced an in-flow of customers (assuming the potential rivals would not have changed their behavior). If rivals had matched the price changes, there would have been no reason for their customers to change operators. It is therefore unlikely that our customer-level figures would be greatly affected by such switches. If at all, the effect would most likely be that our figures present a lower bound. The reason for this is that the most likely customers to change from a rival operator to the operator whose data we have would have been those with above average valuation (and hence usage) of mobile internet services.

$$(A.6) \quad \frac{\partial B(p, K)}{\partial p} = \sqrt{2\gamma\tau(1/K)} + 1/\lambda$$

$$(A.7) \quad \frac{\partial A(p, K)}{\partial K} = \left[ -(1/2)\sqrt{2\gamma\tau(1/K)} \right] (1 - F(\bar{\alpha}(\tau p, \tau K)))$$

$$(A.8) \quad \frac{\partial B(p, K)}{\partial K} = (1/2)\sqrt{2\gamma\tau K} (p - c) + 1.$$

For the log-linear demand function, the corresponding parts take the following forms:

$$(A.9) \quad A(p, K) = \mu \exp(-\lambda\gamma\tau^2 K \exp(\gamma\tau p))$$

$$(A.10) \quad B(p, K) = [(p - c)(\gamma K \tau^2 \exp(\gamma\tau p) + 1/\lambda) \exp(-\gamma\tau p) + K - C]$$

$$(A.11) \quad \frac{\partial A(p, K)}{\partial p} = -\lambda\gamma^2\tau^3 K \exp(\gamma\tau p) [\mu(1 - F(\bar{\alpha}(\tau p, \tau K)))]$$

$$(A.12) \quad \frac{\partial B(p, K)}{\partial p} = \gamma K \tau^2 + \exp(-\gamma\tau p)(1/\lambda)(1 - \gamma\tau(p - c))$$

$$(A.13) \quad \frac{\partial A(p, K)}{\partial K} = -\lambda\gamma\tau^2 \exp(\gamma\tau p) [\mu(1 - F(\bar{\alpha}(\tau p, \tau K)))]$$

$$(A.14) \quad \frac{\partial B(p, K)}{\partial K} = \gamma\tau^2 (p - c) + 1.$$

## B. Survey of expert opinions

*Design and structure of the survey:* In initial discussions with an industry expert it became clear that the relevant marginal costs are, given the circumstances of WAP introduction (i.e., existing networks with high coverage, large capacity and therefore no need to install new capacity/expansion of network coverage), the opening/termination cost of an extra WAP (GSM) connection through the existing network, and the marginal cost of lengthening an existing WAP (GSM) connection by one more minute. It also turned out that it is most likely the case that the operators have not systematically gathered and/or stored data on the two marginal costs.

To obtain information on marginal costs, we designed a structured survey that consisted in total of 13 questions and that was administrated to a number of acknowledged industry experts (more on this below). The structure of the survey was as follows: First, a definition and example of marginal costs were given. Second, the respondents were asked point estimates of the two marginal costs (Q1-Q2). The next two questions were about the marginal costs of GSM calls (Q3-Q4) and the fifth (Q5) was a statement about the (technical) equivalence of WAP connections and

GSM calls. We designed Q5 to be a check-question. The purpose was to be able to separate the answers of those respondents who know that WAP connections were technically identical to GSM calls from the answers of those who didn't know this. The sixth question (Q6) was about the elasticity of demand and the seventh (Q7) about pricing. Finally, we elicited through a series of questions (Q8-Q13) information on the distribution of marginal costs, following the work of Dominitz and Manski (1996, 1997).<sup>39</sup> The survey instrument was pre-tested with an industry expert and modified slightly on the basis of his reactions to it.

*Selection of the survey frame and survey procedure:* We asked for names of people likely to have intimate knowledge of WAP as of 2001 both from an acknowledged industry expert, and from operators and a large telecom equipment company, yielding us in the end 36 names. We sent these people an email explaining our survey (with the survey attached) and requested to call them at their convenience. After a couple of reminder emails we phoned them if we had not received any reaction. When making the survey-call, we explained them (as described in the survey document) the purpose of the study and the survey, and proceeded to ask the questions. We did not systematically record the length of the interviews, but a typical interview lasted 35-45 minutes.

*Response rate and reasons for non-response:* In the end, 18 people agreed to answer our questions, but several of them made various reservations. Those who declined invariably gave as the main reason that they felt that they did not have adequate expertise to answer our questions (which they often called difficult). Only one person declined because of corporate policy. We ended up discarding answers of 5 people because they did not agree at least partly with our technical background question (Q5 in the survey). Some of those who are included in the final survey sample made various reservations which we recorded. In order to be conservative, we decided to keep their answers in the sample as they regularly gave higher estimates of marginal costs than those respondents not making reservations. By doing this we hope to have avoided any downward bias in our estimated marginal costs, as downward bias would strengthen our results vis-à-vis deviations from optimal pricing, and the consequences of this.

### C. Estimation results

---

<sup>39</sup> The complete survey document is available at Otto Toivanen's www-page ([www.hecer.fi](http://www.hecer.fi)).

We present the basic estimation results of the first and second step of the two-step m-estimation procedure briefly here, because the key insights come from the economic implications of the model. Table A.1 displays the cross-validation results, separately for  $\exp(g(z_i, \underline{\mu}))$  and  $k(z_i, \underline{\lambda})$ . The first step estimation results are displayed in Table A.2, which among things shows that Wald-tests indicate that the included variables are jointly highly significant. The results from the second step are displayed in Table A.3, separately for linear and log-linear demand. We report Huber-White standard errors adjusted for clustering within consumers in both tables. Bootstrapping the standard errors in the second step is important, for they are about ten times larger than the unadjusted (incorrect) standard errors.

[Insert Tables A.1, A.2 and A.3 here]

#### D. Back-of-the-envelope i-mode calculation

Let us start by emphasizing that reliable price comparisons between WAP and i-mode are difficult due to differences in the pricing principles (i-mode's package based vs. WAP's two-part metered tariff) and especially due to lack of comparable data: From Kakkori (2001) we can infer that in Spring 2001, the monthly fee for the i-mode service was 2.73 euros. In addition, one packet, i.e., 128 bytes, cost 0.0027 euros. If a representative WAP connection in our data had involved a transfer of 2 kilobytes, the price of making the connection using i-mode would have been around 0.042 euros. Comparing the i-mode price to the prices used and obtained in our counterfactual calculations show that usage would have been higher than it actually was. If one took into account differences in service quality (see the main text), the increase in usage would correspondingly be larger.

#### E. Further robustness checks: Learning, consumption shifting and advertising

In Table A.4 we report reduced form estimations of Poisson models of connection counts, and within estimates of connection length, conditioning naturally on strictly positive connection length. Both specifications include consumer fixed effects and period fixed effects with period 1 as the base period. As can be seen from the table, the number of connections in one of the two post-experiment periods was smaller than in period 1 (the base period), and larger in the other. Although this doesn't strictly rule out experimentation, it does suggest that the consumers did not update their beliefs regarding the quality of their goods. This, coupled with the fact that the number of connections remains high through all three experiment periods suggests that learning or

experimentation is not a culprit: If it were the case that consumers initially experimented, they should have rather quickly realized that their priors hold, in which case the demand during the later parts of the experiment (period 5) should already mostly reflect “true” demand. The demands during the three experimental periods, while decreasing in time, are rather close to each other. While re-estimation of the model using only period 5 (i.e., excluding data from the first two experiment periods 3 and 4) data produces lower elasticities and higher consumer surpluses, the overall picture remains the same. E.g. using the linear model, the mean connection length price elasticity is 1.97 compared to 2.28, and the mean annual consumer surplus 21.89 compared to 16.40 when we do not utilize the (experiment) periods 3 and 4 in estimation.

[Insert Table A.4 here]

As to shifting consumption into the experiment period, the results do not support that alternative either as the coefficients of both the last pre-experiment and the first post-experiment periods are positive, indicating higher usage than during the base period (period 1). Finally, as the post-experiment number and length of connections is not systematically higher than in the pre-experiment periods, there is no evidence that advertising resulted in a shift in demand.

We repeated this analysis using a sample of individuals who early on reacted strongly to the pricing experiment. The reasoning is that if some individuals only learned of the experiment in the 2<sup>nd</sup> or 3<sup>rd</sup> two-week period, they might not have had enough time to experiment/learn. We thus computed for each customer the change in number of connections from the last pre-experiment to the 1<sup>st</sup> experiment period, and excluded all individuals who were below the mean (in the second variant, median) change. We then reran the reduced form fixed effects Poisson and length of call estimations including only the last experiment period and all the non-experiment periods in the estimation sample. The results, available upon request, are very similar to those reported above.

#### F. Results using data from plan “B” consumers

We repeat our empirical analysis using the data on the plan “B” subscribers. For brevity, we only note that cross-validation results suggest that the simplest model (Model 1 in Table A.2) suffices. The estimated latent demand remains strong and price elasticities high: The average number of needs per consumer is 14.36 per a two-week period. The average per-minute price ( $p_t$ ) elasticity of *connection length* ranges from -2.04 (log-linear demand) to -2.91 (linear demand). We also obtain large price elasticities of the number of wireless connections for the per-minute price and the fixed fee. For example, the per-minute price elasticity of the number of wireless connections is on aver-

age -1.78, and ranges from -0.82 to -3.20 for the linear demand. However the conditional probability of opening a connection is (again) small, on average about 9%. When translated into annual numbers, the per consumer surplus is on average 18.33 euros in the case of linear demand and 1.54 euros in the case of log-linear demand. These results are in line with the results we obtained using the data on the more popular plan.

#### G. Reduced form (consumer surplus) estimates

In our case, the relevant consumer surplus/CV -measure is the difference in the consumer's expenditure function between the expenditure at the market prices and at the service's virtual price, which is the price that sets the service's demand to zero. These expenditures are measured at the level of the utility received once the new service is on the market. Hausman (1999) shows that a practical way to compute the welfare gain is to use the approximation  $CV \approx -0.5pY\epsilon^{-1}$ , where  $Y$  is the quantity consumed,  $p$  is the price, and  $\epsilon$  is the own-price elasticity. In our case, direct application of the above approach is complicated for three reasons. First, what is  $Y$ ? Second, what is  $p$ ? Third, what functional form should we use to estimate  $\epsilon$ ? The standard log-log linear form is an option, but it cannot in our case be linearized conveniently by taking logarithms, because  $Y$  is frequently zero, as is the price during the experiment periods.

We proceed by assuming that  $Y$  is the number of connections made. To define  $p$  we compute the average of connection lengths over all consumer-period observations for which the length is positive, using data from the non-experiment periods only. We then take as the imputed price per connection the sum of the fixed connection fee plus the average outlay per connections, defined as the product of the per-minute charge and the average connection length. This imputed total price is 0.42 euros per connection. We then regress the number of connections on the imputed total price using a standard Poisson regression. Table A.5 reports the own-price elasticity and consumer surplus estimates derived from the reduced form regression that we report in the main body of the text. It also gives the underlying regression coefficients.

[Insert Table A.5. here]

With this benchmark reduced form estimate at hand, we perform a number of additional reduced-form estimations: First, we replicate the results of Table A.5 by estimating a standard linear model with ordinary least squares. The coefficient of the price variable is -115.70, which results in even lower consumer surplus. Second, we use a different imputed per connection price: If we simply take the "total" price to be the fixed connection fee, which is 0.09 euros, plus the per-

minute price, which in the more popular plan is 0.12 euros, we obtain an elasticity estimate that is clearly lower, about one fifth of that presented in Table A.5. The estimated consumer surpluses remain, however, negligible at 1.31 euros per consumer per annum. Finally, we allow for consumer heterogeneity: Estimating a fixed-effects Poisson (Hausman, Hall, and Griliches 1984) reinforces the finding of highly elastic demand: At the imputed total price of 0.42 euros per connection, the own-price elasticity is -10.83. Using the demographic variables as an alternative way to control for the heterogeneity produces very similar results. Taken together, these alternative welfare calculations echo the findings reported in the main text.

Table A.1. Cross-validation results

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
CONSTANT	X	X	X	X	X	X
AGE	X	X	X	X	X	X
GENDER	X	X	X	X	X	X
CITY	X	X	X	X	X	X
GENDER*AGE		X	X	X	X	X
CITY*AGE		X	X	X	X	X
GENDER*CITY		X	X	X	X	X
AGE^2			X	X	X	X
GENDER*AGE^2				X	X	X
CITY*AGE^2				X	X	X
AGE^3					X	X
GENDER*AGE^3						X
CITY*AGE^3						X
Cross-validation						
of exp(g(.))	2588.502	2585.69	2588.231	2590.941	2591.215	2601.392
of k(.)	110.1632	110.1178	110.0678	110.0267	109.8905	110.023

Table A.2. Estimation results from the first step ( $\mu$  and  $\lambda$ )

	WAP_COUNT ( $\mu$ )		CALL_DUR ( $\lambda$ )	
	Coefficient	Std. error*	Coefficient	Std. error*
AGE	0.010	0.006	-0.069	0.036
GENDER	0.112	0.291	-0.051	0.325
CITY	0.404	0.268	-0.609	0.313
GENDER*AGE	-3.8E-03	6.5E-03	-7.3E-03	1.6E-02
CITY*AGE	-0.017	0.006	0.025	0.017
GENDER*CITY	0.046	0.161	-0.028	0.077
AGE2	-	-	1.5E-03	7.9E-04
GENDER*AGE2	-	-	1.7E-04	2.0E-04
CITY*AGE2	-	-	-3.1E-04	2.2E-04
AGE3	-	-	-1.1E-05	5.4E-06
CONSTANT	2.133	0.277	2.814	0.522
Obs.	5460		4350	
Wald (joint significance)	36.91		67.61	
d.f.	6		10	
p-value	0.000		0.000	
Log-likelihood	-66434.57		-5555.11	

\*Huber-White heteroscedasticity robust covariance matrix, adjusted for clustering within consumers

Table A.3. Estimation results from the second step ( $\gamma$ )

Model	Linear demand			Log-linear demand		
	Coefficient	Std. error*	Std. error**	Coefficient	Std. error*	Std. error**
AGE	-1.083	0.035	0.381	-0.096	0.004	0.037
GENDER	-41.497	2.073	17.329	-3.796	0.197	1.487
CITY	-5.828	2.398	19.944	-0.912	0.220	1.759
GENDER*AGE	0.965	0.043	0.412	0.090	0.005	0.039
CITY*AGE	-0.263	0.043	0.379	-0.032	0.005	0.040
GENDER*CITY	-8.206	1.354	11.469	-0.296	0.146	1.077
CONSTANT	102.003	1.825	16.204	20.068	0.178	1.409
Obs.	12740			12740		
Log-likelihood	-85963.512			-86007.262		

\*unadjusted standard error

\*\*bootstrap standard error

Table A.4. Reduced form estimation of connection count and call length (cond. > 0)

Period-dummy	Count		Length	
	Coefficient	p-value	Coefficient	p-value
t = 2	0.277	< 0.001	0.142	0.593
t = 3	2.652	< 0.001	2.783	< 0.001
t = 4	2.341	< 0.001	2.653	< 0.001
t = 5	2.25	< 0.001	2.436	< 0.001
t = 6	0.166	0.001	-0.411	0.119
t = 7	-0.085	0.011	0.139	0.606
Consumer fixed-effects	YES		YES	
Tests of equality of period coefficients (p-values)				
2 vs. 6	< 0.001		0.028	
2 vs. 7	< 0.001		0.992	
6 vs. 7	< 0.001		0.025	

Table A.5. Reduced form Poisson estimation and consumer surplus

	Own-price elasticity	Annual per-consumer surplus (euros)
At the imputed price:	-10.84	0.56
	Price-variable	Constant
Coefficient	-25.89	10.95
Standard error	0.85	0.29