Problems in launching the mobile internet:
Evidence from a pricing experiment

Maija Gao, Ari Hyytinen and Otto Toivanen

PROBLEMS IN LAUNCHING THE MOBILE INTERNET:
EVIDENCE FROM A PRICING EXPERIMENT[*]

Maija Gao
Nokia Corporation

Ari Hyytinen
University of Jyväskylä and Yrjö Jahnsson Foundation

Otto Toivanen
Katholieke Universiteit Leuven and
The Research Institute of the Finnish Economy (ETLA)

Abstract

Commercialization of innovations frequently stumbles. A prominent recent example is 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 pricing experiment. The experiment allows 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. We find that pricing reduced usage substantially and explore potential reasons for the high prices.

JEL: L11, D12

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

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).[1]

Wireless Internet generated high expectations that materialized in the hundreds of million euros 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.[2] Soon after its launch, some commentators announced WAP a

---

[1] The wireless technologies currently in use are much more advanced than the first version of WAP: WAP was a first-generation technical standard to provide wireless services that was introduced in Europe in 1999.

[2] 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, were also introduced subsequently in the U.S..

failure.[3] While the current views of WAP clearly challenge the most critical accounts,[4] the question of why the early take-up stumbled still remains. This question has more than historical relevance as any new product launch 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 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 use data from a pricing experiment[5] targeting actual customers of a mobile phone operator, implemented in Fall 2001, and a structural model of demand. Owing to the experiment, we observe the prices charged and the quantity demanded (i.e., the number and average length of connections) for a panel of consumers both during four non-experimental and three exper-

---

[3] 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 at the time 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).

[4] Many see WAP as a technological standard that developed over time and that provided a bridge to newer generations of wireless technologies. It was upgraded (to WAP 2.0) in 2002, which resulted in resurgence around 2003-2004, especially in Europe. For example, the number of WAP page impressions viewed in the UK more than doubled during the nine months prior to May 2003 (the Mobile Data Association, the U.K.). In Finland, usage grew at the time similarly.

[5] The pricing experiment was based on a promotional campaign. Therefore, it was not a natural field experiment.

imental two-week periods. Prior to and after the experiment, prices were at their normal (equilibrium) levels of the time. 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.[6] The operator had two plans which had the same connection fee (0.09 euros), but different per minute prices (0.12 for plan A, 0.17 for plan B) for wireless. One might wonder why anybody would choose plan B given that plan A seems to dominate it. The answer lies in the fact that the vast majority of consumption and revenue during our observation period was from voice services, in which the two plans differed, rather than wireless services. 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 proba-

---

[6] 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 wireless services. At the time of the experiment, 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.

bility as functions of those characteristics.[7] We identify the parameters of our model using variation that the experiment induces and estimate them using a flexible two-step m-estimator (e.g. Newey 1994a). The 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. The story that emerges is that it wasn't (just) the technical features of the new product that lead to its failure in the market place, but that high prices curtailed the demand for this new good.

This key result of ours calls for an investigation as to why prices were so high. We first estimate marginal costs through a structured survey of industry experts in which we (following, e.g., Bajari and Ye 2003, and O'Hagan 1998) elicit beliefs about the marginal costs from a number of industry experts. This evidence suggests that true marginal costs were likely significantly lower than those that would rationalize the actual prices even under a set of conservative assumptions.

The natural next step is an exploration of whether the firm made a mistake in pricing the WAP services. Given that we are studying a good that was new in nature not only to customers, but also to the firm, it is not inconceivable that such a mistake could have happened. To study this, we compare actual prices to those that a static monopolist would have set. A static monopolist is a convenient benchmark as most (if not necessarily all) alternative models, such as dynamic oligopoly mod-

---

[7] Demand estimates for new services and goods are often wrought with empirical difficulties. Lack of detailed data is a primary reason. Bajari and Benkard (2005) and Berry, Linton and Pakes (2004) spell out some methodological difficulties.

els, would yield lower equilibrium prices. While we cannot be conclusive on this, our results, especially those obtained by our preferred demand specification, suggest that optimal (profit-maximizing static monopoly) prices would have been lower than those actually charged. Consistent with this, we find that neither the consumers nor the producer would have suffered from a surplus loss had the prices been lower (profit-maximizing). We also find that the estimated profit-maximizing prices would have delivered 70% of the potential total surplus.

To conclude, high prices – whether optimal or not – appear to be a key reason for why the early take-up of the WAP enabled services stumbled initially. Taken together, our survey evidence and counterfactual analysis speak for the possibility of a pricing mistake, but we discuss below other explanations for the observed high prices.

The remainder of the paper is divided into five sections. In Section 2 we discuss the technology, its relationship to the previous and current 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 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 in Section 5. While our data have some clear strengths, it also has some weaknesses that are due to it being generated from a promotional campaign: There are alternative dynamic stories that could explain the higher usage during the experimental periods, such as the experiment being anticipated, experimentation by consumers, advertising effects on demand, consumers shifting demand in time and our specification choices. We perform a number of robustness tests regarding these alternative explanations in Section 5. We

find that our results are robust to these tests. 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 first-generation wireless Internet. The services were accessed using a mobile phone that relied on WAP and that allowed the user to connect to the Internet. [8] The Finnish operator whose data we use launched its WAP-based wireless Internet service in late 1999. At the time of the experiment in 2001, the operator was one of the two leading firms in a market with four operators and its own service portfolio consisted of 67 services, ranging from news, sports results and weather services to games, betting results and TV-listings.[9] These services were tailored for a wireless user. The users also had access to external Internet sites.[10]

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 leas-

---

[8] 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.

[9] Kakkori (2001) provides a complete account of the types of services that where 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.

[10] 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 much more limited than what are available today. 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 slow. The display sizes were also small. For example, one of the most often used terminals had a display with 96x60 pixels.

ing of handsets, because a Finnish law prohibited the practice.[11] 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.[12] From a production technology point of view, zero marginal costs seem like a good starting point. However, access fees paid by the operator to owners of some antennas and usage charges for content also paid by the operator imply that the marginal costs may have been nonzero (see Sections 4.2 and Appendix for more detail).

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. 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. This advertising campaign raises the question of whether demand during the experiment periods could have been affected by it, thereby biasing our results. We explore this in Section 5.2 below, but find little evidence for advertising having shifted demand. After the campaign, the prices returned to their previous levels.

---

[11] 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 655 000 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.

[12] During 2001, about 60% of the operator sponsored wireless services had an additional contents charge (Kakkori, 2001; see pp. 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.

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.

3.1   Step 1: Demand side analysis

The theoretical model of demand

We use 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.[13] These features are important, because 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 $V(.)$ is concave and twice continuously differentiable in q, and α is a demand parameter explained below.[14]

We introduce two types of demand shocks that determine observed behavior: 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 es-

---

[13] Dubin and McFadden (1984) is an early example of a discrete-continuous demand model to which class our model also belongs.

[14] This is a commonly used transformation both in the Industrial Organization (for its derivation, see e.g. Tirole 1988, pp. 143) and the telecommunications literatures (e.g. Mitchell, 1978).

tablish 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 can e.g. imagine that there is cross-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, α. We assume that $V(q,\alpha)$ is increasing in α and that this shock 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, α, 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 α 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) $q(p,K,\alpha) = \arg\max\left[V(q,\alpha) - (K + pq)\right]$

if consumer surplus gross of the fixed cost is larger than the fixed cost, i.e., $CS(p,\alpha) \equiv \int_p^{\bar{x}} q(x;\alpha)dx \geq K$ and zero otherwise.[15] 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

---

[15] The upper bound of integration, $\bar{\tau}$, depends on the demand function.

wireless service connection with probability $\pi(p,K) = \left[1 - F(\bar{\alpha}(p,K))\right]$, where $\bar{\alpha}(p,K)$ is the unique value of the shock that is implicitly defined by $CS(p, \bar{\alpha}(p,K)) \equiv \int_p^{\bar{x}} q(x;\alpha)dx = K$. It follows that for a sequence of connection needs, the number of connections established follows a binomial distribution (see below).

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:[16]

Assumption 2: The number of connection needs during a given time interval, *W*, is distributed Poisson with mean μ.

Assumption 2 means that the expected number of consumption needs for wireless services during a time interval is μ. 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* ~ Poisson with $E[Y|p,K] = \mu\pi(p,K)$.[17]

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 α 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 α, 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

---

[16] See e.g. Hayes and Ganesh Babu (2004).

[17] 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.

with probability $\pi = 1 - F(\bar{\alpha}(p, K))$, and the number of connections is thus the result of consumer optimization.

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 [μ], the density of the demand shift shock [$f(\alpha)$], and the form of demand determining the optimal connection length [$q(p)$].

Satiation demand is typically unobservable. We can however measure it, because during the experiment periods (periods $t = 3, 4, 5$) $p_t = K_t = 0$. The assumptions we make (see below) on the demand function ensure that $\bar{\alpha}(0, 0) = 0$ and $\pi_{it} = 1 - F(0)$ 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 the number of observed connections, Y_{it}, is equal to the number of connection needs W_{it} during the experiment periods. By Assumption (2), W_{it} is distributed Poisson with mean μ. We allow for heterogeneity and model the mean satiation demand of consumer i flexibly as a function of demographics. We seek to exploit our data, which do not contain too many consumer characteristics, as fully as possible, by letting all parameters of the model be functions of all observed consumer characteristics. We assume, specifically, that $\mu_i = \exp(g(\underset{\sim}{z_i}; \underset{\sim}{\mu}))$, where g($\underset{\sim}{z_i}, \underset{\sim}{\mu}$) is an initially unknown function of the vector of (all observable) consumer demographics $\underset{\sim}{z_i}$ and the associated vector of parameters $\underset{\sim}{\mu}$.

The theory is silent about the distribution of the demand shift parameter α, which assumes a different value for each need that consumer i faces during period t. Following the telecommunications engineering literature, we specify that α has a (stationary) exponential distribution. However, we allow for heterogeneity as

11

follows: $\alpha_{ij} \sim \exp(\lambda_i)$, with $E[\alpha_{ij}] = 1/\lambda_i$, where j indexes connection needs ($j =$

0, ..., W_{it}) of consumer i during period t. The mean depends on consumer de-

mographics, i.e., $\lambda_i = k(z_i, \underset{\sim}{\lambda})$, where $k(z_i, \underset{\sim}{\lambda})$ is an initially unknown function of

the vector of all consumer demographics z_i and the associated vector of parame-

ters $\underset{\sim}{\lambda}$.

Absent an established practice, we consider two different demand specifica-

tions for $q(p)$. The demand for the service (i.e., the length of the connection) is

therefore initially allowed to be 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 γ_i as a

function of consumer demographics ($\gamma_i = \underset{\sim}{\gamma}' z_i$) and that they have been used ex-

tensively in the prior analyses of telecommunications demand.[18] Consistent with

our data, they also allow for a point of satiation. Despite both demand functions

having these merits, it turns out that the log-linear specification fits our data clearly

better (in terms of an out-of-sample test; see section 5.1), provides more precise

estimates for the economic quantities of interest (see Table 4 below) and gives

results that are quantitatively closer to those that we obtain from a reduced form

analysis (see section 5.2). We therefore emphasize more the results from the log-

linear specification in the discussion that follows. However, to allow for a trans-

parent and balanced treatment, we present the results from both specifications side-

by-side in the tables and avoid drawing strong conclusions in cases where the two

specifications diverge in their results.

---

[18] For studies using the linear demand function, see for example Miravete (2002) and Miravete and Röller (2004). The log-linear demand is known as "Perl-demand" in the telecommunications litera-ture; see Taylor (2002) for a review of this literature.

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 $E\left[Y_{it}|p_t, K_t, z_i\right] = \mu_i \pi_{it}$. However, when estimating the model parameters we only make use of the moment condition given by the conditional mean.[19] The conditional mean can be re-written as

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

where $\bar{\alpha}(p_t, K_t) = (\gamma' z_i)p_t + \sqrt{2(\gamma' z_i)K_t}$ in the case of linear demand and $\bar{\alpha}(p_t, K_t) = (\gamma' z_i)K_t \exp((\gamma' z_i)p_t)$ in the case of log-linear demand. Our theoretical model imposes the restriction that the coefficients of the per-minute price and the connection fee are the same.[20]

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,

---

[19] That is, we use only the first moment and do not impose the potentially very restrictive Poisson density or variance assumption; see, for example, Gourieroux, Montfort and Trognon (1984a,b), and Wooldridge (1997). While our data would allow us to estimate a negative binomial count model, it would not bring any further identification power for the structural parameters of our model, as we do not need the second moments (e.g. variance parameters) to calculate our economic objectives of interest. This notwithstanding, we did estimate a negative binomial (type I) model. The point estimates were almost identical to those obtained with the Poisson specification and the differences were not statistically significant. The estimated variance parameter was 2.181 (s.e. 0.043). A consequence of these results is that using a negative binomial would not materially change our counterfactual results.

[20] Notice that in our model, the price elasticities w.r.t. p and K can still differ when we impose this restriction. One might consider relaxing this standard theoretical restriction. Lack of sufficient (inter-temporal or cross-sectional) price variation means that in practice, it is not feasible to relax this restriction in our data.

and whatever the realization of the demand shock, a connection was established. To make full use of this identifying information we adopt a two-step *m*-estimator in which 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 key benefit of this sequential approach is that we can estimate μ_i and λ_i very flexibly as a function of all the demographics that we observe in our data.

The first step identifies and estimates the latent demand as follows: First, because $W_{it} = Y_{it}$ (i.e., number of connections is equal to the number of needs for a connection) for the experiment periods $t = 3, 4, 5$, variation in the number of connections equals variation in the number of needs. This variation identifies $\mu_i = \exp(g(z_i, \mu))$. We run regressions of the form

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

for $t = 3, 4$ and 5, and estimate $g(\underset{\sim}{z_i}, \mu)$ 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 $\underset{\sim}{z_i}$ consist of consumers' age, gender and place of residence, and their powers and interactions.

The other parameter of latent demand is $\lambda_i = k(\underset{\sim}{z}_i, \underset{\sim}{\lambda})$, which determines the mean of the demand shift shocks. Variation in connection lengths across consumers during the experiment identifies $\lambda_i = k(\underset{\sim}{z}_i, \underset{\sim}{\lambda})$, 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) $\quad E(q_{it} \mid p_t = 0, K_t = 0, z_i) = 1/\lambda_i = 1/k(\underset{\sim}{z}_i, \underset{\sim}{\lambda})$,

for $t = 3$, 4 and 5. To estimate $k(\underset{\sim}{z}_i, \underset{\sim}{\lambda})$ flexibly we use a power series estimator and cross-validation, where $\underset{\sim}{z}_i$ includes, as before, all observable consumer characteristics. 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 identify $\underset{\sim}{\gamma}$ from the cross-sectional demand responses of consumers (with plan A; see section 4.1) induced by the two price changes in the data (into and out of the experiment period). This we do by estimating the price effects, parameterized by $\underset{\sim}{\gamma}'\underset{\sim}{z}_i$, 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) $\quad E[Y_{it} \mid p_t, K_t, z_i] = \hat{\mu}_i \left[1 - \hat{F}(\overline{\alpha}(p_t, K_t)) \right] = \hat{\mu}_i \exp(-\hat{\lambda}_i \overline{\alpha}(p_t, K_t))$.

where $\overline{\alpha}(p_t, K_t)$ is a function of $\underset{\sim}{\gamma}'\underset{\sim}{z}_i$. 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 $\underset{\sim}{\gamma}$, we make use of the moment 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.

We acknowledge that one could use a GMM estimator that uses data from all periods and estimates simultaneously all model parameters (instead of doing it sequentially, as outlined above). While this would potentially be a source of efficiency and perhaps a way to allow for correlation between the shocks in the number of needs and their intensity, this approach would bring some non-negligible costs with it.[21] Besides complexity, it would not lend itself to a nonparametric estimation of the unknown functions in the way that the sequential estimation approach we adopt does.

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 from the producer's point of view.

To explore whether the prices were set (sub)optimally relative to a static monopoly 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 static monopoly pricing, to estimate the

---

[21] We leave it for future work to build a discrete-continuous model that allows explicitly for a correlation between the shocks in the length of connections and the number of connection needs. If the shocks are positively correlated, the parameters of demand intercept are likely to be biased toward zero, whereas the direction of the bias, if any, in the slope parameter is unclear. At the request of a referee, we calculated the correlation coefficient between the two shocks in our data, using the uncensored connections from the experiment periods. The correlation is 0.13, with a p-value of 0.00. The correlation is not alarmingly high (i.e., not high enough to distort the computation of optimal prices and surpluses), but it does suggest that those customers who have more needs also have longer connections.

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 online 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 have been too high, the static monopoly assumption is most likely 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 online Appendix.

The second part of our approach, inspired by e.g. O'Hagan (1998), Bajari and Ye (2003) and Manski (2004), makes use of the results of a structured survey that we administered to acknowledged industry experts to elicit their beliefs about the marginal costs. 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, engineers are typically not familiar with the economic

concept of marginal cost.[22] 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 give informative answers. We were able to elicit 18 responses with a 50% response rate (for details, see part B of the online Appendix).

The third part of our approach is to compare the marginal cost estimates of the industry experts with the estimates from our model.

4 Empirical analysis

4.1 Data

The sample of consumers 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 weeks, it may still be possible that not all (potential) users are included in our data.[23] 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.[24] Second, if one drops the maintained assumption of establishing a connection at zero prices given a need, we might miss

---

[22] Indeed, of the 36 experts that we approached, half declined to answer our questionnaire after having read the questions, citing as a reason that they felt they did not have the necessary expertise or information to provide reliable answers.

[23] During our observation period, mobile phone customers in Finland bought 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.

[24] 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: Making the Poisson assumption and using our mean latent demand estimate, the probability of having zero needs during the 2nd 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.

customers with high opportunity costs. Such selection would bias our results to-wards too low price elasticities.[25]

Table 1 displays the descriptive statistics for the sample, separately for the non-experiment and experiment periods, as well as the basic consumer de-mographics. We have a balanced panel with 14 882 consumer-period observations ($N = 2\ 126$ and $T = 7$). Table 1 shows that the experiment increases usage dramati-cally. During the non-experiment periods, the average number of wireless connec-tions (WAP_COUNT) per a two-week period is 1.13. The average connection length (CALL_DUR) is 2.66 minutes. During the experiment periods, the average number of needs (= number of connections during the experiment periods) is much higher, 11.94 per a two-week period. The average satiation connection length (= length of connection during the experiment period) also is longer, 5.54 minutes. There is, however, a lot of variation.

Table 1 shows in addition 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 connec-tion fee (WAP_K) is 0.09 euros per connection. While there is limited temporal price variation in our data, the rich cross-sectional variation in consumer character-istics will be helpful in identifying the effect of consumer characteristics on re-sponses to price changes. This we capture flexibly in estimation, as explained above. We do not resort to using the data from the two plans together even if that would give us cross-sectional variation in prices. The reason is that we want to

---

[25] 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, 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. However, we cannot completely rule out the possibility that some individuals may have changed operators to take advantage of the experiment. These would presum-ably 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 esti-mates of the number of needs and our price elasticity estimates upwards.

avoid any selection bias that this might induce (as selection into the plans is en-dogenous). In the empirics that follow, we thus use data from the larger plan A only. Our results are however robust to using either or both plans (see section 5.2).

[Insert Table 1 here]

It also is the case that any differences in behavior before and after the experiment are small. As an example, for plan A customers, the average conditional connection length before (after) the experiment was 2.72 (2.73) minutes. Below (see section 5.2) we demonstrate that this 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: The first step cross-validation of $\exp(g(z_i, \mu))$ and $k(z_i, \lambda)$, and the estimation results for $\hat{\lambda}_i$, $\hat{\mu}_i$, and $\hat{\gamma}_i$ are presented in part C of the online 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 reported. 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.[26] 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.[27]

Figure 1 shows the age profile of latent demand ($\hat{\mu}_i$) and the conditional probability of opening a connection ($\hat{\pi}_t$), given a need.[28] 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 more than 1.5 times as many connection needs as a 70 year old, but his probability of connection was only two thirds of that of the 70 year old.

[Insert Figure 1 here]

---

[26] 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.

[27] The estimated price elasticities, evaluated at the market price, are high, too, providing further support for this conclusion. A key difference between the linear and log-linear specifications is that the former implies the length of the connections has a high elasticity, whereas the latter implies that it is the number of connections that is price sensitive. To be more specific, we find that the average per-minute price (p_t) elasticity of connection *length* is high, ranging from -1.92 (log-linear demand) to -14.69 (linear demand). The latter is perhaps too high to be credible, and indeed, in an out-of-sample test that we perform later (see section 5.1), we find that the linear specification predicts far too short calls. The per-minute price (p_t) and the fixed fee (K_t) elasticities of the *number* of wireless connections appear to be plausible for both specifications: For the linear (log-linear) demand, the average per-minute price elasticity of the number of wireless connections is -1.70 (-3.44). The average fixed fee elasticity of the number of wireless connections is -0.34 (-1.78) for the linear (log-linear) demand. Note also that our data allow us to uncover short-run elasticities only and that the long-run elasticities may be even higher.

[28] The plot is drawn for the log-linear demand specification using data for males that live in the Helsinki capital area. The pattern for linear demand is similar.

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. These results raise the question of whether prices were optimally set.

Step 2: Supply side implications

As we discuss in more detail in the Appendix, there are a priori reasons to think that the (true) marginal costs are nonzero.[29] We therefore report the marginal costs implied by the model and the (mean and median) marginal cost estimates of the industry experts in Table 3, and compare the two sets of marginal cost estimates.[30] The logic of the comparison is that due to our monopoly assumption, the model gives us *lower bound* estimates of the marginal costs to which the elicited marginal costs can be contrasted.

As can be seen from the Table, the linear model implies somewhat lower marginal costs than the log-linear models: The linear demand model yields an estimated marginal connection cost (C) of 0.07 euros, while that given by the log-linear model is 0.08 euros. As to the marginal cost of connection length (c), the linear (log-linear) model yields an estimate of 0.06 (0.11) euros. These estimates are higher than the elicited marginal costs: The average (median) expert opinion point-estimate of C is 0.02 (0.01) euros and that of c 0.01 (0.01) euros.[31] 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 re-

---

[29] For example, according to the expert we interviewed, the operator had to pay access fees for some connections, and usage fees for some services.

[30] No standard errors for the estimated marginal costs can be provided due to lack of variation in prices during the non-experiment periods.

[31] 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 online 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.

spondents is 0.02 (0.01) and 0.02 (0.01) euros. It is worth emphasizing that also the *maximum* point estimates from the survey ($C^{max} = 0.05$ and $c^{max} = 0.04$ euros) are lower than the model-based estimates.[32]

We can make use of the elicited marginal cost data and the lower bound estimates 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 lognormal distribution and taking the mean and standard error estimates from Table 3 as its first two moments, we find that these probabilities are all well below 5%. This comparison supports the conclusion that the marginal costs implied by both demand models are too high compared to our benchmark static monopoly model.[33] These results suggest, but do not conclusively show, that sub-optimal prices may explain the commercial failure of the early WAP services. They also suggest a number of counterfactual experiments.

[Insert Table 3 here]

4.3   Counterfactual experiments and surplus calculations

We perform three counterfactual experiments. As the first experiment, we infer profit-maximizing 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. The latter two experiments are motivated by possible regulatory intervention. The third one provides also a welfare benchmark. We focus on usage and welfare implications. We use the average marginal costs calculated from the survey respondents' marginal cost distribu-

---

[32] 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.

[33] 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.

tions as these are the highest and thus for our purposes the most conservative estimates.[34]

[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. One should note that our consumer surplus estimates - like those of any exercise of this type (see Petrin 2002) - will be potentially sensitive to us having to integrate over all possible prices, some of which are a lot higher than those observed in the data.

We start by emphasizing the *qualitative* results that do not depend on whether the log-linear or linear specification is used. First, it seems that static monopoly profit-maximizing prices would have been lower than those actually charged. Second, we find that going from actual to the monopoly prices increases the connection probability, producer surplus and consumer surplus in both specifications. Thus, neither the consumers nor the producer would have suffered from a surplus loss had the prices been lower. In particular, both specifications indicate that consumer surplus would have been quite a bit higher had the prices been at their lower, profit-maximizing level. Third, profit-maximizing prices would have delivered about 70% of the potential total surplus (i.e. of those generated by marginal cost pricing).

---

[34] 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 major increase in the number of connections. 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).

The results from both specifications are thus in the same qualitative direction against our conservative benchmark of static monopoly pricing. Most importantly, they support the view that pricing of the services had a strong effect, curtailing usage. However, as Table 4 shows, the magnitudes differ in some cases significantly between the two specifications. In general, the results obtained by our preferred (log-linear) demand specification suggest larger differences between the actual and optimal pricing.[35] For example, the log-linear specification implies that usage, as measured by the connection probability, would be as much as five times higher had the prices at the level charged by a static monopoly. The corresponding number from the linear model is smaller, 1.8. The larger differences of the log-linear model are not just noise, as the reported (bootstrap) standard errors show: All increases suggested by the log-linear model can be measured relatively accurately.

One has to exercise some caution in the interpretation of these counterfactuals for two different reasons: On the one hand, we need to extrapolate far outside the support of observed prices and our consumer surplus results in particular may hence be driven by functional form assumptions. On the other hand, there is the possibility that our results may be influenced by us observing zero prices. There is some evidence (see e.g. Ariely, Gneezy, and Haruvy 2006 Shaman'er and Ariely 2007) that consumers react differently to zero as opposed to other low prices. The Ariely, Gneezy and Haruvy experiment suggests that we may underestimate the price elasticities, while the Shaman'er and Ariely experiment suggests the opposite.

---

[35] Below we report the results of an out-of-sample test that suggest that while neither model is perfect, the prediction error of the log-linear model is significantly smaller than that of the linear model, and of opposite sign.

5   Interpretation and robustness

5.1   Interpretation of the empirical evidence

Choosing between demand specifications

To explore which of the two demand specifications better fits the data, 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 6.57 (-.48) minutes for the linear (loglinear) model, both significant at the 1% level. The difference in the (mean) squared prediction error (between the two models) is 43.83 with p-value 0.00, i.e., the linear model's prediction error is systematically larger.[36] Thus, while the loglinear model slightly underestimates the call length, it produces clearly better out-of-sample estimates than the linear model. This finding suggests, as we have argued already earlier, that we should emphasize the results obtained from the loglinear model over those from the linear model. This conclusion is also supported by a comparison to price elasticities obtained from reduced form models (see part G of the online Appendix.

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

---

[36] The mean squared prediction errors are 53.94 (linear) and 10.11 (loglinear).

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 that quality-adjusted prices were high (see part D of the online Appendix).[37] A back-of-the-envelope price comparison 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 closer to the price level of i-mode.

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 may have been sub-optimally high relative to static monopoly prices. *If* this was the case, they leave open whether the operator systematically failed to maximize its profits when setting its prices or whether we have identified an isolated pricing mistake.

Speculating on the possibility of a 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 formed its main revenue source at the time of the experiment. It thus seems that it knew how

---

[37] Most of the descriptive details in this paragraph are from Kakkori (2001).

27

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 could have 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 5.2) also support this conjecture: Even with the benefit of hindsight, industry experts had widely differing opinions on how elastic the demand for the early WAP services was.

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 (2009) study the effects of price search on the Internet on demand elasticity, and document Internet price

28

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, according to our estimates, the (realized) per-consumer surplus from the early wireless services appears to be a minor fraction of the gains that satellite TV channel buyers experienced in the U.S. (Goolsbee and Petrin 2004).[38] 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.

5.2   Robustness tests

To verify the robustness of our results, we have performed a large number of robustness tests on various aspects of our analysis. We report these briefly in the Appendix (for more detail, see the online Appendix).

First, we looked for alternative explanations, such as learning, advertising and demand shifts, for the existence of latent demand and the high estimated price elasticities and found that the robustness tests support our modelling. Second, we

---

[38] All these numbers are naturally based on short-run elasticities.

utilized the plan B data to see whether our results are sensitive to using just data on plan A customers; we found this not to be the case. We then constrained the model in various ways, and estimated reduced form models to provide a benchmark for our price elasticity results. These results were in line with those from the main analysis.

In addition to these demand side robustness tests, we resorted to interviews and our survey to consider the robustness of our supply side analysis. The survey allowed us to verify our marginal cost estimates, and to compare our pricing results and elasticities to the (necessarily subjective and admittedly ex-post) views of the interviewed experts. These robustness tests also supported our results. For example, we found that experts mostly agreed that the prices were too high.

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 early mobile Internet were plenty, and thus that lack of latent demand cannot explain the sluggish initial adoption. It seems that prices were a main impediment of diffusion and, in particular, that the pricing of the services had a curtailing effect on usage. Consistent with this, we find that the realized consumer surplus to an (early) adopter of these services were on average moderate. This interpretation is

consistent with the early adoption experiences and critical accounts that were aired at the time our experiment was run, and with Grajek and Kretschmer (2012) whose results suggest that price plays a key role in achieving critical mass.

We cannot however draw strong conclusions as to *why* the prices were so high that they curtailed the usage and consumer surplus. First, our demand elasticity estimates may be too high for a number of reasons, such as demand shifts due to exogenous reasons and/or advertising shifting demand. Inelastic demand would be consistent with high prices and thus with the view that the slow initial take up is just an incarnation of the standard deadweight loss in the presence of market power. In robustness tests, we however find little support for biases in our demand estimates or for demand shifts. It is also possible that our results are partly driven by the prices being zero in the experiment periods; earlier research has shown that (changes to) zero prices may induce different behavior from small but strictly positive prices.

Another possible interpretation is that the prices were sub-optimally high and thus that the firm made a pricing mistake: Both our counterfactual analysis as well as the structured survey of industry experts that we conducted for this paper support this interpretation, as they 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. We thus conclude with the conjecture that had the prices of the early of mobile service been lower, it would not have compromised the producer's profits from the new service. Instead, the lower prices would have speeded up the early adoption.

The larger issue that our paper raises 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. A better understanding of how to handle com-

31

mercial risks, such as those arising from pricing, could significantly raise the social

and private returns to innovation.

References

Bajari, Patrick and Benkard, C. Lanier, 2005, Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach, *Journal of Political Economy*, 113(6), 1239-1276.

Bajari, Patrick and Ye, Lixin, 2003, Deciding Between Competition and Collusion, *Review of Economics and Statistics*, Vol. 85, No. 4, pp. 971-989.

Berry, Steve, Linton, Oliver and Pakes, Ariel, 2004, Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems," *Review of Economic Studies*, 71(3), 613-654.

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, Jeffrey and McFadden Daniel, 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., 2009, Search, Obfuscation, and Price Elasticities on the Internet, *Econometrica*, 77 (2), 427-452.

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, Christian, Montfort, Alain and Trognon, Alain, 1984a, Pseudo Maximum Likelihood Methods: Theory, *Econometrica*, Vol. 52, No. 3, 681-700.

Gourieroux, Christian, Montfort, Alain and Trognon, Alain, 1984b, Pseudo Maximum Likelihood Methods: Applications to Poisson Models, *Econometrica*, Vol. 52, No. 3, 701-720.

Grajek, Michal and Kretschmer, Tobias, 2012, Identifying Critical Mass in the Global Cellular Telephony Market, *International Journal of Industrial Organization*,Vol. 30, 496-507.

Hayes, Jeremiah F and, Thimma V.J. Ganesh Babu., 2004, Modeling and Analysis of Telecommunications Networks, Wiley Interscience.

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, Ker-Chau., 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

Manski, Charles F., 2004, Measuring Expectations, *Econometrica*, Vol. 72, No. 5, pp. 1329-1376.

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, 2004, Estimating Markups under Nonlinear Pricing, *Journal of the European Economic Association*, 2, 526-535, April-May 2004.

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.

O'Hagan, Anthony, 1998, Eliciting Expert Beliefs in Substantial Practical Applications, *Statistician*, Vol. 47, Part 1, pp. 21-35.

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.

Rouvinen, Petri and Ylä-Anttila, Pekka, 2003, Case Study: Little Finland's Transformation to a Wireless Giant, in S. Dutta, B. Lanvin and F. Paua (Eds.), *The Global Information Technology Report 2003-2004*, New York: Oxford University Press, pp. 87-108.

Shampan'er, Kristina and Ariely, Dan, 2000, How Small is Zero Price? The True Value of Free Products, *Marketing Science.* Vol. 26, No. 6, 742 - 757.

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, 2007, Do Switching Costs Make Markets More or Less Competitive?: The Case of 800-Number Portability, *RAND Journal of Economics*, Vol. 38, Spring, 146-163.

Table 1. Descriptive statistics

Non-experiment periods 1-2 and 6-7	Obs: N*4	Mean	S.d.	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	S.d.	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	S.d.	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

	Log-linear demand		Linear demand	
	Mean	S.d./S.e.	Mean	S.d./S.e.
Expected number of needs (μ)				
- per two-week period	11.54	1.49	11.54	1.49
- per year	300.01	38.61	300.01	38.61
Prob(connect / need) (π)	0.10	0.04	0.10	0.04
Expected # of connections established ($\mu\pi$)				
- per two-week period	1.13	0.37	1.13	0.41
- per year	29.32	9.65	29.33	10.78
The price elasticity of:				
- wireless connection length				
w.r.t. per-minute price (p)	-1.92	0.02	-14.69	0.93
w.r.t. fixed connection fee (K)	-	-	-	-
- # of wireless connections				
w.r.t. per-minute price (p)	-3.44	0.14	-1.70	0.06
w.r.t. fixed connection fee (K)	-1.77	0.05	-0.34	0.01

Table 3. Marginal cost estimates

Marginal cost of	Model		Expert opinion (survey data)				
	Log-linear	Linear	Aver. 1*	Aver. 2**	Med. 1*	Med. 2**	Std.err.*
- Establishing connection, C	0.08	0.07	0.02	0.02	0.01	0.01	0.02
Prob[C > Model C]***	0.01	0.02	-	-	-	-	-
- Lengthening connection by 1 min., c	0.11	0.06	0.01	0.02	0.01	0.01	0.01
Prob[c > Model c]***	0.01	0.02	-	-	-	-	-

* based on elicited point estimates (survey data)

** based on elicited distribution estimates (survey data)

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

Table 4. Counterfactual experiments

	Connection probability (π)	Average connection length (min)	Producer surplus (euro)	Consumer surplus (euro)
Log-linear model				
Actual prices (p = 0.12, K = 0.09)	0.10	2.61	7.49	1.20
s.e.	(0.01)	(0.02)	(0.57)	(0.16)
Optimal prices (p = 0.066, K = 0.069)	0.55	3.25	21.48	17.92
s.e.	(<0.01)	(0.05)	(0.75)	(0.65)
No connection fee (p = 0.087, K = 0)	1.00	1.34	5.75	22.89
s.e.	-	(0.02)	(0.28)	(0.80)
Marginal cost pricing (p = c, K = C)	0.88	4.20	-	54.62
s.e.	(<0.01)	(0.05)	-	(1.94)
Linear model				
Actual prices (p = 0.12, K = 0.09)	0.10	9.38	22.59	20.76
s.e.	(0.01)	(0.16)	(1.48)	(36.96)
Optimal prices (p = 0.09, K = 0.02)	0.18	7.76	24.27	31.25
s.e.	(<0.01)	(0.13)	(1.28)	(36.93)
No connection fee (p = 0.096, K = 0)	0.25	5.61	23.95	30.90
s.e.	(<0.01)	(0.09)	(1.28)	(36.93)
Marginal cost pricing (p = c, K = C)	0.49	7.57	<0.01	82.19
s.e.	(0.01)	(0.12)	-	(37.17)

Note: c = 0.0201, C = 0.0216. Reported numbers are point estimate and s.e.

Standard errors are produced using a bootstrap with 500 repetitions.

Figure 1. Consumption needs (μ) and the conditional connection probability (π) as

a function of age (male, living in Helsinki)
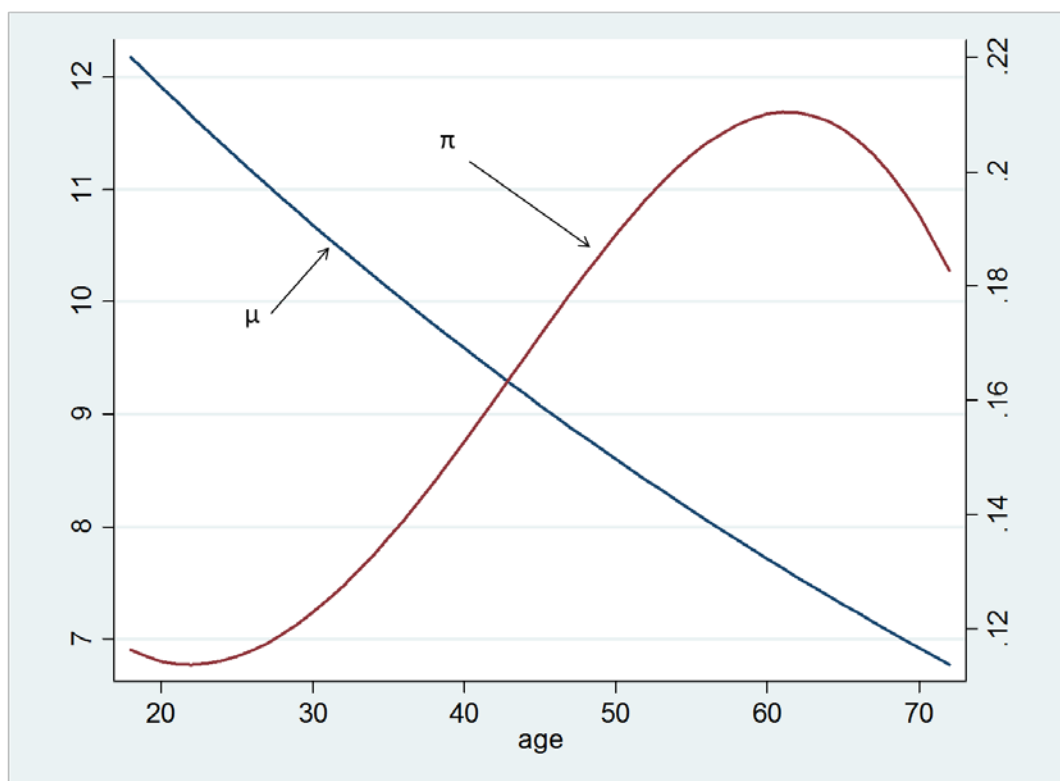
APPENDIX I

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 our results: *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 too high elasticities. We should also observe declining usage during the experiment period if the consumers do not find the quality of the good to exceed their expectations.[39] *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.[40] *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.[41]

To test these implications we have run a number of reduced-form regressions explaining the number of connections and their length by consumer fixed effects and period indicators. While these do not provide a perfect test for the alternative explanations, they do provide some support for our assumptions as we find no evidence for learning, consumption shifting in time or an advertising effect (see part E of the online Appendix for details).

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. Using data on the 306 consumers with plan B either alone or together with data on plan A consumers does not change our results (see part F of the online Appendix).

---

[39] A testable implication of this type of learning is that if the consumers learned that they liked (disliked) the services, the number of connections after the experiment should have been higher (lower) than before the experiment.

[40] 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.

[41] 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.

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 both with this and the other combinations of explanatory variables we tried.

A deeper criticism challenges our entire demand model. As an alternative, 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 that include consumer fixed effects show that the own-price elasticity is -10.84 and the annual per-consumer surplus 0.56 euros (see part G of the online Appendix for details). Estimating these elasticities and consumer surpluses in several alternative ways reinforces this result.

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. While the number of industry experts from whom we personally elicited the marginal cost beliefs is larger than in prior work (including Bajari and Ye 2003) and the method of elicitation follows the practices recommended elsewhere (see, e.g., Dominitz and Maski 1996, 1997, Manski 2004), we have considered the plausibility of these estimates 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 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 undisclosed contracts) 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 provid-

ers 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.[42]

As a second check of the quality of the experts' 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 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:[43] These suggested prices of experts are in fact close to, though *lower* than, 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.

---

[42] 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.

[43] There is only one price suggestion that is strictly *higher* than the actual prices: One respondent suggested that the optimal p is 0.10 euros.

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.[44] 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 are 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 *static monopoly* prices from our preferred model are *lower* than the actual prices that were

---

[44] 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.

charged. This revealed pricing behavior strongly suggests that the operator whose data we have did *not* behave as if it faced significant price competition over the users of the early mobile services.[45]

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

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

where c = marginal cost per minute of connection, C = marginal cost of opening/terminating a connection, $q_{\alpha > \bar{\alpha}}$ = conditional connection length, and τ = the valued added tax (VAT) parameter (as VAT is 22%, $\tau = 1.22$), implying that per-minute price (connection fee) charged from the customer is τp (τK). 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

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

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

The general form of the FOCs is

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

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

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

$$(A.4) \qquad B(p, K) = \left[\sqrt{2\gamma \tau K} + \tfrac{1}{\lambda} \right] (p - c) + K - C$$

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

---

[45] 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.

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

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

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

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

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

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

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

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

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

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

B. Elicitation of expert opinions

Elicitation procedure: 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 had not systematically gathered and/or stored data on the two marginal costs.

To obtain information on marginal costs, we designed a structured survey that consists in total of 13 questions which we personally administrated to a number of acknowledged industry experts (more on this below). The structure of the elicitation procedure is as follows: First, a definition and an example of marginal costs is given and discussed, if needed, at length with the industry expert. Second, the respondents are asked for point estimates of the two marginal costs (Q1-Q2). The next two questions are 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. This question (Q5) is a check-question, as it allows us to separate the answers of those respondents who know that WAP connections were technically identical to GSM calls from the answers of those who do not know this. The sixth question (Q6) is about the elasticity of demand and the seventh (Q7) about pricing. Finally, a series of questions (Q8-Q13) elicits information about the distribution of marginal costs, following the work of Dominitz and Manski (1996, 1997). This elicitation procedure was pre-tested with an industry expert and modified slightly on the basis of his reactions to it.

Selection of the industry experts: 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 the elicitation procedure described) 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 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.

C. Estimation results of the baseline model

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 $\exp(g(z_i, \mu))$ and $k(z_i, \lambda)$.

The first step estimation results are displayed in Table A.2, which among other 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 Table A.2 and boot-strapped standard errors in Table A.3. 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. Comparison to i-mode

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 on whether quality-adjusted WAP prices were high.

First, i-mode's overall usability appears to have been somewhat better than that of WAP. 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.[46] All this suggests that i-mode had fewer substitutes and better quality than WAP.

For a 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.

---

[46] 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).

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 have 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.

[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 2nd or 3rd 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 1st 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 average -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 or compensating variation (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

$CV \approx -0.5pY/\varepsilon$, where Y is the quantity consumed, p is the price, and ε 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 ε? 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 sur-

plus 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.

48

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	<u>2585.69</u>	2588.231	2590.941	2591.215	2601.392
of k(.)	110.1632	110.1178	110.0678	110.0267	<u>109.8905</u>	110.023

Table A.2. Estimation results from the first step (μ and λ)

	WAP_COUNT (μ)		CALL_DUR (λ)	
	Coefficient	Std. error*	Coefficient	Std. error*
AGE	0.01	0.01	-0.07	0.04
GENDER	0.11	0.29	-0.05	0.32
CITY	0.40	0.27	-0.61	0.31
GENDER*AGE	-3.8E-03	6.5E-03	-7.3E-03	1.6E-02
CITY*AGE	-0.02	0.01	0.03	0.02
GENDER*CITY	0.05	0.16	-0.03	0.08
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	2.13	0.28	2.81	0.52
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 (γ)

	Log-linear demand		Linear demand	
	Coefficient	Std. error	Coefficient	Std. error
AGE	-0.10	0.04	-1.64	0.59
GENDER	-3.95	1.74	-62.92	28.36
CITY	-0.95	2.12	-9.00	34.03
GENDER*AGE	0.09	0.04	1.46	0.66
CITY*AGE	-0.03	0.05	-0.40	0.67
GENDER*CITY	-0.30	1.26	-12.12	18.30
CONSTANT	22.28	1.72	155.77	25.51
Obs.	12740		12740	

Note: Standard errors are produced using a bootstrap with 500 repetitions.

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

	Count		Length	
Period-dummy	Coefficient	p-value	Coefficient	p-value
t = 2	0.28	< 0.01	0.14	0.59
t = 3	2.65	< 0.01	2.78	< 0.01
t = 4	2.34	< 0.01	2.65	< 0.01
t = 5	2.25	< 0.01	2.44	< 0.01
t = 6	0.17	0.00	-0.41	0.12
t = 7	-0.09	0.01	0.14	0.61
Consumer fixed-effects	YES		YES	
Tests of equality of period coefficients (p-values)				
2 vs. 6	< 0.01		0.03	
2 vs. 7	< 0.01		0.99	
6 vs. 7	< 0.01		0.03	

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