When Talk is “Free”: The Effect of Tariff Structure on Usage under Two- and Three-Part Tariffs

Eva Ascarza
Anja Lambrecht
Naufel Vilcassim◊

July 2012

(Forthcoming at *Journal of Marketing Research*)

◊ Eva Ascarza is Assistant Professor of Marketing at Columbia University (email: ascarza@columbia.edu). Anja Lambrecht is Assistant Professor in Marketing at London Business School (email: alambrecht@london.edu). Naufel Vilcassim is Professor of Marketing at London Business School (email: nvilcassim@london.edu). The authors thank the London Business School Centre for Marketing for financial support. They would like to thank Rajesh Chandy, Bruce Hardie, Kamel Jedidi, Carl Mela, Oded Netzer, Catarina Sismeiro, Gonca Soysal, participants at the UTD FORMS, the 2009 Marketing Science Conference, the 8th Triennial Invitational Choice Symposium, the 2011 Four School Colloquium, The Psychology and Economics of Scarce Attention conference, and at seminars at INSEAD, Erasmus University, HKUST, Northwestern University, and University of Rochester for their helpful comments.

When Talk is "Free": The Effect of Tariff Structure on Usage under Two- and Three-Part Tariffs

Abstract

In many service industries, firms introduce three-part tariffs to replace or complement existing two-part tariffs. As opposed to two-part tariffs, three-part tariffs offer allowances, or "free" units of the service. Behavioral research suggests that the attributes of a pricing plan may affect behavior beyond their direct cost implications. There is evidence that customers value "free" units above and beyond what would be expected based on the change to their budget constraint. Nonlinear pricing research, however, has not considered such an effect.

The authors consider a market where three-part tariffs were introduced for the first time. They analyze tariff choice and usage behavior for customers who switch from two-part to three-part tariffs. This study finds that switchers significantly "over-use" compared to their prior two-part tariff usage. They attain a level of consumption that cannot be explained by a shift in the budget constraint. The authors estimate a discrete-continuous model of tariff choice and usage that accounts for the valuation of "free" units. Results show that 83.9% of three-part tariff users value minutes on a three-part tariff more than they would on a two-part tariff. The authors derive recommendations for how the provider can exploit these insights to further increase revenues.

Keywords: Pricing, Nonlinear Pricing, Discrete / Continuous Choice Model, Three-Part Tariffs, Uncertainty, Learning, Free products.

Firms in various sectors are complementing or replacing their two-part tariffs with three-part tariffs. For example, telecom providers offer plans with “free” minutes instead of charging for every minute and banks offer a set number of “free” check-writing privileges instead of billing per check. Two-part tariff customers pay a regular, often monthly, access price and a usage price for every unit of consumption, whereas under a three-part tariff a usage price applies only to consumption in excess of the usage allowance. Within the allowance there is no usage charge; usage is “free.”

Recent evidence illustrates that consumers generally respond to free products or services differently from how they respond to the same good if the firm charged for it. Specifically, when evaluating free products or services, consumers do not simply subtract costs from benefits but instead perceive the benefits associated with free products as being higher than they would otherwise. This leads to increased demand (Shampanier, Mazar, and Ariely 2007), which has important implications for firms. For example, when AOL replaced their pay-per-use plans for dial-up internet access with flat-rate plans, demand at a zero usage price was far greater than AOL had forecasted based on the income effect of the price change alone (Cnet.com 1996; Messay 1998).[1] Behavioral research suggests that this increased demand can be attributed to a positive affective response to the offer of a zero usage price—a “free” component of the tariff. This affective response increases usage above and beyond what would be predicted based on the change of the budget constraint alone (Shampanier, Mazar, and Ariely 2007).

Importantly, research demonstrates that this positive affective response may likewise increase the valuation of other, even unrelated products. For example, Isen et al. (1978) show that individuals who obtain a small free gift (notepad or nail clipper) subsequently evaluate the performance of unrelated products (their own car and TV) as significantly higher than individuals

[1] As a result, AOL had to manage significant congestion and dissatisfied customers.

who do not obtain a free gift. Similarly, field data show that coupons can increase purchases beyond the expected income effect for other products than the coupon was issued for (Heilman, Nakamoto, and Rao 2002).[2] These insights go against standard economic theory that assumes that a change to a price would affect demand only through a change in the budget constraint.

In this research we examine how consumer demand changes when consumers switch from a two-part to a three-part tariff. We build on Shampanier, Mazar, and Ariely (2007) and argue that the free component of a three-part tariff leads to a positive affective response. This positive affective response increases the valuation of the service as such. Building on research that illustrates persistence of demand effects from positive affect across products, we argue that the positive affective response to a tariff with free units persists even after consumers have exceeded their usage allowance. As a result, we propose that when consumers switch from two-part to three-part tariffs, demand should increase beyond what would be predicted based on the change in the budget constraint alone. This demand effect should hold even for usage beyond the allowance.

We use data on tariff choice and usage of customers of a mobile phone company. An important feature of our data is the introduction of three-part tariffs in addition to the existing two-part tariffs during our observation period. This allows us to observe the same set of customers under two different pricing regimes: first, when only two-part tariffs were available, and then when customers were free to switch to three-part tariffs. In this market, the three-part tariffs were largely designed to increase customer acquisition. Thus, their addition to the choice set for existing customers is close to a natural experiment.

[2] The authors attribute this to elevated mood and additionally propose that a psychological income effect from receiving an unexpected coupon could lead consumers to spending more than the windfall gain from the coupon. They do not test which of the two explanations is more relevant. Golden and Zimmer (1986), Sherman and Smith (1987) and Donovan et al. (1994) provide further evidence that positive affect increases product demand, even for products that did not directly stimulate the positive affect.

An initial exploration of the data shows that customers who switched to a three-part tariff significantly "over-use" after switching: their level of consumption cannot easily be explained by the change to the budget constraint only, or by other plausible alternative explanations. To disentangle the effect of free minutes on customers' valuation of the service from the change in the budget constraint that arises from the new pricing structure, and from preferences at tariff choice, we jointly estimate each customer's tariff choice and usage decision conditional on the chosen tariff. In the utility function, we explicitly allow for greater demand on three-part tariffs. Since three-part tariffs are new to this market, customers are likely to be initially unaware that free minutes might affect their demand beyond the budget constraint. Hence, we allow for the possibility that customers learn about their three-part tariff usage which means that they learn about their valuation of "free" minutes.

Our results indicate that on three-part tariffs, 83.9% of customers use more than expected based on their previous usage. We interpret this as these customers having a greater valuation of the service than under the two-part tariff. This effect increases the provider's revenue from three-part tariff customers by 19.7%. We find that by reducing the fee charged for switching between tariffs, the provider could increase total revenues by 3.9%, and even further if it discontinued the option to switch to another two-part tariff. In both instances, customers' greater usage on three-part tariffs is key to any revenue increase.

Our findings are managerially relevant from three perspectives. First, projecting customer usage based solely on observed usage under existing two-part tariffs could lead firms to incorrectly determine the optimal tariff structure and prices, with potentially serious consequences for the firm's profits. Our analysis shows that firms may significantly underestimate the revenue effect from introducing three-part tariffs if they do not sufficiently account for the effect on preferences arising from the change in tariff structure. Second, our findings show that ignoring the effect of a greater valuation for "free" minutes underestimates three-part tariff usage by 14.9%. This result suggests that, when changing their tariff structure, firms may also need to adjust their service capacity. Third, and more broadly, our results illustrate that the attributes of a pricing plan do not only change its monetary value but also affect the perceived characteristics of the service.

Our findings add to the nonlinear pricing literature that has recognized behavioral preferences at tariff choice: customers choose flat-rate or three-part tariffs with large allowances even when these entail a greater bill than tariffs with a lower allowance (Nunes 2000; DellaVigna and Malmendier 2006; Lambrecht and Skiera 2006; Lambrecht Seim, and Skiera 2007). Our work confirms that such deviations from standard economic theory are not limited to the choice of a product or service but also affect its usage. The present work likewise complements a recent literature on choice and consumption under three-part tariffs (Jensen 2006; Bagh and Bhargava 2007; Iyengar, Ansari, and Gupta 2007; Grubb 2009; Grubb and Osborne 2011), which has so far abstracted from potential effects of the tariff structure on usage. As an exception, Iyengar et al. (2011) explore how tariff structure affects usage on two-part versus pay-per-use tariffs. They find that customers' marginal utility of consumption is lower on a two-part tariff than on a pay-per-use tariff.

More broadly our work contributes to research that explores behavioral effects of pricing. This includes the insight that attributes of a price or a pricing plan can affect behavior beyond their direct cost implications (Bertini and Wathieu 2008), systematic effects of price endings on consumers' purchase decisions (Anderson and Simester 2003; Thomas and Morwitz 2005), or a valuation of discounts that goes beyond the change in prices (Darke and Dahl 2003).

We next present our data, then provide evidence that customers "over-use" on three-part tariffs and discuss possible explanations. We then develop a joint model of tariff choice and usage that allows for a greater valuation of the service on a three-part tariff, resulting in greater usage. We present the results of the estimation and of our counterfactual analyses, and conclude with a summary of findings and implications of our work.

DATA

Our data includes a random sample of 5,831 individual customers (i.e., non-corporate) of a South Asian mobile telephony provider. We observe customers for up to 12 months, commencing May 2006. The focal firm had 16% of the installed base. Mobile phone service penetration was 35% in May 2007. On average, customers had been with the firm for 23.5 months before the start of our observation period. The data contain information on the tariff chosen and monthly usage of outgoing calls. During the observation period, 3.7% of customers left the firm, resulting in a total of 69,878 monthly usage and tariff choice observations. During the first three months, customers were offered a choice of two-part tariffs (Tariff_2_1 to Tariff_2_4 in Table 1; we refer to its currency as MU for "monetary units"). For each two-part tariff, the provider charges four different per-minute prices, all greater than zero, depending on the time of day and the call destination. Our data include the total number of minutes used per month, but not by time of day or destination. In addition, the firm provided us with the number of minutes used across all customers in each tariff by time of day and call destination. As a result, we use the weighted average of usage prices per tariff as a measure of the usage price.

On average, a customer uses 297 minutes a month and has a bill of MU 17.15 (Table 1). Customers are able to check their usage and bill by text-message, by phone or over the Internet. They are free to leave the provider at any time—there are no contractual obligations—or to switch to another tariff of the same provider. Customers can switch tariffs by calling into the firm's customer service center, visiting one of the firm's retail outlets or through an authorized agent. The provider charged MU 10 for switching to another tariff. This fee is high compared to the two-part tariff access prices and represents a significant expenditure in this emerging market where customers are cash flow-constrained.

We examine whether customers choose the *ex post* cost-minimizing two-part tariff based on the average and standard deviation of their usage in the first three months of our data. We find that only 10.9% of customers would have paid less on a different two-part tariff (see web appendix).

Three months after the start of our observation period, the company added three three-part tariffs to the existing two-part tariffs (Tariff_3_1 to Tariff_3_3 in Table 1). Under a three-part tariff, the marginal price is zero for usage within the allowance. The provider charges a single price for usage above the allowance. The new tariffs were heavily advertised in print and on TV. The provider introduced three-part tariffs to differentiate its offerings from its competitors and to increase customer acquisition, not as a recognition of limitations of two-part tariffs in sorting customers (Wilson 1993; Jensen 2006). Hence, for existing customers, the introduction of three-part tariffs was close to a natural experiment. The firm was unaware that a change in the tariff structure could potentially change demand above and beyond what would be expected from the change in the budget constraint.

Our panel covers customers who were subscribers of the firm before three-part tariffs were introduced but not newly acquired customers. Thus, we do not address market expansion effects. The two competitors offered similar two-part tariffs as the focal firm but no three-part tariffs.

DESCRIPTIVE ANALYSES

Switching from two- to three-part tariffs

While the focus of this research lies on three-part tariff usage, we also provide an overview of tariff-switching behavior. In our data, 13.7% of customers switched between tariffs: 5.8%

switched to a two-part tariff and 7.9% switched to one of the three-part tariffs, resulting in a total of 2,357 three-part tariff observations. A key strength of our data is that we observe customers on a two-part tariff prior to the introduction of three-part tariffs and, subsequently, several months of usage behavior under three-part tariffs. This provides us with a high number of observations per tariff and allows us to identify the effect of tariff structure on three-part tariff usage within individual customers.

As Table 2 illustrates, customers mostly switched from any of the two-part tariffs to Tariff_3_1. We ask whether a customer's decision to switch to a three-part tariff could have been predicted from her previous consumption. We find that of all customers that would have benefited from switching, 9.0% did switch, whereas among those customers that would not have benefited, only 4.7% switched (see web appendix for details). This is consistent with generally low switching rates in telecom services and likely a result of the high switching fee that may have deterred customers from switching.[3]

To assess whether the high switching fee might have deterred customers from switching, we replicate the analysis above, now considering a switch to be beneficial only when savings in the first month would compensate for the switching fee. In this case, a greater proportion of customers who would benefit from switching to a three-part tariff did switch (13.2% versus 9.0%). These results provide some indication that while customers behave optimally when choosing tariffs, the switching fee could have prevented customers from switching to three-part tariffs. Our econometric model will account for this.

[3] The share of customers switching to a three-part tariff is significantly lower than the share of customers that Grubb (2009) observes initially choosing a three-part tariff. Multiple factors likely contribute to this difference. First, our customers are required to *switch* to a three-part tariff, so switching costs may deter customers from switching. Second, our customers had a long-time experience with the service so over-confidence, which Grubb identifies as an important factor in choice, may have been less important. Third, in our market three-part tariffs are new hence customers might feel reluctant to try new plans.

Finally we examine subsequent tariff choices among customers who switched to a three-part tariff. We find that 11.7% later switched to another tariff: 8.3% switched back to a two-part tariff and 3.4% switched to a different three-part tariff. In contrast, none of the customers who switched to a two-part tariff in the first place later switched again. The difference in subsequent switching behavior between customers who switch to two-part tariffs and those who switch to three-part tariffs suggests they had difficulty predicting their own usage under three-part tariffs but not under two-part tariffs. Furthermore, customers typically do not switch immediately after their first three-part tariff choice. Rather, they spend on average 2.7 months on a three-part tariff before switching again. This indicates that customers learn about their usage on three-part tariffs before adjusting their tariff choice. Our econometric model will incorporate this learning process.

Usage under three-part tariffs

We turn to customers' usage behavior after the introduction of the three-part tariffs. A unique aspect of our data is that we observe all customers on two-part tariffs before the three-part tariffs were introduced. Moreover, not all customers switched to three-part tariffs after the introduction. We can therefore analyze whether switching to a tariff with "free" minutes affects consumption.

We compare the average monthly usage before the three-part tariff introduction to usage in the last month of our data. Customers who switched to a three-part tariff increased their usage by 15.1%, while customers who remained on a two-part tariff increased their usage by .9% (Figure 1(a)). This indicates a change in usage of three-part tariff customers that goes beyond a general time trend. This pattern is persistent over time (see web appendix) and is consistent across all three-part tariffs: On Tariff_3_1, usage increased by an average of 15.5%, on Tariff_3_2 by 16.2% and on Tariff_3_3 by 19.2%. Furthermore, actual usage often significantly exceeds the allowance: 72% of three-part tariff observations exceed the usage allowance, on average by 88.4%.

At first glance, this increase in usage could be a result of the change in tariffs' marginal prices; a utility-maximizing customer may use more on a three-part tariff simply because of the change in the budget constraint. To explore whether the change to the budget constraint can explain the increased three-part tariff usage, we estimate a linear demand model of monthly usage. Using the observations before the three-part tariff introduction, we estimate an individual-level demand intercept and a price coefficient for monthly consumption. We then predict usage conditional on the chosen tariff in the last month in our data, and compare these predictions with the actual behavior. This approach accounts for changes to the budget constraint since the prediction is based on prices and allowances of the tariff that each customer faces in the last month (see web appendix for details of the analysis and results).

Figure 1(b) shows the actual and predicted average usage in the last period. For customers who stayed on a two-part tariff predicted usage is 98.9% of observed usage. However, for customers who switched to a three-part tariff, predicted usage is only 85.9% of actual usage. In other words, the demand model that accounts for the change in the budget constraint predicts two-part tariff usage very accurately, but underestimates three-part tariff usage by almost 15%. This result is consistent across all three-part tariffs, and is independent of the month in which customers switch to a three-part tariff. This indicates that the effect persists over time (see web appendix).

Moreover, this result still holds when we relax model assumptions that could lead us to systematically over-predict usage. First, if demand was not linear in price but convex, then imposing a linear demand specification could possibly underestimate usage in regions where price is very low (or zero), hence systematically underestimating three-part tariff usage. We relax the linearity assumption and find that non-linear utility specifications lead to qualitatively the same results (see web appendix). Second, it is possible that customers who switch to a three-part tariff had different usage price sensitivity than customers who do not switch to a three-part tariff. As a consequence, the assumption of homogenous usage price sensitivity could lead us to under- or overestimate their price response and to under-predict three-part tariff usage. Similar to existing research on multi-part tariffs (Iyengar, Ansari, and Gupta 2007; Lambrecht, Seim, and Skiera 2007; Narayanan, Chintagunta, and Miravete 2007), we cannot estimate an individual-level price coefficient due to lack of individual-level price variation in the data. We can check, however, whether the homogeneity assumption is driving our results. We estimate a demand specification in which customers who switched to a three-part tariff have a different set of parameters than customers who did not switch to a three-part tariff. Again, we obtain qualitatively similar results (see web appendix).

In sum, the large increase in usage after switching to a tariff with "free" minutes cannot be explained by the difference in the budget constraint. It is consistent, however, with previous research that shows that free goods lead to a positive affective response, hence increase the valuation of the product or service, and consequently, its demand (Shampanier, Mazar, and Ariely 2007). Furthermore, this positive response towards "free" minutes can increase customers' valuation of other goods (Isen et al. 1978; Heilman, Nakamoto, and Rao 2002), thus affecting the entire consumption experience. We therefore conclude that customers who switch to a three-part tariff may assign greater value to the minutes of that tariff, including those above the allowance. Our data support this: 72% of three-part tariff observations exceed the usage allowance, on average by 88.4%. In other words, we observe that most customers increase their usage after switching to a three-part tariff, even when that implies that monthly consumption exceeds the tariff allowance.

Alternative explanations

We acknowledge that other explanations could plausibly lead to a similar pattern of usage.

Risk aversion: Risk aversion might lead customers to choose tariffs with large allowances where greater usage may be optimal. This may then result in greater consumption. However, risk-averse customers would also be more likely to keep their usage at or slightly below the allowance, which is not consistent with our data: 72% of three-part tariff observations exceed the usage allowance, on average by 88.4%, and 92% of three-part tariff customers exceed their usage allowance at least once. Furthermore, we observe that for 61% of the three-part tariff customers, usage lies above their two-part tariff satiation level. Risk aversion would not be able to explain such a change to the satiation level.

Regret avoidance: On a three-part tariff, customers might want to use "their money's worth", i.e., entirely use their allowance even if this exceeds the optimal usage of a rational utility-maximizing consumer. However, this explanation is not consistent with our data since most three-part tariff customers exceed the allowance by a large amount (on average 88.4%).

Within-day price variation: On the firm's two-part tariffs, usage prices vary with the time of day (peak and off-peak rates) and call destination (within and out of network). Since we do not have access to usage data by call type or records of individual calls, our previous analysis relied on the average usage price provided to us by the provider. However, in theory, all observed two-part tariff calls could have been made during peak hours when a greater usage price was charged, in which case the previous analysis may have overestimated the usage price sensitivity and hence under-predicted three-part tariff usage.

As a robustness check, we re-estimate the model presented above, now assuming that all two-part tariff calls were made at the highest marginal price, i.e. during peak hours out of network. In doing so, we possibly underestimate customers' price sensitivity and possibly over-predict, but would not under-predict, three-part tariff usage. Again only 85.8% of usage of three-part tariff customers can be explained by the shift in the budget constraint.

Self-selection: Customers may have switched to a three-part tariff because they anticipated greater usage in future periods. Econometrically, self-selection should lead to a high correlation between factors that affect three-part tariff choice, above and beyond expected savings, and factors that affect three-part tariff usage, beyond what would be predicted based on a consumer's demand parameters from two-part tariff usage. Investigating whether such a correlation exists is not feasible in a purely descriptive way, or by looking at demand alone. This can only be tested in a joint model of usage and tariff choice. We will therefore check for such a correlation when estimating our full discrete-continuous choice model in our model section.

Alternatively, self-selection could arise if usage followed an autoregressive process. If this were the case, customers who switch to a three-part tariff because they had a positive usage shock in the last period should also be more likely to increase their usage in future periods. We rule out this possibility in two ways. First we check for serial correlation among monthly usage shocks. Second, we investigate whether customers who get (higher) positive usage shocks are indeed more likely to switch. Neither of the analyses support that an autoregressive process leads to self-selection in our data (see web appendix for details).

Tariff-specific services or marketing activities: Usage behavior could change if the three-part tariff offered other services unavailable on the two-part tariff. For example, if text messages were offered for free on the two-part but not on the three-part tariff, three-part tariff customers may substitute calls for text messages. In our setting, text messages are rarely used because the language of conversation is different from the language script on the handset and because neither calling plan is connected with other services that would lead to such an increase in usage. Similarly, marketing activities aimed at switching customers to three-part tariffs and simultaneously increasing usage on these tariffs could explain high three-part tariff usage, e.g., advertising or the introduction of handsets with new features that stimulate usage and were limited to three-part tariff users. We know from the provider that no such activities were undertaken in this market.

Awareness of usage: Customers might "over-use" because they are unaware of their usage level. There are several indications, however, that suggest that over-usage is not due to a lack of awareness. First, customers had used the service for, on average, 23.5 months before the start of our data and the firm provides many possibilities for monitoring usage. Hence, it seems unlikely that three-part tariff switchers are completely unaware of their usage level. Even if the availability of "free" minutes initially resulted in unintended over-usage, it seems reasonable to assume that customers would adjust their usage within the next months, a pattern not observed in the data (see web appendix). Second, our data show a mass point of usage observations at the allowance, providing further support that customers track their usage (see web appendix). Third, if over-usage was largely a result of a lack of awareness, customers should stop using at their satiation point. However, 61.0% of customers use more than the satiation level we predict based on their two-part tariff usage. This provides further support that a lack of awareness is likely not the main reason behind the usage increase, but that instead the satiation level changed. Still we acknowledge that since our data does not include information on when or how often customers check their usage levels we cannot fully rule out that a lack of awareness contributed to the increase in usage we observe.

Intra-month usage uncertainty: On a three-part tariff, a customer's decision to make a call depends on her expected valuation of future calls during that month. For instance, a customer would prefer to use the entire allowance today even if today's calls are of low value to her, as long as the expected value of tomorrow's calls are of even lower value. However, if tomorrow's calls are unexpectedly of greater value than the three-part tariff's usage price, then the customer will still make these calls, possibly resulting in "over-usage". As a consequence, intra-month usage uncertainty could lead to the “over-usage” we observe in the data. To investigate this possibility, we analyze whether customers with greater intra-month usage uncertainty also show greater “over-usage” once they switch to a three-part tariff.

Since our data is limited to aggregate monthly usage and does not contain information on individual calls, we cannot measure the level of uncertainty within an individual month. However, the level of usage variation across months provides a measure of the degree of uncertainty a consumer faces in her overall usage. We check whether the degree of usage variation before the three-part tariffs were introduced is correlated with “over-usage” after switching. We measure “over-usage” as the ratio of actual to predicted three-part tariff usage. We find no correlation between these two measures (correlation=0.080, p-value=0.098). Alternatively, we measure over-usage as the difference between actual and predicted three-part tariff usage and correlate this with usage variation. Again, we find no correlation with the degree of usage variation (correlation=0.038, p-value=0.432). We conclude that the degree of intra-month uncertainty is unlikely to explain the increase in usage we observe in the data.

To conclude, even if we cannot simultaneously rule out all alternative accounts, they can hardly explain the full extent of the increase in usage we observe. Instead, the increase in usage appears consistent with the interpretation that customers have a greater valuation when a tariff offers “free” minutes. There are, however, difficulties in precisely pinning down the effect of “free” minutes in the descriptive model presented so far. First, we cannot identify a systematic increase in usage separately from random usage shocks and time-varying demand shifters. As a result we are so far unable to precisely estimate by how much customers’ demand changes due to the change in tariff structure. This is important for firms who need to accurately forecast their revenues and capacity needs. Second, modeling usage independently from tariff choice could lead to selection bias (Dubin and McFadden 1984) and provide biased estimates of demand parameters. Since the same set of parameters affects choice and usage, jointly modeling both should yield unbiased and more reliable parameter estimates. Third, in order to conduct policy simulations, we need a consistent set of parameters that fully describes customers' usage and tariff choice decisions and so need to incorporate factors that affect tariff choice alone (e.g., switching costs, tariff preferences) which have been overlooked so far. Finally, only jointly modeling choice and usage will allow us to conclusively rule out self-selection as an alternative explanation. As a consequence, we next build a joint model of usage and tariff choice that enables us to estimate the effect of "free" minutes and allows us to make inferences about customers' behavior on three-part tariffs.

MODEL DEVELOPMENT AND ESTIMATION

Model set-up

At the beginning of each month, a customer chooses one of the available tariffs or leaves the provider based on her expected usage for that month. Conditional on her tariff choice, she next determines her monthly usage based on the utility she derives from the service. We capture this behavior with a discrete/continuous choice model (Hanemann 1984; Dubin and McFadden 1984).

Building on our descriptive analyses, we incorporate a factor in the utility function to capture the possibility of greater utility if the tariff provides "free" minutes. This valuation of the "free" minutes affects three-part tariff usage directly, and choice indirectly, through expected usage. Since three-part tariffs were completely new to this market, we assume that a customer is initially unaware of a possible effect of "free" minutes on her consumption, above and beyond what the change to her budget constraint would account for.[4] Only when she first experiences a three-part tariff does she become aware of her valuation of "free" minutes. Consistent with tariff

[4] The firm did not expect any change to the usage from the introduction of three-part tariffs beyond what the change to the cost structure would entail.

switching patterns in the data, we assume that a customer learns about her three-part tariff usage, and thus the value of consuming on a three-part tariff to her, over time.

Mobile phone penetration is increasing and the firm's customers have used the service for a long time, so we assume that a customer who leaves the provider switches to a competitor rather than disconnecting the service. The competitors' tariff offerings did not change during the observation period, so explicitly accounting for their tariffs in estimation would not differ greatly from normalizing the price of the outside option to one, which we do for simplicity.

We model a customer's tariff choice based on the expected utility in the next period only (Iyengar, Ansari, and Gupta 2007; Lambrecht, Seim, and Skiera 2007; Narayanan, Chintagunta, and Miravete 2007). Alternatively, we could assume that customers trade off current-period switching cost against all future utility gains (Goettler and Clay 2011). However, accounting for future periods would require assumptions about the discount rate and complicate the estimation. Note that if consumers were in fact forward-looking, a static model might potentially overestimate consumers' switching costs but will not bias our main parameter of interest that captures the additional value of a three-part tariff.

Utility function

The customer chooses among a set of *J* tariffs. Each tariff *j* is defined by a monthly access price, denoted by F_j, an allowance measured in minutes of usage, $\tilde{q}_j$, and a marginal price, p_j, charged for each minute that exceeds the tariff's monthly allowance. A higher access price is associated with a higher usage allowance, so that $F_j < F_{j'}$ if $\tilde{q}_j < \tilde{q}_{j'}$. A two-part tariff is similar to a three-part tariff, but its allowance, $\tilde{q}_j$, is by definition set to zero. For each two-part tariff, a higher access price is associated with a lower usage price so that $p_j > p_{j'}$ if $F_j < F_{j'}$.

We assume that customer *i* at time *t* chooses the optimal tariff *j* and consumption levels for minutes of calls, q^*_{ijt}, and the outside good, q^*_{i0t}, to maximize her utility subject to the budget constraint. We choose a quadratic utility function as it allows for satiation (Iyengar, Ansari, and Gupta 2007, Lambrecht, Seim, and Skiera 2007). This is important since it reflects the behavior in our data where some customers use less than the allowance, the maximum possible usage at a zero usage price. It also assumes that customers are risk-averse agents. Since the utility function is linear in q_{i0t}, it does not capture nonlinearities in the outside good. Utility on tariff *j* is represented by

(1) $$U_{ijt}(q_{ijt}, q_{i0t}) = c_i \left[\frac{1}{b_i} \left(d_{ijt} q_{ijt} - \frac{\left(q_{ijt}^2 + d_{ijt}^2 \right)}{2} \right) + q_{i0t} \right] + \varsigma_{ijt}, \quad b_i, c_i, d_{ijt} > 0,$$

where c_i represents the marginal utility of income and d_{ijt} the satiation level, i.e., demand at a zero usage price. The demand slope, b_i, measures sensitivity to the usage price. The term ς_{ijt} captures observable and unobservable characteristics that affect tariff choice but not consumption. Customer *i*'s budget constraint is given by

(2) $$y_{it} = q_{i0t} + F_j + (q_{ijt} - \tilde{q}_j) I_{q_{ijt} \geq \tilde{q}_j} p_j,$$

where the price of the outside good has been normalized to one. Under a three-part tariff, the usage price, p_j, is strictly positive only for $q_{ijt} > \tilde{q}_j$, instances that we capture with the indicator variable $I_{q_{ijt} \geq \tilde{q}_j}$ set to one if $q_{ijt} \geq \tilde{q}_j$ and zero otherwise. For two-part tariffs, $\tilde{q}_j$ is by definition set to zero, hence $I_{q_{ijt} \geq \tilde{q}_j}$ is always one.

From Equations (1) and (2), we derive the customer's optimal two-part tariff usage as

(3) $$q^*_{ijt} = \begin{cases} 0 & \text{if } d_{ijt} \leq b_i p_j \\ d_{ijt} - b_i p_j & \text{if } d_{ijt} > b_i p_j \end{cases}$$

and under a three-part tariff as

$$(4) \qquad q_{ijt}^{*}=\begin{cases} d_{ijt} & \text{if } d_{ijt}<\tilde{q}_j \\ d_{ijt}-b_i p_j & \text{if } d_{ijt}-b_i p_j>\tilde{q}_j \\ \tilde{q}_j & \text{if } d_{ijt}-b_i p_j\le \tilde{q}_j\le d_{ijt}. \end{cases}$$

The first part of Equation (4) reflects consumption when usage is below the allowance. The second part determines usage when consumption exceeds the allowance and a strictly positive usage price, p_j, applies. The last part accounts for situations where the optimal usage would exceed the allowance of $\tilde{q}_j$ at a marginal price of zero, but falls short of the allowance at the positive marginal price. Since the incremental value of usage beyond the allowance is not justified by the additional usage charges that accrue abruptly, q_{ijt}^{*} must be equal to $\tilde{q}_j$.

Substituting the optimal demand for the outside good and usage into the utility function yields the conditional indirect utility function under a two-part tariff

$$(5) \qquad V_{ijt}(y_{it},p_j,F_j)=\begin{cases} c_i\left[y_{it}-F_j\right]+\varsigma_{ijt} & \text{if } q_{ijt}^{*}=0 \\ c_i\left[y_{it}-F_j-\left(d_{ijt}-\frac{b_i p_j}{2}\right)p_j\right]+\varsigma_{ijt} & \text{if } q_{ijt}^{*}>0 \end{cases}$$

and under a three-part tariff

$$(6) \qquad V_{ijt}(y_{it},p_j,F_j)=\begin{cases} c_i\left[y_{it}-F_j\right]+\varsigma_{ijt} & \text{if } q_{ijt}^{*}\le\tilde{q}_j \\ c_i\left[y_{it}-F_j+p_j\tilde{q}_j-\left(d_{ijt}-\frac{b_i p_j}{2}\right)p_j\right]+\varsigma_{ijt} & \text{if } q_{ijt}^{*}>\tilde{q}_j. \end{cases}$$

We decompose ς_{ijt} into three observed tariff preference shifters: (i) the cost of switching to another tariff of the same provider, (ii) the cost of switching to the outside option (i.e., churn) and (iii) the preference for choosing a three-part as opposed to a two-part tariff,

$$(7) \qquad \varsigma_{ijt}=\rho_1 SC^T\times I_j^T+\rho_2 I_j^P+\lambda_i I_j^{3pt}+\varepsilon_{ijt}.$$

The term SC^T reflects the provider's fee for switching to one of its own tariffs, ρ_1 reflects the sensitivity to this switching cost, and I_j^T is an indicator for switching to another tariff of the same provider. I_j^P is an indicator for switching to another provider, and ρ_2 captures non-monetary costs of switching to a provider other than the focal firm. The indicator I_j^{3pt} is one only under a three-part tariff, so λ_i captures unobserved factors that affect three-part tariff choice independent from usage expectations. It determines the individual-specific propensity to choose a three-part tariff. For customers who do not switch to three-part tariffs, it may, for example, capture a preference for a familiar tariff structure, or aversion to a high access price. For switchers to a three-part tariff it may capture a willingness to experiment with new tariffs that dominates the cost of the switching fee. Thus, it is a factor that is known to customers, even though it is unobserved to the researcher. Importantly, this is a preference that is reflected in choice, and explains switching to three-part tariffs, but not the changes in usage. We assume λ_i to be normally distributed across the population with mean and variance $\left(\mu_\lambda, \sigma_\lambda^2\right)$. The term ε_{ijt} is an unobserved preference shifter that the customer knows at the time of tariff choice. It is assumed to follow a Type 1 extreme value distribution.

Last, we specify the factors that determine the satiation level d_{ijt}. Our challenge is to model increased usage on a three-part versus a two-part tariff beyond what is due to the change in the budget constraint. Empirically, only a change to the demand intercept, d_{ijt}, could explain the observed increase in usage at any level of consumption. (A change to the demand slope could not explain a change in usage while usage is below the allowance.) We specify δ_i as the additional value from the service when consuming on a three-part tariff. It is independent from the change of the budget constraint and measures a change in behavior conditional on tariff choice. Note that while the parameter λ_i captures unobserved factors that lead to the *choice* of a three-part tariff, δ_i captures how the valuation of "free" minutes changes customers' *usage* behavior.

Since three-part tariffs are new to this market, customers are not yet aware of their positive affective response to the "free" allowance and its subsequent effect on usage at their initial three-part tariff choice. This means that they do not yet know that access to "free" minutes may change their usage behavior. Therefore, δ_i does not affect the initial three-part tariff choice and enters the demand intercept, d_{ijt}, only after a customer initially chooses a three-part tariff. Note that we would be unable to identify both δ_i and λ_i if the consumer was aware of δ_i at her initial three-part tariff choice.

The parameter δ_i is assumed to be normally distributed across the population with mean and variance $\left(\mu_\delta, \sigma_\delta^2\right)$ and takes the same value for any three-part tariff. An individual-specific parameter η_i captures differences among customers' demand that are constant over time and known to the customer but unknown to the econometrician. It is assumed to be normally distributed with mean and variance $\left(\mu_\eta, \sigma_\eta^2\right)$.

It is possible that similar factors drive a customer's choice of a three-part tariff and her usage on a three-part tariff. This would induce a correlation between the three-part tariff choice preference, λ_i, and the valuation of the three-part tariff's allowance that affects usage, δ_i. Such a correlation could come from self-selection of customers who switch to a three-part tariff because they plan to use more, or from tariff-specific marketing activities (e.g., handset subsidies), or differences in tariff-specific services (e.g., included text messages). As we discuss above, our conversations with the provider confirm that there are no such policies, and our analyses of the data provide no support for self-selection. Nevertheless, we revisit this possibility in the results section and look at the correlation between posterior estimates for parameters λ_i and δ_i.

We allow for uncertainty over usage at the time of tariff choice (Narayanan, Chintagunta, and Miravete 2007; Lambrecht, Seim, and Skiera 2007). We model an unobserved usage shock ϕ_{it} that reflects random usage variation. We observe in our data that individual-level usage variation is correlated with individual average consumption (correlation=0.77 p-value<0.0001), i.e., heavy users have a higher variance of usage than light users. Thus, we assume a multiplicative usage shock, ϕ_{it}, which is gamma distributed with equal shape and scale parameter r, hence with mean 1, and variance $1/r$. At the moment of tariff choice, the customer knows this shock only in distribution. Following the tariff choice but prior to making her usage decision, she observes her usage shock and consumes accordingly. Unobserved tariff-specific preferences, ε_{ijt}, drive tariff choice, but do not affect the distribution of demand. The two sets of unobservables, ε_{ijt} and ϕ_{it}, are assumed to be independent. Correlation could arise from user- and plan-specific advertising or customer-specific promotions but we know from the provider that such campaigns were not present.

To ensure a positive demand intercept, we specify d_{ijt} in exponential form

(8) $$d_{ijt} = \phi_{it} e^{h_t a_1 + \eta_i + \delta_i I_j^{3pt}},$$

where h_t is a dummy for holiday periods and a_1 is a parameter to be estimated. Note that even though the variance of the usage shock is homogeneous across customers, the effect of the shock on usage is heterogeneous because of its multiplicative interaction with mean usage, $e^{h_t a_1 + \eta_i + \delta_i I_j^{3pt}}$.

Demand uncertainty and tariff choice

A customer chooses the tariff that yields the highest expected indirect utility. Customers are experienced users of a two-part tariff: They know their two-part tariff usage preferences and the distribution of the usage shock, ϕ_{it},[5] but are uncertain about its exact realization. In other words, at tariff choice, customers do not know their exact usage on each tariff, q_{ijt}^{*}, since the usage shock ϕ_{it} has not been yet realized. However, they know their *expected* usage on each tariff, since they know the distribution of the usage shock. We obtain the expected indirect utility of a two-part tariff by taking expectations over the unknown usage quantity in Equation (5), that is over the distribution of the uncertain shock ϕ_{it} (see also Iyengar, Ansari, and Gupta 2007; Lambrecht, Seim, and Skiera 2007; Goettler and Clay 2011)

$$\begin{aligned} \mathrm{E}[V_{ijt}] &= \mathrm{P}(q_{ijt}^{*}=0)\ \mathrm{E}[V_{ijt} \mid q_{ijt}^{*}=0] + \mathrm{P}(q_{ijt}^{*}>0)\mathrm{E}[V_{ijt} \mid q_{ijt}^{*}>0] \\ &= \mathrm{P}(q_{ijt}^{*}=0_{j})c_i(y_{it}-F_j) + \mathrm{P}(q_{ijt}^{*}>0) \\ &\quad \times c_i\left[y_{it} - F_j + \frac{1}{2} b_i p_j^{\ 2} - \mathrm{E}\left(\phi_{it} e^{h_t a_1 + \eta_i} \mid q_{ijt}^{*} > 0 \right) p_j \right] + \varsigma_{ijt}. \end{aligned} \tag{9}$$

Similarly we obtain the expected indirect utility of a three-part tariff by taking expectations of the three-part tariff optimal usage over the unknown quantity. While our customers are experienced users of two-part tariffs, they do not have past experience on three-part tariffs. As such, when a customer experiences a three-part tariff, she observes a realization of $\phi_{it}e^{\delta_i}$ but cannot separate the effect of the usage shock, ϕ_{it}, from that of the new tariff structure. In other words, her three-part tariff choice is guided by uncertainty about the usage shock, ϕ_{it}, and the value of δ_i. Hence we obtain the expected indirect utility of a three-part tariff by taking expectations of Equation (6) with respect to the distribution of $\phi_{it}e^{\delta_i}$

---

[5] Iyengar, Ansari, and Gupta 2007 find that consumers learn about their usage uncertainty within nine months.

$$\mathrm{E}[V_{ijt}] = \mathrm{P}(q^*_{ijt} \le \tilde{q}_j)\ \mathrm{E}[V_{ijt} \mid q^*_{ijt} \le \tilde{q}_j] + \mathrm{P}(q^*_{ijt} > \tilde{q}_j)\mathrm{E}[V_{ijt} \mid q^*_{ijt} > \tilde{q}_j]$$

$$(10) \qquad = \mathrm{P}(q^*_{ijt} \le \tilde{q}_j) c_i (y_{it} - F_j) + \mathrm{P}(q^*_{ijt} > \tilde{q}_j)$$

$$\times c_i \left[y_{it} - F_j + p_j \tilde{q}_j + \frac{1}{2} b_i p_j{}^2 - \mathrm{E}\left(\phi_{it} e^{h_t a_1 + \eta_i + \delta_i} \mid q^*_{ijt} > \tilde{q}_j\right) p_j \right] + \varsigma_{ijt}.$$

Note that the expected indirect utility of an *initial* three-part tariff choice is slightly different since, as discussed earlier in this section, δ_i does not affect a customer's initial three-part tariff choice. As a consequence, when computing the expected indirect utility of a three-part tariff for customers who have not yet experienced it, delta drops from Equation (10) and hence we take expectations over the shock ϕ_{it}. In the appendix we derive the close form expressions for the expected indirect utility in all three cases: a two-part tariff, an initial three-part tariff, and subsequent three-part tariff choices.

Finally, if a customer decides to leave the provider, her expected indirect utility is

$$(11) \qquad \mathrm{E}[V_{i0t}] = c_i\, y_{it} + \varsigma_{i0t}.$$

Summarizing, a customer evaluates the expected indirect utility of each available option and chooses the option with the highest expected indirect utility. When evaluating a two-part tariff, a customer knows her preferences and the distribution of the shocks that affect future demand, but not the exact consumption next period. The same considerations affect her initial three-part tariff choice. Once a customer has experienced a three-part tariff, she evaluates subsequent three-part tariff choices taking into account her beliefs about her three-part tariff usage, that is her beliefs about her own valuation of "free" minutes. We next explain how customers learn about that value over time.

Learning

At the start of our data period, customers have been with the provider for on average 23.5 months, exclusively on two-part tariffs. Consistent with prior research that has found that customers learn about their usage within approximately 9 months (Iyengar, Ansari, and Gupta 2007), we assume that by the time we observe them customers have already learned about their two-part tariff usage. However, once they switch to a three-part tariff, customers' usage behavior changes and customers cannot easily infer their three-part tariff usage from prior two-part tariff consumption.

Two mechanisms could describe the process by which customers become aware of their three-part tariff usage. Either, customers become instantly knowledgeable as they make their first usage decision on a three-part tariff, or they learn gradually as they observe their usage behavior. Our data shows that customers who switch from three-part tariffs do so after an average of 2.7 usage periods, which suggests that they learn about three-part tariff usage over time.

We therefore extend our model to accommodate customers' learning. Specifically, we allow customers to learn about how their usage on a three-part tariff differs from their usage on a two-part tariff. This corresponds to learning about their preference for "free" minutes, and thus about the parameter δ_i which captures the differential usage.

We assume a Bayesian learning process (Erdem and Keane 1996). A customer learns about her true value of δ_i using her own usage as the signal of her preferences. Econometrically, a customer knows that the unknown usage quantity on a three-part tariff, $\phi_{it}e^{\delta_i}$, is gamma-distributed with parameters $\left(r, \beta_i\right)$ where $\beta_i = \frac{r}{e^{\delta_i}}$.[6] The customer knows r since $\frac{1}{r}$ is the variance of the usage shock. Thus, learning about δ_i translates into finding the true value of the scale parameter β_i. After first switching to a three-part tariff, she forms a belief about the true value of β_i. We denote the time-varying individual beliefs by $\tilde{\beta}_{it}$. At the end of a period she

6 This result follows from the properties of the gamma distribution: Let be $X \sim$ gamma$\left(\alpha_1, \alpha_2\right)$ with $\alpha_1, \alpha_2 > 0$. For any $k > 0$, $kX \sim$ gamma$\left(\alpha_1, \alpha_2 / k\right)$.

observes her three-part tariff usage and updates the belief. Figure 5 summarizes the decision process.

We assume that the initial beliefs are gamma-distributed. This is a less restrictive specification than the commonly used normal distribution, since the gamma distribution allows for non-symmetry in the distribution of the unknown usage quantity. In our multiplicative setting, normally distributed beliefs would not be easily tractable. Since the noise that affects the learning process (usage shock) is not normally, but gamma-distributed, normal beliefs would not be conjugate priors. Moreover, the gamma specification allows customers to learn at different rates in a relatively parsimonious model since a customer's variance of the belief, and thus her speed of learning, depends on her own signal (Equation (16). More generally, whereas most learning models reflect learning about an additive shock or shift, we suggest an approach that allows customers to learn about a shock with multiplicative nature.

Following the customer's switch to a three-part tariff, she forms an initial belief over the distribution of β_i such that

(12) $$\tilde{\beta}_{i0} \sim \text{gamma}\left(\alpha_0, \beta_0\right).$$

The customer knows the tariffs' characteristics ($p_j, \tilde{q}_j$), her preferences a_1 and η_i that affect her demand intercept, and her demand slope, b_i. At the end of the first period on a three-part tariff, τ_1, she observes her consumption $q^*_{ij\tau_1}$ and receives the signal $s_{i\tau_1}$

(13) $$s_{i\tau_1} = \begin{cases} \dfrac{q^*_{ij\tau_1}}{e^{h_t a_1 + \eta_i}} & \text{if } q^*_{ij\tau_1} \le \tilde{q}_j \\ \dfrac{q^*_{ij\tau_1} + b_i p_j}{e^{h_t a_1 + \eta_i}} & \text{if } q^*_{ij\tau_1} > \tilde{q}_j \end{cases},$$

which she knows is gamma-distributed, with known shape parameter r and scale parameter β_i.[7] She updates her prior belief about β_i, $\tilde{\beta}_{i1}$, which then enters the next period's tariff choice,

(14) $$\tilde{\beta}_{i1} \sim \text{gamma}\left(\alpha_0 + r, \beta_0 + s_{i\tau_1}\right).$$

More generally, a customer who has spent n periods on a three-part tariff has the following belief about the scale parameter

(15) $$\tilde{\beta}_{i\tau_n} \sim \text{gamma}\left(\alpha_0 + nr, \beta_0 + \sum_{t=1}^{n} s_{i\tau_t}\right),$$

with mean and variance

(16) $$E(\tilde{\beta}_{i\tau_n}) = \frac{\alpha_0 + nr}{\beta_0 + \sum_{t=1}^{n} s_{i\tau_t}} \qquad Var(\tilde{\beta}_{i\tau_n}) = \frac{\alpha_0 + nr}{\left(\beta_0 + \sum_{t=1}^{n} s_{i\tau_t}\right)^2}.$$

Since customers are unaware of the value of "free" minutes before experiencing the three-part tariffs, we set $\alpha_0 = r\beta_0$. This leads to an expected value of the initial belief of r and reduces Equation (15) to Equation (16). The variance of the initial belief is $\frac{r}{\beta_0}$. For any value of the (α_0, β_0), the expected value of the belief converges to the true value and the variance goes to zero as the customer gets more experience on a three-part tariff (see web Appendix for proof).

Identification and estimation

There are three groups of parameters: (i) preferences for usage and sensitivity to the usage price (η_i, b_i, δ_i), (ii) preferences for tariffs, marginal utility of income, and switching costs ($\lambda_i, c, \rho_1, \rho_2$), and (iii) parameters capturing the distribution of uncertainty including beliefs

---

7 If a customer's usage is zero or equal to the allowance, $\tilde{q}_j$, there is not a one-to-one relationship between the shock and realized usage so the customer cannot infer the signal s_{it}. We have no observations of zero usage under three-part tariffs. In 11 instances, we observe usage equal to the allowance, resulting in a many-to-one mapping between ϕ_{it} and q^*_{ijt}. We assume that customers do not update their beliefs in such cases.

(β_0) and usage shocks (r). For each customer we observe tariff choice and usage. Usage is governed by the preference for minutes, the price sensitivity and the distribution of the usage shock, while tariff choice is governed by *expected* usage (determined by known preferences and beliefs about the unknowns) and tariff choice specific parameters. For 13.7% of customers we observe usage on at least two different tariffs.

Given that many customers stay on the same tariff for the entire observation period (i.e., they face the same marginal price), we cannot identify unobserved individual-level heterogeneity in both η_i and b_i. As a consequence, we estimate η_i with unobserved individual-level heterogeneity which we identify from differences in usage across customers, while we specify the price coefficient b_i as a linear function of individual-level (gender and location) and district-level (labor and literacy levels) demographic variables. That is, $b_i = b + a_2 d_i$, where d_i is a vector containing demographic information and b and a_2 are parameters to be estimated.

We identify the price coefficient from differences in usage prices across customers, that is the different usage prices across tariffs (ranging from MU 0 to .079), and within customers, that is the change in usage price when a customer switches between tariffs (13.7% of customers are on at least two tariffs) and the different marginal prices on a three-part tariff (MU 0 for usage below and MU .050 for usage beyond the allowance). Observing usage under a variety of different usage prices allows us to precisely pin down the price coefficient. Previous research was restricted to a much lesser degree of price variation, even across customers (Iyengar, Ansari, and Gupta 2007; Lambrecht, Seim, and Skiera 2007; Narayanan, Chintagunta, and Miravete 2007).

We can disentangle δ_i and λ_i since we observe both tariff choice and usage behaviors: λ_i does not enter the usage decision, and δ_i does not affect the initial three-part tariff choice. We identify λ_i on an individual-level from observed *choices* between two-part and three-part tariffs because we observe that customers who are otherwise similar in their parameter values have a different propensity to switch to a three-part tariff. For switchers to a three-part tariff, we identify δ_i from the individual-level differences in *usage* levels on two-part versus three-part tariffs that are not explained by changes to the budget constraint. Note that we can separate δ_i from η_i for all three-part tariff switchers since we observe the same set of customers on both pricing regimes: first on a two-part tariff and then on a three-part tariff.

Churn is not prominent, so we are unable to identify the marginal utility of income, c, which we set to 1 (Narayanan, Chintagunta, and Miravete 2007). We identify the sensitivity to switching costs, ρ_1, from differences in the propensity to switch between customers with otherwise similar parameters. We can identify the sensitivity to switching cost, ρ_1, separately from the three-part tariff preference, λ_i, since some customers switch between two-part tariffs only.

The parameter β_0 reflects the variability of the initial belief about δ_i. It is identified from differences in choice after having switched to a three-part tariff (11.7% of the three-part tariff switchers later switch to a different tariff), that is from variability across (whether they switch again) and within three-part tariff customers (when they switch). We identify the parameter r, which drives usage uncertainty, from tariff choice and usage variation across people and periods.

The expression of the likelihood function is derived in the appendix. It entails the joint probability of tariff choice and usage decisions. The model is estimated in a Hierarchical Bayes framework. We use a data augmentation approach to model the unobserved individual-level parameters as well as the time-variant beliefs (see web appendix)

In addition to the full model just described (denoted by Model 3 hereafter), we also estimate two restricted versions: Model 1, which assumes that customers do not value three-part tariff minutes any differently from two-part tariff minutes and so the same parameter set governs behavior on two-part and three-part tariffs, and Model 2, which accounts for a greater valuation of minutes on a three-part tariff but assumes that, after switching to a three-part tariff, customers immediately acquire full knowledge about their value of δ_i.

RESULTS

Estimation results

Table 3 summarizes the posterior distributions of the parameter estimates of the three models laid out in the previous section (Model 1 where δ_i is set to zero, Model 2 which includes δ_i but no learning, and Model 3 which accounts for individual-level learning about δ_i). Using each set of estimates, we predict usage levels for all customers in our data (Table 4). Predicted two-part tariff usage of 282 minutes in Model 3 compares to observed usage of 294 minutes. The predicted three-part tariff usage of 434 minutes very accurately reflects the observed three-part tariff usage of 434 minutes. In contrast, Models 1 and 2 predict considerably lower three-part tariff usage (369 and 410 minutes respectively). Similarly, fit measures suggest that Model 3 best reflects usage behavior (Table 5). Compared to Model 1, Model 3 reduces the Mean Squared Error (MSE) for usage, conditional on observed tariff choice, by 3.0%. If we consider three-part tariff observations only, the MSE reduces by 72.1%. By contrast, Model 2 reduces the MSE by 1.3% for all observations and by 21.5% for three-part tariff observations relative to Model 1. The Mean Absolute Percentage Error (MAPE) confirms that Model 3 performs better than either Model 1 or Model 2. The percentage of correctly predicted choices (hit rate) indicates that all three models predict choice well. All fit measures confirm that Model 3 captures customers' tariff choice and usage behavior better than Models 1 and 2.

We next turn to the parameter estimates of Model 3 in more detail. The individual-level preference for "free" minutes, δ_i, has a mean of .218 and a standard deviation of .384. It is positive for 83.9% of three-part tariff customers, indicating that a large majority of customers value the "free" minutes beyond their direct cost implications. Similarly, Model 2 finds a positive valuation for 84.8% of customers. The variance of the usage shock, *r*, has a posterior mean of .205, which translates into a significant effect on usage volatility. For example, a value of the usage shock that is equal to its standard deviation shifts the average two-part tariff usage by 45.3%. The negative coefficient for the sensitivity to switching costs between tariffs, ρ_1, indicates that the switching fee notably reduces switching. Similarly, the negative coefficient for the sensitivity to cost of switching to other providers, ρ_2, shows that customers also have high non-pecuniary costs of leaving the provider.

The parameter β_o relates to the learning process: the variance of customers' initial belief, r / β_o, reflects the extent of over- or underestimation of the true value of δ_i when a customer first switches to a three-part tariff. For the sample of 247 customers who were observed on a three-part tariff for at least four periods, we compute the individual-level deviation between the true value of $\beta_i = r / e^{\delta_i}$ and the belief, $\tilde{\beta}_{it}$. In period 1, the deviation is 32.8% and it reduces to 25.5% after experiencing the three-part tariff for 3 months. This result indicates that customers learn about their three-part tariff usage over time. Not surprisingly, estimation of the learning process leads to a slightly lower variance of the usage shock, $1/r$, than in Model 2. Variation in usage that was previously attributed to the usage shock is now partly captured by learning about δ_i.

Finally, we look at the correlation between the posterior means of the individual-level parameters λ_i and δ_i (Figure 2). The additional valuation of consuming on a three-part tariff is uncorrelated with the customer's choice preference for three-part tariffs (correlation=0.004 p-value=0.930). As a robustness check, we compute the correlation between the posterior means of λ_i and δ_i considering only customers for whom we observe at least 5, 6 or 7 periods on a three-part tariff.[8] In all cases we find the correlation to be small and insignificant (Table 6).

In sum, our results provide strong evidence that accounting for the value of "free" minutes explains behavior significantly better than ignoring such an effect. We show that the large majority of customers value "free" units above their costs implications, and that customers learn about this valuation over time.

Customer sensitivity to prices and allowances

We evaluate the sensitivity of customers' behavior with respect to tariff attributes. Based on the estimates of Model 3, we compute the elasticity of choice and usage to changes of prices and allowances (Table 7). The two-part tariff's choice elasticity with respect to changes in the access price is −2.13. The three-part tariff's choice elasticity with respect to the access price is −13.28. That is comparable to previously estimated three-part tariff elasticities of up to −10.15 (Lambrecht, Seim, and Skiera 2007). These estimates reflect that the elasticity increases in the access price and imply that the access price elasticity is, on average, about four times larger for the three-part tariff relative to a two-part tariff. For the two-part tariff we find a particularly high choice elasticity with respect to the usage price of −11.39. This is likely a result of the structure of the two-part tariff menu where the individual usage prices are relatively comparable. Three-part tariff choice is less elastic to the usage price (−5.05). Its elasticity to changes in the usage allowance is 4.30, reflecting that the allowance plays an important role in consumer tariff choice.

We now turn to the effect on usage. Similarly to previous results (Park, Wetzel, and Mitchell 1983), two-part tariff usage is relatively inelastic to changes in the usage price (−.10).

[8] By doing so we confirm that the absence of correlation is not driven by the prior assumption of independence, since we exclude customers whose individual-level estimates would mostly rely on the prior distribution.

Usage is even less elastic on three-part tariffs (−.03) as not every usage observation lies above the usage allowance. The elasticity of usage with respect to the allowance is .84. Overall, the prices and allowances are more important in determining choice than usage conditional on choice.

Impact of the "free" minutes to the provider's forecasts and revenues

Our results so far show clear evidence for the additional value of "free" minutes. A firm benefits from knowing that customers value usage on a three-part tariff more than on a two-part tariff along several dimensions. First, if a firm was not aware of such greater usage, it would under-estimate usage under a three-part tariff, with likely significant impact on capacity planning. As Table 4 illustrates, a model ignoring the effect of a greater valuation for "free" minutes (Model 1), underestimates three-part tariff usage by 14.9%. If a provider is operating close to its capacity limits, such prediction error may easily cause a drop in call quality and customer satisfaction, which eventually could lead to higher churn. Finally, systematic errors in usage predictions will lead to misleading revenue forecasts. This would have a negative impact on management decisions such as budgeting, customer resource allocation (based, for example, on customer lifetime value calculations) or targeting.

To measure the effect of δ_i for our provider's revenue, we simulate three-part tariff revenue based on the individual-level estimates of the full model (Model 3) and compare it to the revenue the company would obtain if customers did not attach greater value to the free minutes—i.e., we set δ_i to zero. We find that the mean expected revenue per three-part tariff customer decreases from MU 21.3 to MU 17.1, which implies that the preference for "free" minutes, δ_i, accounts for 19.7% of the revenues obtained from three-part tariff customers and so represents a significant fraction of a firm's three-part tariff revenues.

RECOMMENDATIONS TO THE FIRM

Our results indicate that three-part tariffs can increase customers' usage and a firm's revenues. We next explore whether and how the firm can exploit this insight by encouraging switching to three-part tariffs. We investigate the impact of lowering the switching fee on firm's revenues, whether the firm can increase revenues by increasing customers' choice preferences for three-part tariffs and whether the firm would benefit from eliminating two-part tariffs. Since we do not have data about customer acquisition we restrict our analysis to the current set of customers.

Lowering the switching fee

Our estimates show that the switching fee strongly affects switching ($\rho_1 = -1.025$). To check whether the firm could further increase its revenues by reducing the switching fee, we simulate revenues in the period following our observation window under different levels of the switching fee using the estimates from Model 3. Figure 3 shows that the firm would maximize its revenues if it reduced the switching fee from the current level of MU 10 to MU 3.6. We find that this change would increase the firm's revenues by 3.9% (95% posterior interval [1.9%, 5.8%]). [9]

We next examine what share of this revenue increase is due to the effect of δ_i. We compute expected revenues using the same model estimates but now set δ_i to zero. Figure 3 (a) illustrates that the revenue increase from lowering the switching fee is largely due to customers' preference for free minutes. Figure 3(b) illustrates that at the optimal level of the switching fee of MU 3.6, this effect is significantly different from zero. In the absence of δ_i, reducing the switching fee to MU 3.6 would not significantly increase revenues since the posterior mean of the revenue increase is 1.5% with a 95% posterior interval of [−.2%, 3.3%] (Figure 3(c)). In conclusion, the increase in revenue obtained by reducing the switching fee hence encouraging

---

[9] An increase in revenue may not fully translate into an increase in profits if the firm does not have excess capacity. However, in this case the company's network was not operating at or close to capacity constraints.

customers to switch to three-part tariffs is mainly driven by customers' greater valuation for three-part tariff minutes.[10]

Adjusting the tariff structure

A firm might further consider removing two-part tariffs from their product offering while allowing current two-part tariff customers to remain on their tariff. We investigate whether such a decision would be beneficial if the firm also lowered the switching fee. We find an optimal switching fee of MU 2.2, which implies an expected revenue increase of 4.1% (95% posterior interval [1.8%, 6.4%]) and an average increase in customer surplus of .4% (95% posterior interval [.2%, .6%]). The revenue effect would be negligible, and even negative, if customers did not have a preference for "free" minutes (Figure 4). This result highlights the risk for a firm to misalign its pricing structure if it were to abstract from the value of "free" minutes to consumers.

Interestingly, the optimal level of the switching fee is lower than the optimal switching fee if the firm continued to offer two-part tariffs. This difference is driven by customers who previously would have switched to a different two-part tariff, provided the switching fee was sufficiently low. Now, since two-part tariffs are no longer available, these customers either remain on their current tariff, keeping their revenues constant, or switch to a three-part tariff where their revenue is likely to increase due to the preference for "free" minutes.

Since consumers may appreciate a simplification of a provider's pricing offering, we lastly examine the consequences of eliminating the switching fee entirely. If, as illustrated in Figure 4, the firm eliminated the switching fee and also removed the two-part tariffs, expected revenue would increase by 4.0% (95% posterior interval [1.5%, 6.5%])—only slightly less than at

[10] Note that our econometric model assumes that customers' choice decisions are based on the next period only. If customers were in fact forward-looking, our model would potentially over-estimate customers' sensitivity to switching costs and the effect of lowering the switching fee on provider revenues may be lower than predicted here. We run a sensitivity analysis in which we decrease the estimate for customers' sensitivity to the switching fee, ρ_1, by 5%, 10%, and 20%. Our findings are very robust to those changes (see web appendix).

the optimal level of the switching fee—and customer surplus would increase by .6% (95% posterior interval [.5%, .8%]). In contrast, if two-part tariffs were still on offer, eliminating the switching fee would lead to a revenue loss for the firm since consumers would be more likely to switch to two-part than to three-part tariffs (Figure 3).

CONCLUSION

Compared to two-part tariffs, three-part tariffs introduce "free" units of consumption. Behavioral research suggests that free products affect customers' behavior beyond what the change to the budget constraint would predict. In this research we examine how consumer demand changes when consumers switch from a two-part to a three-part tariff. We argue that the free component of a three-part tariff leads to a positive affective response increasing the valuation of the service, an effect that persists even after consumers have exceeded their usage allowance.

To identify the effect of tariff structure separately from the shift to the budget constraint, we jointly estimate tariff choice and usage. We explicitly model the effect of the usage allowance on the valuation of a three-part tariff. Our results confirm that the structure of the three-part tariff affects behavior. They indicate that the large majority of customers value "free" units beyond the change to their cost structure, significantly increasing firm's revenue. The additional valuation of "free" minutes accounts for 19.7% of the revenue generated from three-part tariff customers. We provide evidence that after switching to a three-part tariff, customers learn about their valuation over time. We find that the provider would benefit from reducing the switching fee which would lead more customers to switch to three-part tariffs, and from discontinuing the option to switch to two-part tariffs. Customers' valuation for "free" units is always key to any revenue increase.

Our findings have important implications. First, they suggest that companies should conduct field tests of new tariff structures to better understand how new tariffs affect customer behavior. This is important to set optimal prices and switching fees. Second, it may be beneficial for providers to advertise “free” minutes more intensely since these are highly valued by consumers. Third, companies that face capacity constraints should take into account that changing the tariff structure could lead to greater than expected usage of current capacity.

More broadly, our results point to a dual role of a nonlinear pricing plan: It not only determines the cost to the customer but also alters the perceived characteristics of the service, and so influences customers’ choice and usage. Overall, our findings add a new dimension to nonlinear pricing research that typically assumes that the difference in tariff structure affects usage exclusively through the budget constraint.

Our results motivate a more extensive study of how different tariff structures affect usage. A limitation of our study is that the data only spans customers who were with the firm before three-part tariffs were introduced. Future work could address customer acquisition and market expansion effects of introducing three-part tariffs and could examine how a firm can optimally combine three- and two-part tariffs. Furthermore, future research should investigate behavior when customers switch back to two-part tariffs. Since we have few periods in which we observe customers switching back, our data do not speak to whether usage declines beyond what a change to the budget constraint would predict, or whether greater usage alters preferences to such an extend that it becomes persistent, even once a customer leaves a three-part tariff. Additionally, while we have run multiple robustness checks and sensitivity analyses, we cannot conclusively rule out that other factors contribute to the phenomenon we observe. Future research could address this issue with experimental data, where all alternative accounts can be controlled for at the same time. Finally, future research could examine the effect of other tariff structures, such as bucket pricing that strictly limit consumption to the usage allowance (Schlereth and Skiera 2012), on consumers valuation of a service.

REFERENCES

Anderson, Eric and Duncan Simester (2003), "Effect of $9 Price Endings on Retail Sales: Evidence from Field Experiments," Quantitative Marketing and Economics, 1 (1), 93-110.

Bagh, Adib and Hemant K. Bhargava (2007), "Doing More with Less: Price Discrimination with Three-Part vs. Two-Part Tariffs," University of California, Davis.

Bertini, Marco and Luc Wathieu (2008), "Research Note - Attention Arousal Through Price Partitioning " Marketing Science, 27 (2), 236-46.

Cnet.com (1996), "New usage plan swamps AOL," http://news.cnet.com/New-usage-plan-swamps-AOL/2100-1023_3-251324.html.

Darke, Peter R. and Darren W. Dahl (2003), "Fairness and Discounts: The Subjective Value of a Bargain," Journal of Consumer Psychology, 13 (3), 328-38.

DellaVigna, Stefano and Ulrike Malmendier (2006), "Paying Not to Go to the Gym," American Economic Review, 96, 694-719.

Donovan, Robert J., John R. Rossiter, Gilian Marcoolyn, and Andrew Nesdale (1994), "Store Atmosphere and Purchasing Behavior," Journal of Retailing, 70 (3), 283-94.

Dubin, Jeffrey A. and Daniel L. McFadden (1984), "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," Econometrica, 52 (2), 345-62.

Erdem, Tülin and Michael P. Keane (1996), "Decision-Making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," Marketing Science, 15 (1), 1-20.

Golden, Linda and Mary R. Zimmer (1986), "Relationship Between Affect, Patronage, Frequency and Amount of Money Spent with a Comment on Affect Scaling and Measurement," Advances in Consumer Research, 13, 53-57.

Goettler, Ronald L. and Karen. B. Clay (2011), "Tariff Choice with Consumer Learning and Switching Costs," Journal of Marketing Research, 48 (4), 633{652.

Grubb, Michael D. (2009), Selling to Overconfident Consumers," American Economic Review, 2009, 99 (5), 1770-1807.

Grubb, Michael and Matthew Osborne (2011), "Tariff choice, usage uncertainty, biased beliefs, and learning," Working paper.

Hanemann, W.Michael (1984), "Discrete/Continuous Models of Consumer Demand," Econometrica, 52 (3), 541-62.

Heilman, Carrie M., Kent Nakamoto, and Ambar G. Rao (2002), "Pleasant Surprise: Consumer Response to Unexpected In-Store Coupons," Journal of Marketing Research, 39 (May), 242-52.

Isen, Alice M., Margaret Clark, Thomas E. Shalker, and Lynn Karp (1978), "Affect, Accessibility of Material in Memory, and Behavior: A Cognitive Loop?," Journal of Personality and Social Psychology, 36 (1), 1-12.

Iyengar, Raghuram, Kamel Jedidi, Skander Essegaier, and Peter J. Danaher (2011), "The Impact of Tariff Structure on Customer Retention, Usage and Profitability of Access Services," Marketing Science, 30 (5),820-836.

Iyengar, Raghuram, Asim Ansari, and Sunil Gupta (2007), "A Model of Consumer Learning for Service Quality and Usage," Journal of Marketing Research, 44 (4), 529-44.

Jensen, Sissel (2006), "Implementation of Competitive Nonlinear Pricing: Tariffs with Inclusive Consumption," Review of Economic Design, 10, 9-29.

Lambrecht, Anja, Katja Seim, and Bernd Skiera (2007), "Does Uncertainty Matter? Consumer Behavior under Three-Part Tariffs," Marketing Science, 26 (5), 698-710.

Lambrecht, Anja and Bernd Skiera (2006), "Paying Too Much and Being Happy About It: Existence, Causes and Consequences of Tariff-Choice Biases," Journal of Marketing Research, 43 (May), 212-23.

Messay, Cade. (1998), "America Online," mimeo.

Narayanan, Sridhar, Pradeep K. Chintagunta, and Eugenio J. Miravete (2007), "The Role of Self Selection and Usage Uncertainty in the Demand for Local Telephone Service," Quantitative Marketing and Economics, March (5), 1-34.

Nunes, Joseph C. (2000), "A Cognitive Model of People's Usage Estimations," Journal of Marketing Research, 37 (November), 397-409.

Park, Rolla Edward, B.M. Wetzel, and B.M. Mitchell (1983), "Price Elasticities for Local Telephone Calls," Econometrica, 51 (6), 1699-730.

Schlereth Christian and B. Skiera (2012), "Measurement of Consumer Preferences for Bucket Pricing Plans with Different Service Attributes," International Journal of Research in Marketing. Forthcoming.

Shampanier, Kristina, Nina Mazar, and Dan Ariely (2007), "Zero as a Special Price: The True Value of Free Products," Marketing Science, 26 (6), 745-57.

Sherman, Elaine and Ruth B. Smith (1987), "Mood States of Shoppers and Store Image: Promising Interactions and Possible Behavioral Effects," Advances in Consumer Research, 14, 251-54.

Thomas, Manoj and Vicky Morwitz (2005), "Penny wise and pound foolish: The left-digit effect in price cognition," Journal of Consumer Research, 32 (1), 54-64.

Wilson, Robert (1993), Nonlinear Pricing. New York: Oxford University Press.

Figure 1: USAGE BEFORE AND AFTER THE INTRODUCTION OF THE THREE-PART TARIFFS

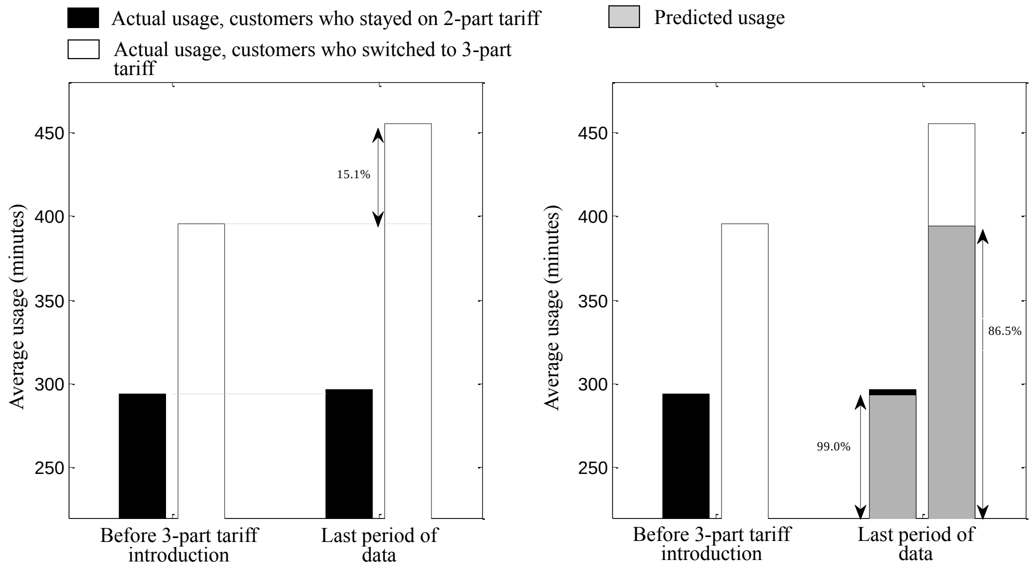


Figure 2: CORRELATION AMONG THE POSTERIOR INDIVIDUAL-LEVEL ESTIMATES FOR δ_i AND λ_i

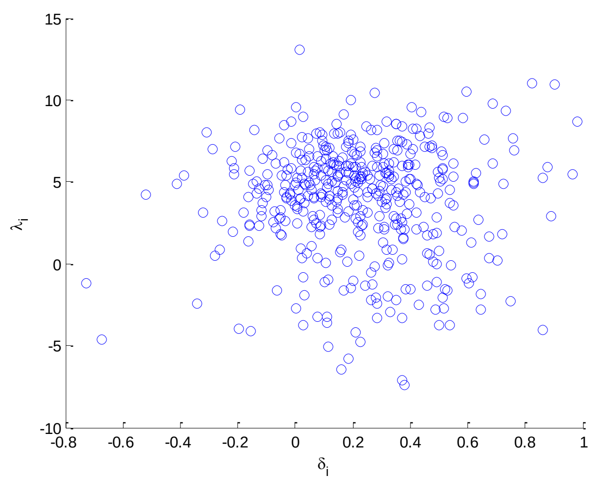


Figure 3: CHANGE IN REVENUE AND SURPLUS DUE TO REDUCTION OF THE SWITCHING FEE

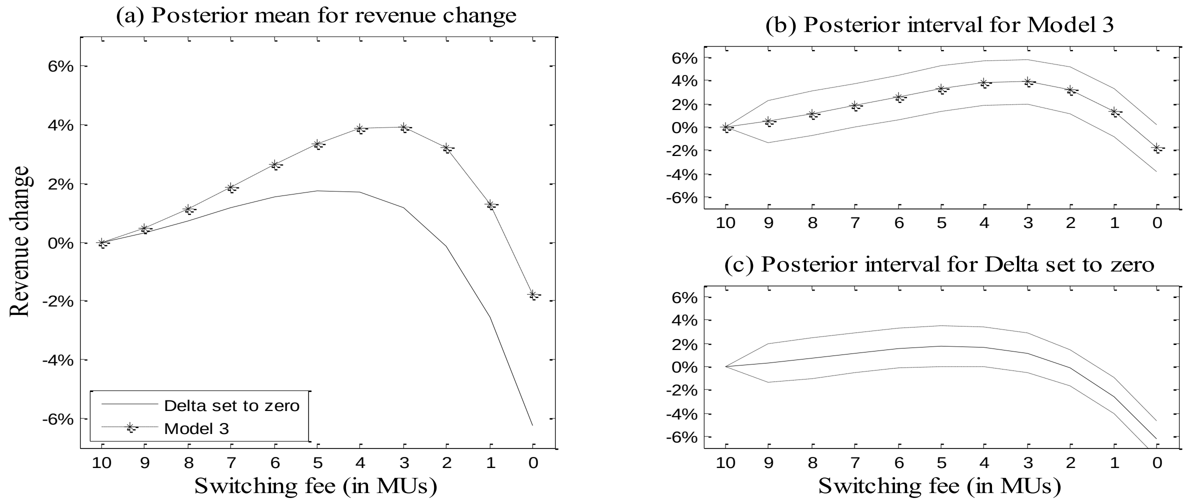


Figure 4: CHANGE IN REVENUE AND SURPLUS DUE TO CHANGES IN SWITCHING FEE AND TARIFF OFFERINGS

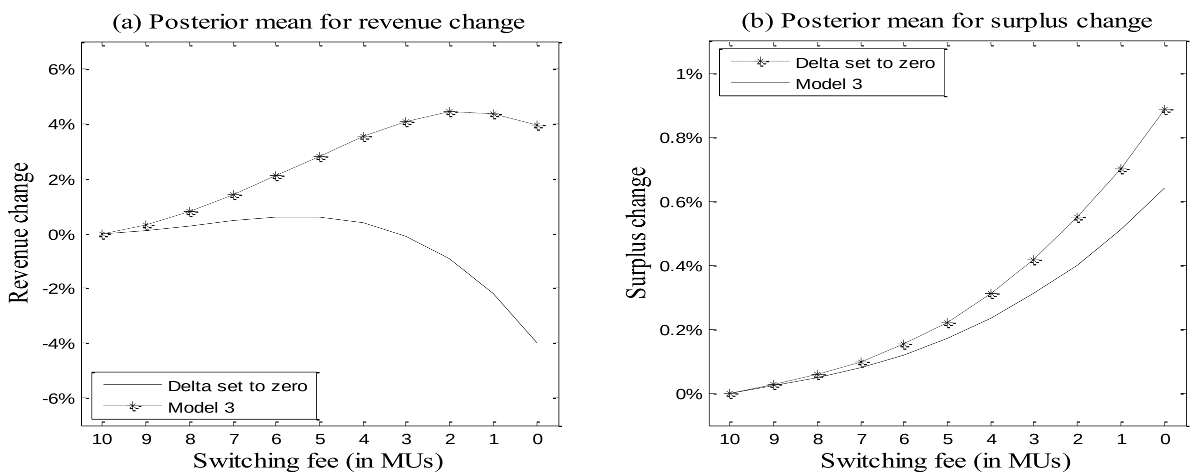


Figure 5: SEQUENCE OF MODEL DECISIONS

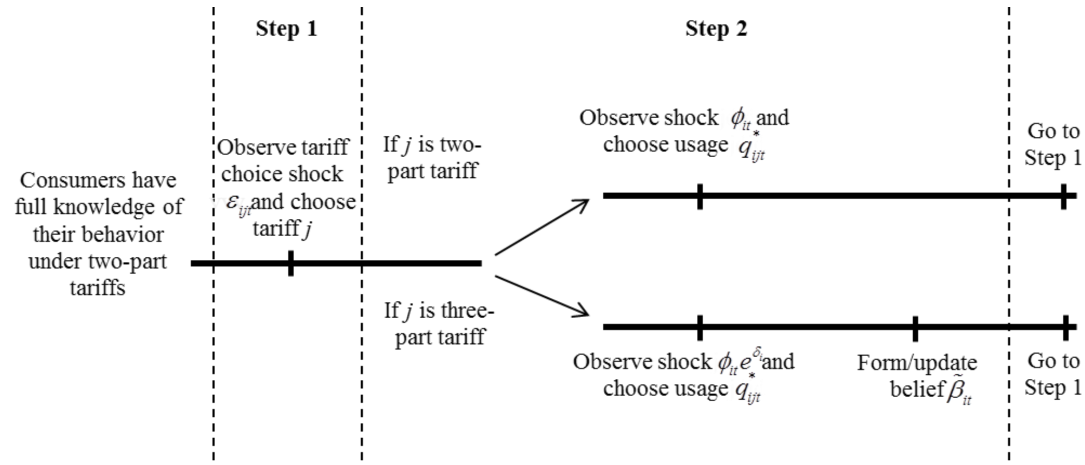


Table 1: TARIFF CHARACTERISTICS

Package	Access Price (MU)	Allowance (minutes)	Usage Price (MU)	Average Usage (Minutes)	Average Bill (MU)	Observations (Number)	Customers April 2007 (Number)
Tariff_2_1	1	0	.079	157.17	12.09	9786	766
Tariff_2_2	2	0	.056	220.45	13.19	9512	752
Tariff_2_3	3	0	.055	267.97	15.93	32166	2541
Tariff_2_4	5	0	.042	470.20	24.52	16057	1346
Tariff_3_1	10	200	.050	324.97	16.95	1896	321
Tariff_3_2	20	500	.050	730.39	33.25	231	47
Tariff_3_3	30	900	.050	1142.86	48.75	230	58

'MU' refers to the local currency

Table 2: TARIFF SWITCHING MATRIX

		Switching to ….						
		Tariff_2_1	Tariff_2_2	Tariff_2_3	Tariff_2_4	Tariff_3_1	Tariff_3_2	Tariff_3_3
Switching from ….	Tariff_2_1	87.3%	.0%	3.2%	2.8%	6.5%	.1%	.1%
	Tariff_2_2	.0%	91.1%	3.2%	1.8%	3.8%	.1%	.0%
	Tariff_2_3	.1%	.1%	85.6%	6.0%	7.1%	.6%	.6%
	Tariff_2_4	.2%	.0%	1.3%	89.0%	4.4%	2.2%	2.8%

*For customers who switched more than once, only the first switch is considered

Table 3: POSTERIOR DISTRIBUTIONS OF PARAMETER ESTIMATES

	Model 1			Model 2			Model 3		
	Mean	95% Interval		Mean	95% Interval		Mean	95% Interval	
Demand intercept									
Individual-level intercept									
Mean, μ_η	5.471	5.455	5.487	5.466	5.450	5.482	5.472	5.457	5.488
Std. dev., σ_η	.694	.682	.705	.694	.683	.705	.687	.676	.699
Dummy for holiday season	.039	.035	.044	.036	.032	.041	.036	.032	.040
Demand slope, b_i									
Intercept	239.913	223.661	256.461	244.791	229.500	261.232	271.026	255.085	285.935
Female	22.626	9.936	34.704	22.590	9.876	35.279	22.829	8.479	37.983
District-level literacy rate	-25.876	-34.992	-17.653	-26.412	-34.524	-17.274	-28.456	-37.191	-19.073
District-level employment rate	15.206	7.262	23.452	14.207	5.782	23.345	14.573	5.991	23.229
District is capital district	-16.042	-28.230	-3.084	-17.138	-30.327	-3.931	-16.755	-29.627	-3.839
Variance of usage shock, $1/r$	.217	.214	.220	.213	.210	.216	.205	.203	.208
Variance of initial belief, r/β_0							.212	.147	.296
Valuation of free units									
Mean, μ_δ				.211	.175	.246	.218	.176	.257
Std. dev., σ_δ				.373	.346	.403	.384	.351	.416
Preferences in tariff choice, ζ_{ijt}									
SC bw. tariffs, ρ_1	-1.025	-1.037	-1.012	-1.051	-1.065	-1.037	-1.025	-1.036	-1.014
SC to other provider, ρ_2	-47.651	-48.027	-47.290	-47.766	-48.094	-47.397	-47.573	-47.949	-47.181
Preference for the three-part tariff									
Mean, μ_λ	-4.916	-5.485	-4.417	-5.257	-5.853	-4.676	-6.024	-6.738	-5.367
Std. dev., σ_λ	5.714	5.392	6.076	6.310	5.937	6.705	6.412	6.032	6.832
Log Marginal Density	-394837			-394197			-394325		

N = 5,831 customers, 63,449 usage and 63,616 choice observations

Table 4: (IN-SAMPLE) PREDICTED VS. ACTUAL USAGE LEVELS (IN MINUTES)

	Model 1 – Prediction	Model 2 - Prediction	Model 3 - Prediction	Observed
- Two-part tariffs	**283**	**283**	**282**	**294**
Tariff_2_1	154	154	153	158
Tariff_2_2	230	230	229	221
Tariff_2_3	267	266	266	269
Tariff_2_4	267	266	266	269
- Three-part tariffs	**369**	**410**	**434**	**434**
Tariff_3_1	292	323	326	322
Tariff_3_2	600	642	722	721
Tariff_3_3	812	931	1084	1125

Table 5: INDIVIDUAL-LEVEL FIT MEASURES

	Model 1	Model 2	%Diff*	Model 3	%Diff*
MSE ('000)					
Full sample	46.94	46.33	-1.3	45.54	-3.0
Sample of two-part tariff usage	46.67	46.38	-.6	46.3	-.8
Sample of three-part tariff usage	56.85	44.6	-21.5	15.84	-72.1
MAPE					
Full sample	73.38	72.4	−1.2	71.87	-2.1
Sample of two-part tariff usage	73.91	72.9	−1.2	72.59	-1.8
Sample of three-part tariff usage	53.97	54.15	−1.0	43.54	-19.3
Hit rate (%)**	.98	.98	--	.98	--

* Percent change compared to Model 1
** We compare actual and predicted individual tariff choice in each period and report the average across all observations.

Table 6: CORRELATION BETWEEN λ_i AND δ_i

Number of three-part tariff observations per customer	Correlation	P-value	Number of customers
5	.007	.917	195
6	-.015	.864	139
7	-.153	.135	97

Table 7: SUMMARY OF ELASTICITIES – Model 3

Elasticity of	With respect to	All tariffs	2-part tariffs	3-part tariffs
Tariff choice	Access price	-2.955	-2.127	-13.276
Tariff choice	Usage price	-10.924	-11.395	-5.052
Tariff choice	Allowance			4.300
Usage	Usage price	-.094	-.100	-.027
Usage	Allowance			.843

APPENDIX

In this appendix we derive in greater detail the expression of the expected indirect utility for two- and three-part tariffs as well as the overall likelihood function.

Expected indirect utility:

Equation (9) shows the expected indirect utility of choosing a two-part tariff. In the estimation, we evaluate the probability of consuming zero minutes and taking expectations over the distribution of ϕ_{it}. From the demand Equation (3), it follows that the probability of zero usage is $\mathrm{P}(d_{ijt} \le b_i p_j) = \mathrm{P}\left(\phi_{it} \le \frac{b_i p_j}{e^{h_t a_1 + \eta_i}}\right)$, which corresponds to the cumulative distribution function (CDF) of a gamma distribution with parameters (r, r) evaluated at $\frac{b_i p_j}{e^{h_t a_1 + \eta_i}}$. To simplify notation, hereafter we denote $\mathrm{G}(x \mid \alpha_1, \alpha_2)$ as the CDF of a gamma distribution of shape parameter α_1 and scale parameter α_2, evaluated at x, and $\mathrm{g}(x \mid \alpha_1, \alpha_2)$ as the PDF of a gamma distribution defined as $\mathrm{g}(x \mid \alpha_1, \alpha_2) = \alpha_2^{\alpha_1} x^{\alpha_1 - 1} e^{-\alpha_2 x} / \Gamma(\alpha_1)$ The probability of observing zero usage is therefore $\mathrm{P}(q^*_{ijt} = 0) = \mathrm{G}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \mid r, r\right)$. The term $\phi_{it} \middle| \phi_{it} > \frac{b_i p_j}{e^{h_t a_1 + \eta_i}}$ follows a truncated gamma distribution. We therefore can express the expected value of the demand intercept on a two-part tariff, $\phi_{it} e^{h_t a_1 + \eta_i}$, conditional on usage greater than zero, as $\mathrm{E}\left(\phi_{it} e^{h_t a_1 + \eta_i} \mid q^*_{ijt} > 0\right) = e^{h_t a_1 + \eta_i} \frac{1 - \mathrm{G}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \mid r + 1, r\right)}{1 - \mathrm{G}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \mid r, r\right)}$. We substitute the above equations into (9) and obtain the expected indirect utility of a two-part tariff which determines a customer's two-part tariff choice,

(App-1)
$$\mathrm{E}[V_{ijt}] = \mathrm{G}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r, r\right) c_i (y_{it} - F_j) + \left(1 - \mathrm{G}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r, r\right)\right) \times c_i \left[y_{it} - F_j + \frac{1}{2} b_i p_j{}^2 - e^{h_t a_1 + \eta_i} \frac{1 - \mathrm{G}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r+1, r\right)}{1 - \mathrm{G}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r, r\right)} p_j \right] + \varsigma_{ijt}.$$

Similarly, we derive the expression of the expected indirect utility of choosing a three-part tariff (from Equation 10). Substituting Equation ((*8*) into Equation ((*4*), we obtain that $\mathrm{P}(q^*_{ijt} \le \tilde{q}_j) = \mathrm{P}\left(\phi_{it} e^{\delta_i} \le \frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}}\right)$. Thus, the probability that a customer's three-part tariff usage is equal to or below the allowance is $\mathrm{P}(q^*_{ijt} \le \tilde{q}_j) = \mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r, \beta_i\right)$. The term $\phi_{it} e^{\delta_i} \left| \phi_{it} e^{\delta_i} > \frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}} \right.$ follows a truncated gamma distribution with parameters (r, β_i). We express the expected indirect utility of a three-part tariff as

(App-2)
$$\mathrm{E}[V_{ijt}] = \mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r, \beta_i\right) c_i (y_{it} - F_j) + \left(1 - \mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r, \beta_i\right)\right) \times c_i \left[y_{it} - F_j + p_j \tilde{q}_j + \frac{1}{2} b_i p_j{}^2 \frac{\mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}}\right)}{\mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}}\right)} - e^{h_t a_1 + \eta_i} \frac{r}{\beta_i} \frac{1 - \mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r+1, \beta_i\right)}{1 - \mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}} \,|\, r, \beta_i\right)} p_j \right] + \varsigma_{ijt}.$$

A customer who has not yet switched to a three-part tariff is not aware of its additional value so her choice is unaffected by δ_i. Hence, at her initial three-part tariff choice, δ_i does not enter her expected indirect utility of a three-part tariff. She therefore takes expectations over the shock ϕ_{it}, and the expected indirect utility of a three-part tariff simplifies to

(App-3)

$$
\begin{aligned}
\mathrm{E}[V_{ijt}] = {} & \mathrm{G}\left(\frac{\tilde{q}_j + \ \ p_j}{e^{h_t a_1 + \eta_i}} \mid r, r\right) c_i (y_{it} - F_j) + \left(1 - \mathrm{G}\left(\frac{\tilde{q}_j + \ \ p_j}{e^{h_t a_1 + \eta_i}} \mid r, r\right)\right) \\
& \times c_i \left[y_{it} - F_j + p_j \tilde{q}_j + \frac{1}{2} b_i p_j^{\ 2} \frac{\mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}}\right)}{\mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i}}\right)} \right. \\
& \left. - e^{h_t a_1 + \eta_i} \frac{1 - \mathrm{G}\left(\frac{\tilde{q}_j + \ \ p_j}{e^{h_t a_1 + \eta_i}} \mid r + 1, r\right)}{1 - \mathrm{G}\left(\frac{\tilde{q}_j + \ \ p_j}{e^{h_t a_1 + \eta_i}} \mid r, r\right)} p_j \right] + \varsigma_{ijt}.
\end{aligned}
$$

Likelihood function:

For every customer i and time t we observe usage, q_{ijt}, and tariff choice, k_{it}, which are outcomes of the customer and time-specific covariates, $Z_{it} = \{h_t, d_i\}$, the tariff-specific characteristics, $X_j = \left\{p_j, F_j, \tilde{q}_j\right\}$, the population parameters, $\Phi = \left\{b, \rho_1, \rho_2, \beta_0, a_1, a_2, \exp(r)\right\}$, the time-invariant individual parameters, $\varpi_i = \left\{\eta_i, \delta_i, \lambda_i\right\}$, and the individual-specific time-variant beliefs $\tilde{\beta}_{it}$.

<u>Likelihood of usage:</u> For a consumer on a two-part tariff, the probability of observing a particular usage level, given the tariff choice, is

(App-4)

$$
\begin{aligned}
f(q_{it} \mid k_{it}, \Phi, \varpi_i, Z_{it}, X_j) &= P(q_{ijt}^* = q_{ijt} \mid k_{it} = j, \Phi, \varpi_i, Z_{it}, X_j) \\
&= P(q_{ijt}^* = 0) I^{\{q_{it} = 0\}} + P(q_{ijt}^* = \phi_{it} e^{h_t a_1 + \eta_i} - b_i p_j) I^{\{q_{it} > 0\}} \\
&= \mathrm{G}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \mid r, r\right) + \Im_{it} \ \mathrm{g}\left(\frac{b_i p_j}{e^{h_t a_1 + \eta_i}} \mid r, r\right) I^{\{q_{it} > \tilde{q}_j\}}
\end{aligned}
$$

where the first term corresponds to the probability of observing zero usage, expressed as the CDF of a gamma distribution with shape and scale parameter r, evaluated at $\frac{b_i p_j}{e^{h_t a_1 + \eta_i}}$. The term $I^{\{A\}}$ is the indicator function that takes value 1 if statement A is true, zero otherwise. The second term is divided into two parts, the Jacobian of the transformation from q_{ijt} to ϕ_{it}, $\Im_{it} = \frac{1}{e^{h_t a_1 + \eta_i}}$, and the PDF of ϕ_{it}, which is distributed $\text{Gamma}(r, r)$.

For a consumer on a three-part tariff, the probability of observing a particular usage level is

$$f(q_{it} \mid k_{it}, \Phi, \varpi_i, Z_{it}, X_j) = P(q_{ijt}^* = q_{ijt} \mid k_{it} = j, \Phi, \varpi_i, Z_{it}, X_j)$$

$$= P(q_{ijt}^* = \phi_{it} e^{z_{it}' a_i + \eta_i + \delta_i}) I^{\{q_{it} < \tilde{q}_j\}} + P(q_{ijt}^* = \tilde{q}_j) I^{\{q_{it} = \tilde{q}_j\}}$$

$$+ P(q_{ijt}^* = \phi_{it} e^{z_{it}' a_i + \eta_i + \delta_i} - b_i p_j) I^{\{q_{it} > \tilde{q}_j\}}$$

(App-5)

$$= \Im_{it} \ \mathrm{g}\left(\frac{q_j}{e^{h_t a_1 + \eta_i + \delta_i}} \mid r, r \right) I^{\{q_{it} < \tilde{q}_j\}}$$

$$+ \left\{ \mathrm{G}\left(\frac{\tilde{q}_j + b_i p_j}{e^{h_t a_1 + \eta_i + \delta_i}} \mid r, r \right) - \mathrm{G}\left(\frac{\tilde{q}_j}{e^{h_t a_1 + \eta_i}} \mid r, r \right) \right\} I^{\{q_{it} = \tilde{q}_j\}}$$

$$+ \Im_{it} \ \mathrm{g}\left(\frac{q_j + b_i p_j}{e^{h_t a_1 + \eta_i + \delta_i}} \mid r, r \right) I^{\{q_{it} > \tilde{q}_j\}},$$

where $\Im_{it} = \frac{1}{e^{h_t a_1 + \eta_i + \delta_i}}$ is the Jacobian of the transformation from q_{ijt} to ϕ_{it}.

__Likelihood of tariff choice__: The tariff-specific shock ε_{ijt} is assumed to follow a Type 1 extreme value distribution. Therefore, the probability of choosing a particular tariff *j* is given by

(App-6)

$$f(k_{it} \mid \alpha, \varpi_i, \tilde{\beta}_{it}, Z_{it}, X_j) = P(k_{it} = j \mid \Phi, \varpi_i, \tilde{\beta}_{it}, Z_{it}, X_j) = \prod_{j=0}^{J} \frac{e^{\tilde{V}_{ijt}}}{\sum_{g=0}^{J} e^{\tilde{V}_{igt}}},$$

where the term $\tilde{V}_{ijt}$ denotes the expected indirect utility of each tariff *j*. We obtain the likelihood function by integrating the customer's tariff choice and usage decisions:

(App-7)

$$L = \prod_{i=1}^{I} \prod_{t=1}^{T_i} [f(q_{it} \mid k_{it}, \alpha, \varpi_i, Z_{it}, X_j) f(k_{it} \mid \alpha, \varpi_i, \tilde{\beta}_{it}, Z_{it}, X_j)].$$

When Talk is "Free": The Effect of Tariff Structure on Usage under Two- and Three-Part Tariffs

Eva Ascarza, Anja Lambrecht, and Naufel Vilcassim

Web Appendix

In this appendix, we present a detailed description of the analyses performed to obtain certain results discussed in the main manuscript.

DESCRIPTIVE ANALYSES

Analysis of tariff choice

We analyze whether customers' bills would have been lower on another than their chosen tariff at the time that three-part tariffs were introduced. Based on the three available usage periods prior to the three-part tariff introduction, we compute the individual-level average usage and standard deviation. For simplicity, we exclude customers who have switched more than once as well as the 1.1% of customers who switched within these three months.

To account for deviations from average usage due to random usage shocks, we then compute the bill for the usage level of [average usage +/− 1 standard deviation] under the current tariff, and the bill for the average usage under each of the remaining tariffs. We conclude that a customer would have had a lower bill on a different tariff if the bill for their average usage on a tariff other than the chosen tariff was below the lower bound of the bill-interval that accounts for variation in usage on the chosen tariff. Note that this analysis focuses on potential savings and does not account for the fact that customers may, on the same bill, be able to use more on a different tariff. The next section will discuss this aspect in detail.

Table A1 illustrates that based on their average usage and standard deviation of usage before the introduction of three-part tariffs, the large majority of customers chose the tariff that minimizes their bill. In total, 26.2% of customers would pay less on a different tariff. For customers that would pay less on a different tariff average savings were between MU 4.1 and MU 7.7. As a result, it would take customers more than one period on average to amortize the switching fee of MU 10.

We then exclude three-part tariffs from this analysis and limit the analysis to whether customers would have paid less on a different two-part tariff. We find that only 10.9% of customers would have paid less on a different two-part tariff. This further confirms that two-part tariff customers largely chose the bill-minimizing tariff.

Table A1: Potential savings when three-part tariffs were introduced

	Tariff with lowest bill (in %)								
Chosen tariff	T_2_1	T_2_2	T_2_3	T_2_4	T_3_1	T_3_2	T_3_3	N	Avg. savings (in MU)*
Tariff_2_1	78.8	5.5	0.0	0.0	13.4	1.7	0.6	850	7.7
Tariff_2_2	0.3	87.0	0.0	0.0	9.1	1.8	1.8	814	6.3
Tariff_2_3	2.5	3.7	72.5	0.0	13.9	4.8	2.7	2,815	5.7
Tariff_2_4	1.4	9.4	0.0	65.0	12.4	6.2	5.6	1,253	4.1

Excluding customers who switched within the first 3 months of our data and customers that in our data switch more than once

* Average savings on tariff with lowest bill, computed only for customers that would have had a lower bill on a different tariff

Detailed analysis of switching from two- to three-part tariffs

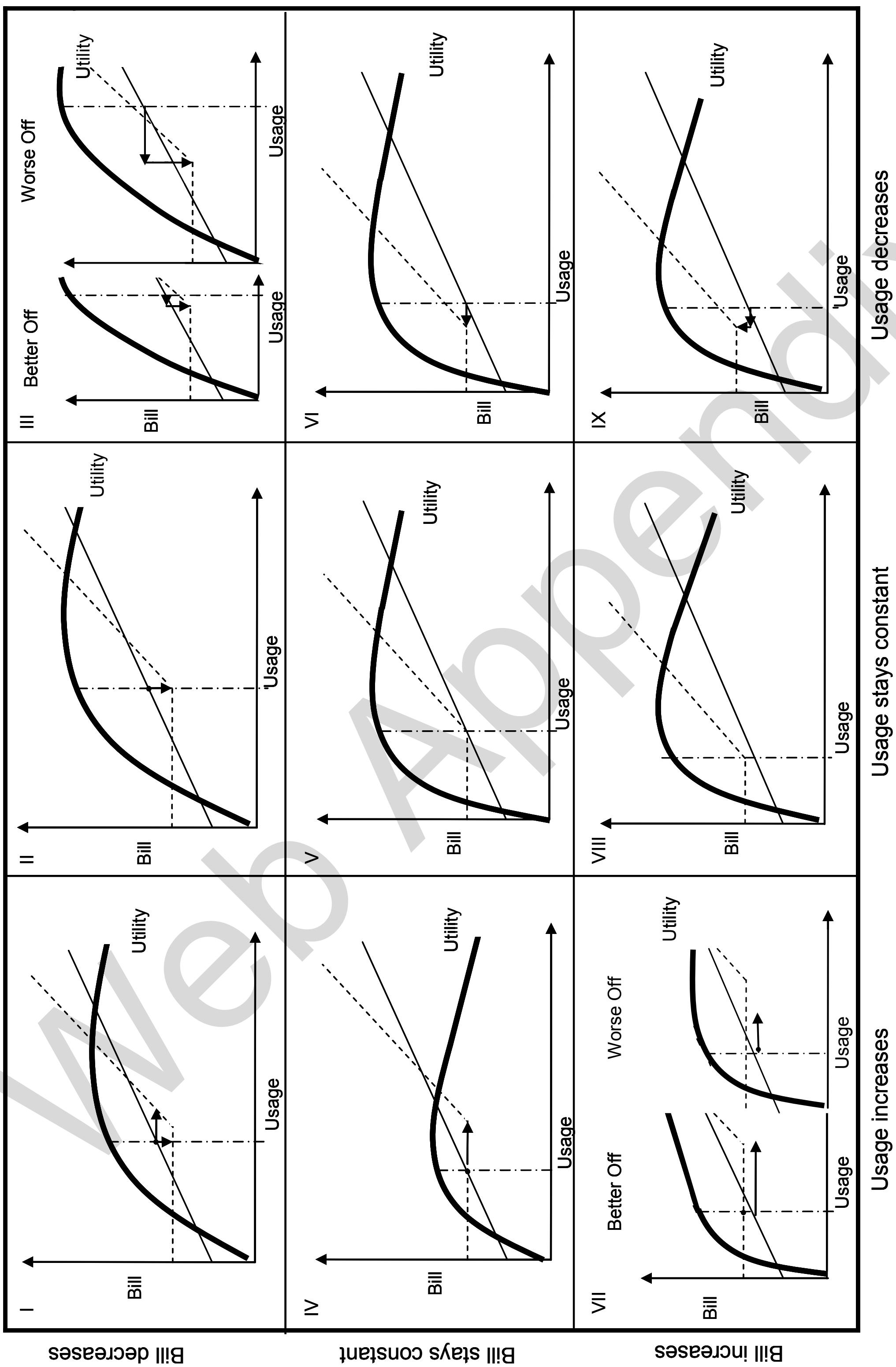
The previous section focused on whether customers would have *paid less* on a different tariff. We now focus on three-part tariffs and analyze in more detail whether customers would benefit from switching to a three-part tariff, accounting for both whether customers would have *paid less* on a different tariff and whether they would have been able to *use more* for the same bill.

Figure A1 illustrates in which situations a customer should or should not switch to a three-part tariff. We abstract from switching costs and assume that a customer knows her optimal usage under a two- and a three-part tariff. We assume a utility function which is quadratic in usage (bold curve; the Model section of the main paper justifies the choice of utility function). The bill on a two-part tariff (straight line) increases in the customer's usage. The bill on a three-part tariff (dashed line) remains flat as long as usage remains within the allowance and then increases linearly in usage. The maximum distance between the utility function and the bill indicates a customer's surplus on that tariff. A rational customer should switch to a three-part tariff if that entails a greater surplus than on a two-part tariff.

The vertical (horizontal) arrows indicate how such a switch would affect a customer's bill (usage). A customer should switch to a three-part tariff if for the same optimal usage, she pays less on a three- than on a two-part tariff (II), for the same bill, her optimal usage is greater on a three- than on a two-part tariff (IV), or if she can increase her optimal usage and still pay less on a three-part tariff (I). A customer should not switch if for the same bill, her optimal usage on a three-part tariff would decrease (VI), for the same optimal usage her bill would increase (VIII) or if her bill would increase while decreasing optimal usage (IX). She is indifferent if the same optimal usage entails the same bill (V). If under a three-part tariff, both optimal usage and the bill would decrease (III) or increase (VII), switching may or may not be beneficial, depending on the curvature of the utility function.

To determine which customers in our sample should or should not switch to a three-part tariff, we compare actual usage and expenditures on a two-part tariff to (a) how much a customer could use under a three-part tariff for the same bill and (b) how much she would pay under a three-part tariff for the same usage (Figure A1). To account for deviations from average usage due to random usage shocks, the interval of [average usage +/− 1 standard deviation] and the interval of the bill of [average usage +/− 1 standard deviation] serve as a reference point. For example, we classify a customer as being indifferent between switching to a three-part tariff and staying on a two-part tariff (Case V) if the same optimal usage entails a bill in the same interval on a two- and a three-part tariff.

Figure A1: Predicted switching from two- to three-part tariffs



The maximum distance between the utility function and a tariff's bill indicates maximum surplus.
The vertical dotted line represents the optimal level of usage on a two-part tariff.
Note: Marginal utility and the price of the outside good are set to 1, so utility represents willingness to pay.

Table A2 summarizes the results of this analysis. The first four columns correspond to the results when the switching fee is not taken into consideration. They indicate that customers who, according to our analysis, should switch to a three-part tariff were far more likely to switch to a three-part tariff than customers who according to our analysis should not switch to a three-part tariff. The next set of results accounts for the fee the customer has to pay for switching. Here we consider a switch to be beneficial if savings in the first month would compensate for the switching fee. Since the fee increases the bill, the share of customers classified as "unknown", i.e., those for whom both optimal usage and the bill would increase on a three-part tariff, is larger than when abstracting from the switching fee.

Table A2: Predicted and actual switching behavior

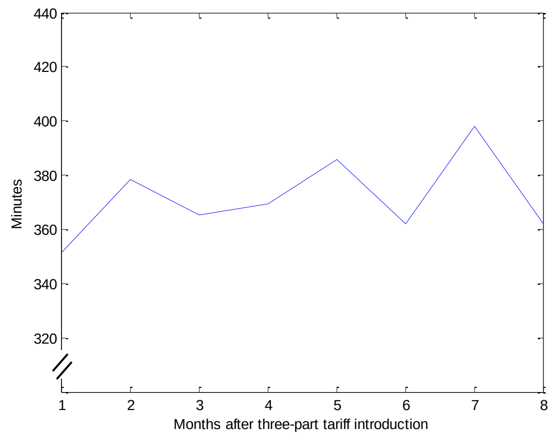
	Not considering switching fee				**Considering switching fee**			
Category	No. of customers	% of sample	% of customers in that group who switched	% of total switchers belonging to category	No. of customers	% of sample	% of customers in that group who switched	% of total switchers belonging to category
Should switch	3,710	63.7%	8.95%	71.7%	980	16.8%	13.2%	27.9%
Should not switch	85	1.5%	4.71%	0.9%	704	12.1%	6.4%	9.7%
Indifferent[(a)]	765	13.1%	7.19%	11.9%	211	3.6%	10.4%	4.8%
Unknown[(b)]	1271	21.8%	5.66%	15.6%	3936	67.5%	6.8%	57.7%

(a) A customer is indifferent if the same optimal usage entails the same bill on a two- and a three-part tariff.
(b) If under a three-part tariff, both optimal usage and the bill would decrease or increase, switching may or may not be beneficial depending on the curvature of the utility function.

Persistence of three-part tariff usage over time

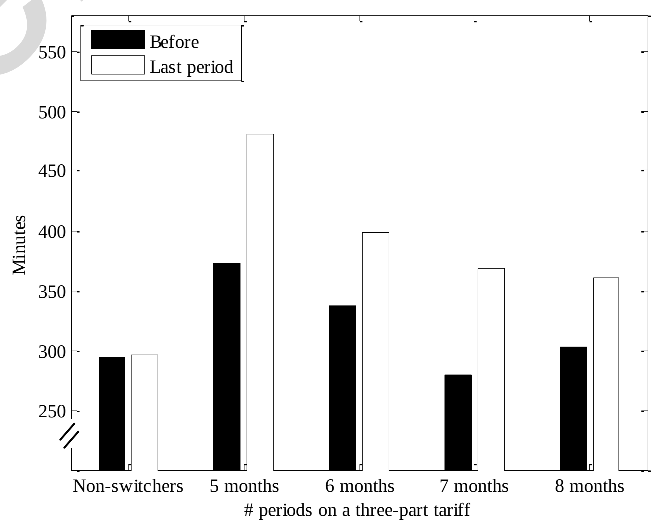
We next check whether the increase in three-part tariff usage persists over time. We focus on customers for whom we observe at least six months of three-part tariff usage and plot the aggregate three-part tariff usage over time. Figure A2 illustrates that, apart from the holiday seasons in months 5 and 7 after the introduction of the three-part tariffs, there are no clear trends of increasing or decreasing three-part tariff usage.

Figure A2: Monthly average usage after switching to a three-part tariff



Second, we compare average usage before and after the three-part tariff introduction, as we do in the Descriptive Analysis section of the main manuscript, but now analyze differences by cohorts (i.e., groups of customers who switched to a three-part tariff in the same month). Figure A3 shows, for each cohort, the average usage before the three-part tariffs were introduced and the average usage in the last period of our data and compares it to customers who did not switch to a three-part tariff. We observe a consistent increment in usage among three-part tariff switchers, regardless of how long customers have been on a three-part tariff.

Figure A3: Average usage before and after the introduction of three-part tariffs, by cohorts

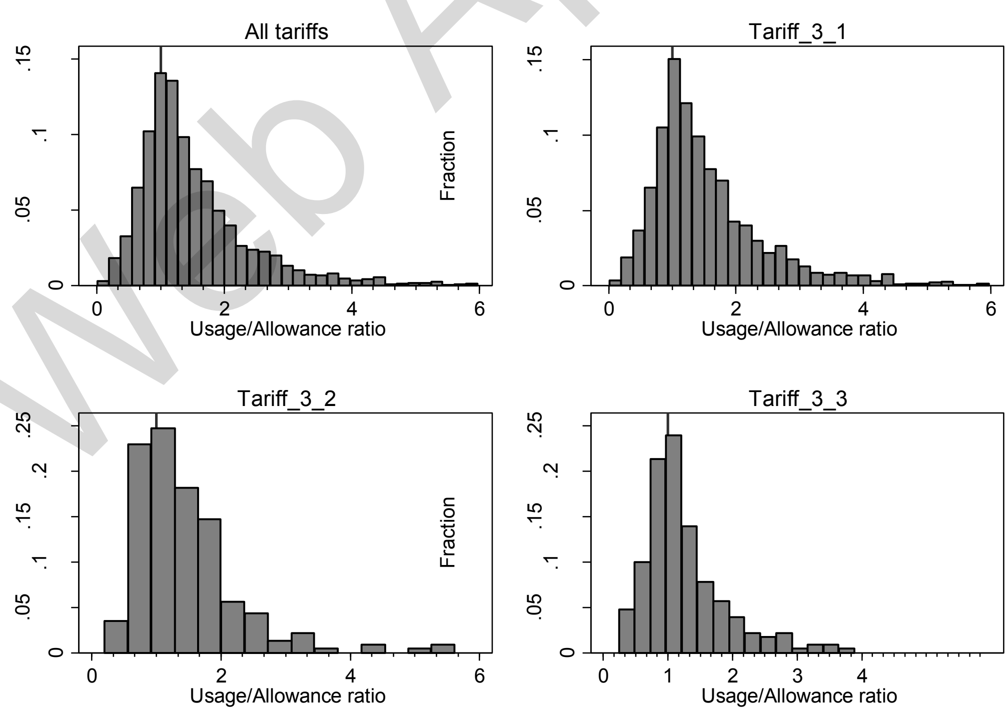


Further details on three-part tariff usage behavior

We summarize information on customers' usage behavior on three-part tariffs. First, we analyze the distribution of usage as percentage of the allowance (Figure A4). Across all three-part tariffs, we observe a mass point of usage observations when usage is approximately equal to the allowance. This mass point results from the type of budget constraint imposed by a three-part tariff that implies bunching of usage observations at 100% of the allowance (see equation (4) in the main manuscript). It is reassuring that we indeed find such a mass point in our data since it provides additional evidence that customers are aware of their usage behavior.

Figure A4 also illustrates that many customers use more than their usage allowance. This is in line with the behavioral motivation that leads to greater three-part tariff usage as discussed in the main paper. It outlines that the positive effect from a three-part tariff should persist when consumers have exceeded their allowance.

Figure A4: Usage as a percent of allowance



Second, we analyze whether three-part tariff customers had chosen the ex-post bill-minimizing tariff based on their first three months of three-part tariff usage. As in the first section of this web appendix, we rely on the bill for the usage level of [average usage +/- 1 standard deviation] under the current tariff, and the bill for the average usage under each available tariff. Table A3 illustrates that overall 86.8% of customers chose the three-part tariff that minimizes their bill based on their *ex post* usage. Since the differences between access prices and allowances between the three-part tariffs are large, even customers that use more than their allowance are largely in the bill-minimizing tariff.

Table A3: Optimality of chosen three-part tariff (based on first three periods on a three-part tariff)

	Tariff with lowest bill (in %)				
Chosen tariff	Two-part tariff	T_3_1	T_3_2	T_3_3	N
Tariff_3_1	5.7	86.8	5.1	2.4	296
Tariff_3_2	0.0	6.1	81.8	12.1	33
Tariff_3_3	0.0	0.0	8.8	91.2	34

Includes all customers with at least three periods on a three-part tariff, excludes customers who switched again in their first three periods on a three-part tariff

DEMAND ESTIMATION

Linear demand estimation of three-part tariff usage

We compare actual usage on two- and three-part tariffs to predicted usage for the last month in our data. We estimate a linear demand function for two-part tariff usage, $q_{ijt} = d_{it} - bp_j$, where q_{ijt} denotes the number of minutes that individual *i* consumes on tariff *j* at time *t*, d_{it} denotes the satiation level, or demand intercept, b refers to the price coefficient and p_j is the usage price of tariff *j*. Since we have little within-customer variation of the usage price, the price coefficient is assumed to be homogenous across customers. We incorporate an individual-level preference, η_i, and a multiplicative shock, ϕ_{it}, into the demand intercept, $d_{it} = \phi_{it} e^{\eta_i}$. We assume that η_i follows a normal distribution with mean and variance $\left(\mu_\eta, \sigma_\eta^2\right)$ and that ϕ_{it} is distributed lognormal with parameters $\left(-0.5\sigma_\phi^2, \sigma_\phi^2\right)$, such that $E(\phi_{it}) = 1$. We use MCMC methods to estimate the model. We choose diffuse hyperpriors for b, μ_η, σ_η ,and σ_ϕ. We burn-in 90,000 iterations and use the next 10,000 to sample from the posterior distributions of the parameters of interest and to predict consumption in the last period of data. The parameters estimates are shown in Table A4.

Table A4: Estimation Results (Homogeneous price coefficient)

	Mean	95% Interval	
b	421.158	371.841	478.932
μ_η	5.519	5.498	5.545
σ_η	0.679	0.662	0.696
σ_ϕ	0.308	0.300	0.316

For customers who remained on a two-part tariff, we predict consumption in the last period of the data as:

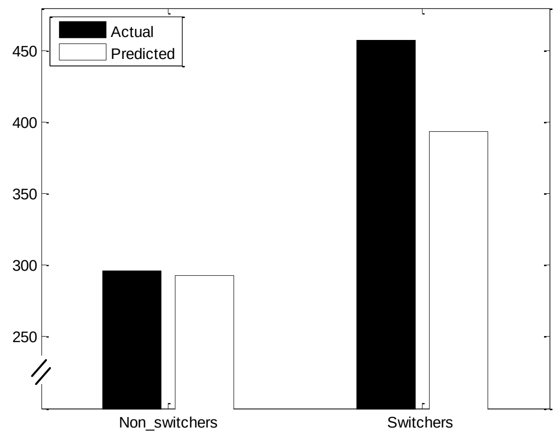
$$q_{ijt}^* = \begin{cases} 0 & \text{if } d_{ijt} \le bp_j \\ d_{ijt} - bp_j & \text{if } d_{ijt} > bp_j. \end{cases}$$

For customers who have switched to a three-part tariff, we predict consumption in the last period of the data as

$$q_{ijt}^* = \begin{cases} d_{ijt} & \text{if } d_{ijt} < \tilde{q}_j \\ \text{Max}(\tilde{q}_j, d_{ijt} - bp_j) & \text{if } d_{ijt} > \tilde{q}_j \end{cases}$$

Figure A5 illustrates that the model accurately predicts usage for customers who remain on a two-part tariff while notably underpredicting consumption for customers who switch to a three-part tariff. In other words, the model does not capture the increment in usage observed for three-part tariff customers.

Figure A5: Usage predictions using linear model (all customers)



Persistence of over-usage over time

We next check whether the unpredicted increase in three-part tariff usage persists over time. We use the estimates obtained in the analysis presented in the previous section but now analyze three-part tariff customers in cohorts of customers who switched to a three-part tariff in the same month. For each cohort, we predict usage in the last month of the data and compare it with actual usage in that month. The model under-predicts three-part tariff usage regardless of how long customers have been on a three-part tariff. Specifically, we under-predict usage by 22.1% for the five-month cohort, by 12.7% for the six-month cohort, by 19.8% for the seven-month cohort, and by 12.1% for the eight-month cohort.

Robustness to non-linear demand specifications

If customers' usage followed a convex demand function, our linear demand model in the previous section would predict demand accurately in the area of usage prices similar to those of the two-part tariffs, i.e., 0.042–0.079 MU, but would possibly underpredict usage at a zero price. As a consequence, the over-usage we find in the descriptive analysis presented in the main manuscript could simply be due to the specification of the demand function. We rule out this possibility by estimating two additional demand specifications.

First, we use a polynomial specification (as a Taylor approximation to the true demand function) to estimate demand. We build on the demand function presented in the previous section, $q_{ijt} = d_{it} - bp_j$, and include a quadratic term, $b_2 p_j^2$, and a cubic term $b_3 p_j^3$. We estimate demand as $q_{ijt} = d_{it} - b_1 p_j - b_2 p_j^2 - b_3 p_j^3$. If the quadratic and cubic terms do not significantly differ from zero, that would support the choice of a linear demand function.

We replicate the analysis presented in the main manuscript. The results show that the quadratic and cubic terms of the demand function are not significantly different from zero (Table A5).

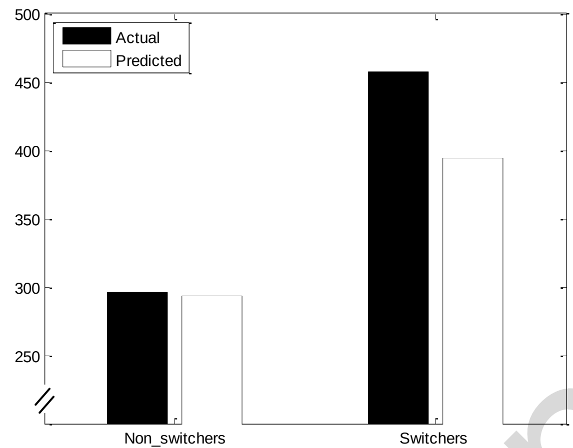
We next use the parameter estimates to predict usage in the last period.

Figure A6 displays the results. Similarly to our main specification, predicted usage of customers who switched to a three-part tariff is only 86.4% of their actual usage while the model predicts 98.9% of actual usage for customers who remain on a two-part tariff. This provides evidence that the increase in usage is not due to the specific form of the demand function.

Table A5: Estimation results (quadratic and cubic terms)

	Mean	95% Posterior Interval	
μ_η	5.519	5.504	5.534
σ_η	0.674	0.659	0.689
b_1	442.398	396.798	486.488
b_2	4.742	-49.264	60.743
b_3	-1.984	-63.673	62.446
σ_ϕ	0.306	0.299	0.315

Figure A6: Usage predictions using quadratic and cubic terms



Second, we estimate an additional model specification that allows for convex demand: $q_{ijt} = \frac{e^{\eta_i + \nu_{it}}}{p_j - \gamma} - \beta$. This demand specification is obtained by maximizing the utility function $U_{ijt}(q_{ijt}, q_{i0t}) = \alpha \log(q_{ijt} + \beta) + \gamma q_{ijt} + q_{i0t}$, with $\alpha, \beta > 0$ and $\gamma < 0$. The term q_{i0t} denotes the outside good, when its price is being normalized to 1.

To empirically disentangle η_i and $, \gamma$ the data needs to have individual-level variation of the usage price. However, in our data there is little tariff switching before the three-part tariffs were introduced. An alternative is to fix the value of γ at a reasonable level and estimate the remaining parameters based on the first two periods and predict usage for the last period. We proceed in three steps:

1. To avoid having to arbitrarily set γ, we estimate the demand model using all observations from the first six periods of data. We obtain an estimate of γ (-0.049).[1]

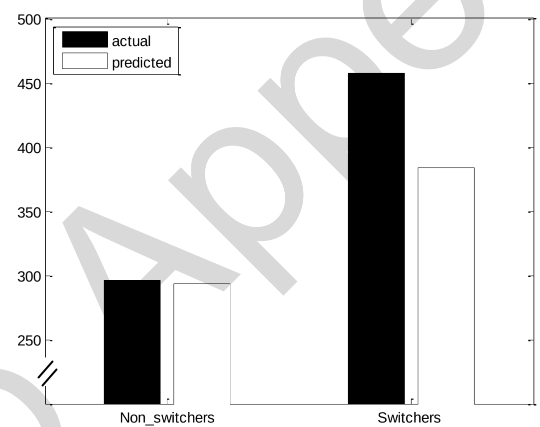
---

1 We conduct two sets of robustness checks to our estimate of γ. First, we estimate γ based on a different number of periods (4 and 6 periods). Second, we estimate γ based on a random subsample of 50% of the customers in our dataset. We find that our estimate of γ is robust to these alternative specifications.

2. We then set γ and estimate the remaining parameters, including η_i, using two-part tariff usage observations prior to the three-part tariff introduction.
3. We then use the set of estimated parameters to predict usage in the last period of our data.

Figure A7 illustrates predicted versus actual usage. Consistent with the results obtained in the previous section, we under-predict three-part tariff usage by 19.2% while predicting two-part tariff usage very accurately (under-prediction of only 0.9%). This provides further evidence that the assumption of linear demand does not lead us to artificially under-predict three-part tariff usage.

Figure A7: Usage prediction convex demand function



Robustness to non-homogeneous price sensitivity

It is possible that customers who switch to a three-part tariff differ in their usage price sensitivity from customers who remain on a two-part tariff. Given the limited within-customer price variation in our data, we cannot estimate a model with an individual-level price coefficient, b_i. Nevertheless, we conduct an *ad hoc* analysis in which we allow for a different set of parameters for switchers to a three-part tariff compared to all other customers. We then test whether this specification still under-predicts three-part tariff usage.

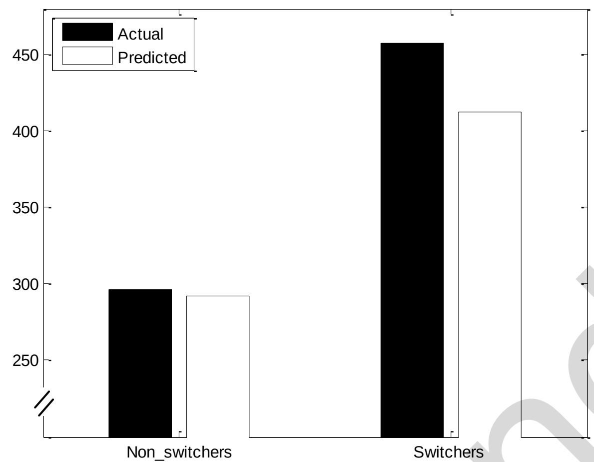
As in the Descriptive Analysis section in the main manuscript, we estimate a demand model using the two-part tariff periods prior to the three-part tariff introduction and then predict usage in the last period of our data. We now estimate two sets of coefficients, one for customers who remained on a two-part tariff and one for customers who switched to a three-part tariff. The same diffuse priors were chosen for both sets of parameters. Table A6 summarizes the posterior distributions and Figure A8 shows the model predictions.

The model with heterogeneity in price sensitivity under-predicts three-part tariff usage by 9.8% while two-part tariff usage is predicted very accurately. We conclude that while heterogeneity in usage price sensitivity may possibly contribute to greater three-part tariff usage, it does not explain the large increase in usage we observe in the data. To capture some degree of heterogeneity in usage price sensitivity, our full model specification (the Model section of the main manuscript) incorporates observed heterogeneity in the usage price sensitivity.

Table A6: Estimation results (heterogeneous price coefficient)

	Customers who do not switch to a three-part tariff			Customers who switch to a three-part tariff		
	Mean	95% Interval		Mean	95% Interval	
b	422.556	368.723	476.531	764.364	542.254	1027.587
μ_η	5.501	5.478	5.525	5.817	5.733	5.895
σ_η	0.668	0.651	0.686	0.700	0.649	0.739
σ_ϕ	0.307	0.299	0.314	0.294	0.268	0.323

Figure A8: Usage predictions for heterogeneity in price sensitivity (linear model)



Analysis of autocorrelation in the usage process leading to self-selection

As discussed in the main manuscript, autocorrelation in the usage process could be a possible explanation for the usage increase we observe. If usage followed an autoregressive process and customers switched to a three-part tariff after having received a positive usage shock, then we would expect that customers increase their consumption after switching to a three-part tariff.

However, we find that this pattern of behavior is not consistent with our data.

We first investigate the level of autocorrelation among the usage shocks. Given that our demand is specified with multiplicative usage shocks in the demand coefficient, shocks do not enter in a linear way. Hence, we cannot run simple autocorrelation tests using usage observations. To isolate the usage shocks, one would need to take logs of the quantity $\left(q_{ijt} + bp_j\right)$, which is not feasible since b is one of the parameters to be estimated. To overcome this issue, we consider sub-samples of customers for which p_j does not vary, reducing the term pb_j to a constant, and then estimate the degree of autocorrelation in each sub-sample. We do so by successively limiting the sample to customers who are on the same tariff and do not switch to a different tariff. Then we run a fixed effect linear regression for the whole history of each set of customers, using $\log\left(q_{ijt}\right)$ as dependent variable and its lagged value as independent variable.[2] For each of the subset of customers, we find no evidence of strong autocorrelation among the usage shocks (ρ ranges from 0.16 to 0.35 across all tariffs).

We then perform further analyses to ensure that the weak serial correlation we find does not bias our model estimates. We first simulate tariff choice and usage behavior for a synthetic panel of customers where we use the estimated parameters from our main model as the data generating process. We incorporate weak autocorrelation (values of 0.2, 0.3 and 0.4) into the usage process through autocorrelated usage shocks. We estimate all parameters using our main model. We find that in all cases the simulated values lay within the posterior interval of the estimated parameters. This provides further confirmation that our results are not affected by a possible weak autocorrelation.

Second, we investigate whether past usage shocks affect switching behavior. We estimate a logistic regression with 'switching to a three-part tariff' as dependent variable.[3] As independent variables, we use past usage, dummy variables for the current two-part tariff, and the ratio of usage in the last period to usage in the period before last. The latter variable serves as a proxy for the usage shock received in the previous period. If past usage shocks affected switching to three-part tariffs, then the "shock" variable should be significant. We find that this is not the case.

Table A7 summarizes the results of three different specifications. In the first specification, we include the usage shock in the last period as a predictor for switching behavior, controlling for

---

[2] We use the method proposed by Blundell and Bond (1998) to correct for the Nickell bias induced by the fixed effect.

the chosen tariff. In the second specification, we also control for the average usage level previous to the three-part tariff introduction, and in the third specification, we add a quadratic term for average usage.[4] In all specifications, the proxy for a past usage shock is not significant. We therefore conclude that autocorrelation does not explain the over-usage we observe in the data.

Table A7: Logistic regression results for switching to three-part tariffs

Variable	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Constant	-5.020	0.000	-5.496	0.000	-5.582	0.000
Previous usage (avg.)			0.001	0.008	0.001	0.130
Previous usage (avg.) ^2					0.000	0.645
Past usage shock	0.034	0.206	0.034	0.226	0.035	0.233
Dummy for previous tariff 2_1	0.635	0.160	0.966	0.045	1.000	0.039
Dummy for previous tariff 2_2	-1.037	0.186	-0.763	0.338	-0.746	0.349
Dummy for previous tariff 2_3	0.755	0.041	0.978	0.012	0.988	0.011

MODEL

Asymptotic properties of the learning model

We show that for any value of the initial parameters (α_0, β_0), the expected value of the belief $\tilde{\beta}_i$ converges to the true value, β_i, and its variance goes to zero as the consumer gets more experience on a three-part tariff (i.e., the number of periods on a three part tariff goes to infinity). We compute the limit of the mean and the variance of the beliefs, as shown in equation (22), when n goes to infinity:

[3] We estimate tariff choice in the fourth month of data. As a robustness check we also estimate the same model using months 5, 6, etc. and in all cases, obtain qualitatively the same results.

[4] We perform the same analysis using (1) current usage, and (2) lagged usage. We obtain the same qualitative results.

$$\lim_{n\to\infty} E(\tilde{\beta}_{i\tau_n}) = \lim_{n\to\infty} \frac{\alpha_0 + nr}{\beta_0 + \sum_{t=1}^{n} s_{i\tau_t}}$$

(A1)

$$= \lim_{n\to\infty} \frac{\frac{\alpha_0}{n} + r}{\frac{\beta_0}{n} + \frac{1}{n}\sum_{t=1}^{n} s_{i\tau_t}}.$$

We know from equation (19), that $s_{i\tau_t}$ is gamma-distributed with shape and scale parameters $\left(r, r/e^{\delta_i}\right)$. Thus, as $n \to \infty$, we know that $\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} s_{i\tau_t} = e^{\delta_i}$.

Therefore, substituting this result into (A1), we obtain that $\lim_{n\to\infty} E(\tilde{\beta}_{i\tau_n}) = \frac{r}{e^{\delta_i}}$.

$$\lim_{n\to\infty} Var(\tilde{\beta}_{i\tau_n}) = \lim_{n\to\infty} \frac{\alpha_0 + nr}{\left(\beta_0 + \sum_{t=1}^{n} s_{i\tau_t}\right)^2}$$

(A2)

$$= \lim_{n\to\infty} \frac{\frac{\alpha_0}{n} + r}{n\left(\frac{\beta_0}{n} + \frac{1}{n}\sum_{t=1}^{n} s_{i\tau_t}\right)^2}$$

$$= 0.$$

Posterior distributions for the full model

The model is estimated using a Bayesian framework. We obtain estimates of all model parameters by drawing from the marginal posterior distributions. Given the nonlinearities of our likelihood function and the complexity of the hierarchy in the parameters, most conditional distributions do not have conjugate posteriors. We use the Metropolis-Hasting (MH) algorithm to draw from these conditional posterior distributions. We use a data augmentation approach to include the unobserved individual-level parameters as well as the time-variant beliefs.

We denote Ω as all parameters in our model, including the population parameters $\Phi = \{b, \rho_1, \rho_2, \beta_0, a_1, a_2, \exp(r)\}$, the individual-level parameters $\varpi_i = \{\eta_i, \delta_i, \lambda_i\}$, the mixing parameters $\alpha = \{\mu_\eta, \sigma_\eta, \mu_\delta, \sigma_\delta, \mu_\lambda, \sigma_\lambda\}$, and the individual specific time-variant beliefs $\tilde{\beta}_{it}$. The full joint posterior distribution can be written as:

$$
\begin{aligned}
f(\Omega \mid Data) &\propto L(Data \mid \Omega) f(\Omega) \\
&= \left\{ \prod_{i=1}^{I} \left\{ \prod_{t=1}^{T_i} f(q_{it} \mid k_{it}, \Phi, \eta_i, \delta_i, Z_{it}, X_j) \right.\right. \\
&\qquad \times f(k_{it} \mid \tilde{\beta}_{it}, \Phi, \eta_i, \delta_i, \lambda_i, Z_{it}, X_j) \\
&\qquad \left. \times f(\tilde{\beta}_{it} \mid \Phi, \eta_i, \beta_0, Z_{it}) \right\} \\
&\qquad \left. \times f(\eta_i \mid \mu_\eta, \sigma_\eta) f(\delta_i \mid \mu_\delta, \sigma_\delta) f(\lambda_i \mid \mu_\lambda, \sigma_\lambda) \right\} \\
&\qquad \times f(\mu_\eta) f(\sigma_\eta) f(\mu_\delta) f(\sigma_\delta) f(\mu_\lambda) f(\sigma_\lambda) f(\Phi).
\end{aligned}
$$

where $f(q_{it} \mid k_{it}, \Phi, \eta_i, \delta_i, Z_{it}, X_j)$, $f(k_{it} \mid \tilde{\beta}_{it}, \Phi, \eta_i, \delta_i, \lambda_i, Z_{it}, X_j)$, and $f(\tilde{\beta}_{it} \mid \Phi, \eta_i, Z_{it})$ are the expressions derived in the appendix, (App-1), (App-2), and equation (21) in the main paper. Expressions $f(\eta_i \mid \mu_\eta, \sigma_\eta)$, $f(\delta_i \mid \mu_\delta, \sigma_\delta)$, and $f(\lambda_i \mid \mu_\lambda, \sigma_\lambda)$ correspond to the mixing distribution for the population parameters, as specified in the Model section. We choose diffuse prior distributions for all population parameters. We use a normal distribution with mean and standard deviation (0,100) for $\mu_\eta, \mu_\delta, \mu_\lambda$, and inverse-gamma with shape and scale parameters $(1, \sqrt{10})$ for $\sigma_\eta, \sigma_\delta, \sigma_\lambda$. We assume that $\Phi = \{b, \rho_1, \rho_2, \beta_0, a_1, a_2, \exp(r)\}$ follows a multivariate normal distribution with parameters $\mu_\Phi = [\varnothing_{n\Phi-1}, 3]$ and $\text{diag}(\Sigma_\Phi) = [100 \times \mathrm{I}_{n\Phi-1}, 1]$, where $n\Phi$ is the dimension of Φ, $\varnothing_{n\Phi-1}$ is a $1 \times n\Phi$ vector of zeros, and I_n is the identity matrix of dimensions

$n \times n$. (The values of μ_Φ and Σ_Φ were chosen to ensure uninformative priors in the transformed space.) We draw recursively from the following posterior distributions:

1. (Gibbs) Parameters $\mu_\eta, \sigma_\eta, \mu_\delta, \sigma_\delta, \mu_\lambda, \sigma_\lambda$ are obtained by sampling from the following distributions:

   - $f(\mu_\eta \mid \sigma_\eta^2, \eta_i) = \text{Normal}\left(\frac{\sum_{i=1}^{I} \eta_i}{I}, \left(\frac{1}{10000} + \frac{I}{\sigma_\eta^2} \right)^{-2} \right).$

   - $f(\sigma_\eta \mid \mu_\eta, \eta_i) = \text{Inverse Gamma}\left(1 + \frac{I}{2}, \sqrt{10} + \frac{\sum_{i=1}^{I} (\eta_i - \mu_\eta)^2}{2} \right).$

   We proceed similarly for parameters $\mu_\delta, \sigma_\delta, \mu_\lambda, \sigma_\lambda$.

2. (MH) Draws for Φ are obtained by sampling from

   - $f(\Phi \mid \mu_\Phi, \Sigma_\Phi, \varpi_i, \beta_{it}, \text{data}) \propto \exp\left(.5 (\Phi - \mu_\Phi)' \Sigma_\Phi^{-1} (\Phi - \mu_\Phi) \right) P(\text{data} \mid \varpi_i, \beta_{it}, \Phi)$

3. (MH) Draws for η_i are obtained by sampling from:

   - $f(\eta_i \mid \mu_\eta, \sigma_\eta, \Phi, \delta_i, \gamma_i, \beta_{it}, \text{data}) \propto \exp\left(.5 \frac{(\eta_i - \mu_\eta)^2}{\sigma_\eta^2} \right) P(\text{data} \mid \Phi, \varpi_i, \beta_{it})$

   - We proceed similarly for λ_i.

4. (MH) Draws for β_{it} are obtained by sampling from:

   - $f(\beta_{it} \mid \Phi, \delta_i, \gamma_i, \beta_{it}, \text{data}) \propto g\left(\beta_{it} \mid r\beta_0 + nr, \beta_0 + \sum_{t=1}^{n} s_{i\tau_t} \right) P(\text{data} \mid \Phi, \delta_i, \gamma_i, \beta_{it})$,

   where $g\left(\beta_{it} \mid r\beta_0 + nr, \beta_0 + \sum_{t=1}^{n} s_{i\tau_t} \right)$ is the gamma pdf as derived in (21).

Since there is no closed-form expression for the posterior distributions of Φ and ϖ_i, we use a Gaussian random-walk Metropolis-Hasting algorithm to draw from these distributions. Following the Metropolis-Hasting procedure proposed by Atchade (2006), for each iteration, *s*, we draw a proposal vector of parameters $\zeta^{(s)}$ (either for α and ϖ_i):

$$\zeta^{(l)} \sim \text{Normal}\left(\zeta^{(l-1)}, \sigma^{(l-1)}, \Delta^{(l-1)}\right)$$

and then accept the vector using the Metropolis-Hastings acceptance ratio. The tuning parameters $\sigma^{(l-1)}$ and $\Delta^{(l-1)}$ are adapted in each iteration to get an acceptance rate of approximately 20%. We ran the simulation for 30,000 iterations. The first 20,000 iterations were used as a ``burn-in'' period, and the last 10,000 iterations were used to estimate the conditional posterior distributions. Figure A9 and Figure A10 show the posterior draws obtained in the simulation.

Figure A9: Posterior draws for the population parameters (MH steps)

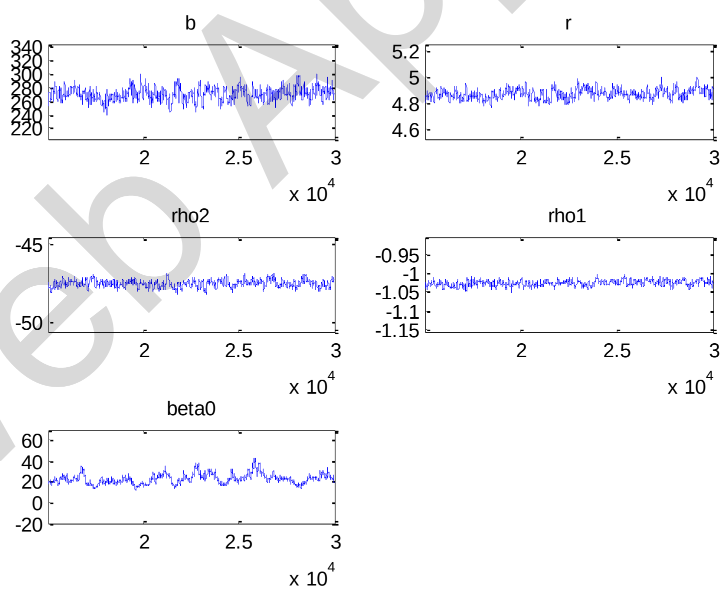
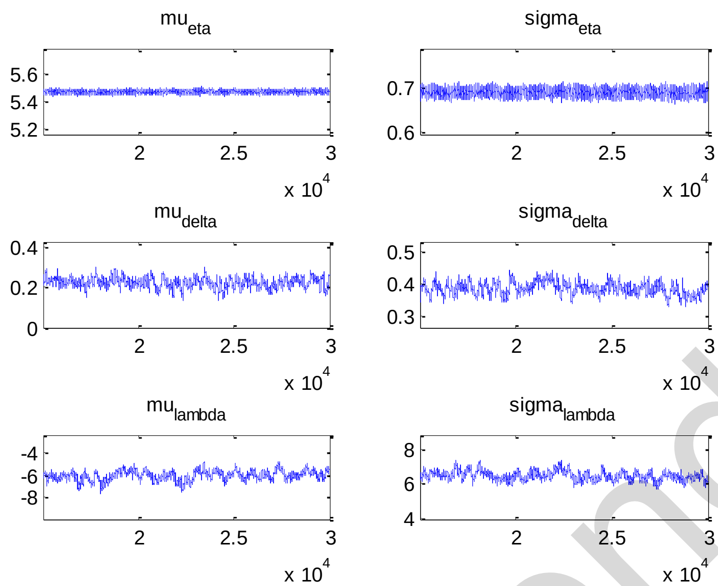


Figure A10: Posterior draws for mixing individual-level parameters (Gibbs)



FURTHER ROBUSTNESS CHECKS

Sensitivity analysis for the effect of switching costs on counterfactual analyses

Our econometric model assumes that customers' choice decisions are based on the next period only. This assumption does not affect the estimates of our main variable of interest, δ_i, but could potentially lead us to overestimate consumers' sensitivity to the switching fee, ρ_1. If this were the case, the effect of lowering the switching fee on provider revenues could be lower than what our results about recommendations to the firm suggest. We run a sensitivity analysis to measure whether the effect of reducing the switching fee, as presented in in the main manuscrip, is robust to lower levels of ρ_1. We reduce the estimate of ρ_1 by 5%, 10%, and 20%.

Figure A11, Figure A12 and Figure A13 replicate the results obtained in the main manuscript (see Figure 3 of the main manuscript) for lower levels of ρ_1. We find that the revenue impact from lowering the switching fee is very robust to lower levels of ρ_1. In an additional analysis, we similarly vary the level of the sensitivity to cost of switching to a different provider,

ρ_2. We find that while a lower sensitivity to cost of leaving the provider affects the level of provider revenues, it does not change the optimal level of the switching fee. Hence, we are confident that the assumption that customers make tariff choice decisions taking into account their usage in the next period only does not significantly bias our policy simulations.

Figure A11: Change in revenue due to reduction of the switching fee if ρ_1 is reduced by 5%

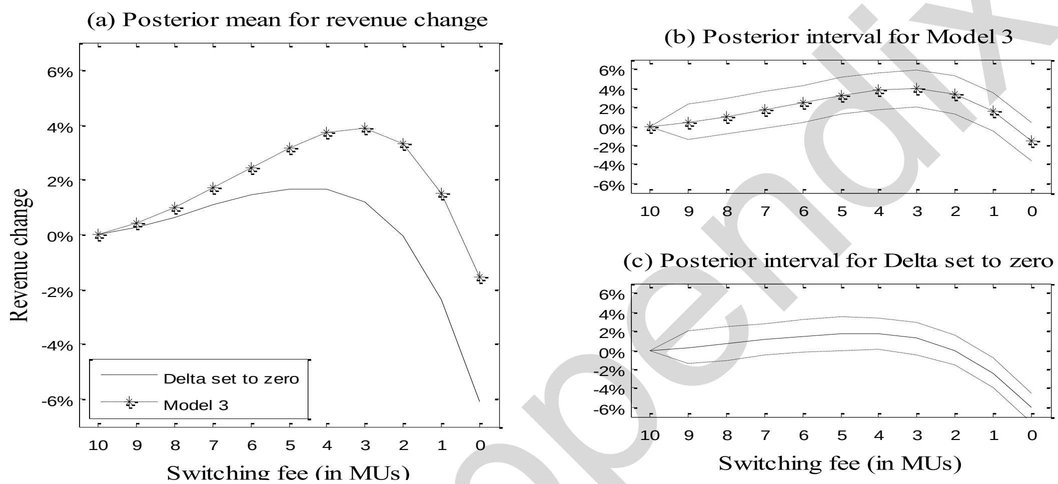


Figure A12: Change in revenue due to reduction of the switching fee if ρ_1 is reduced by 10%

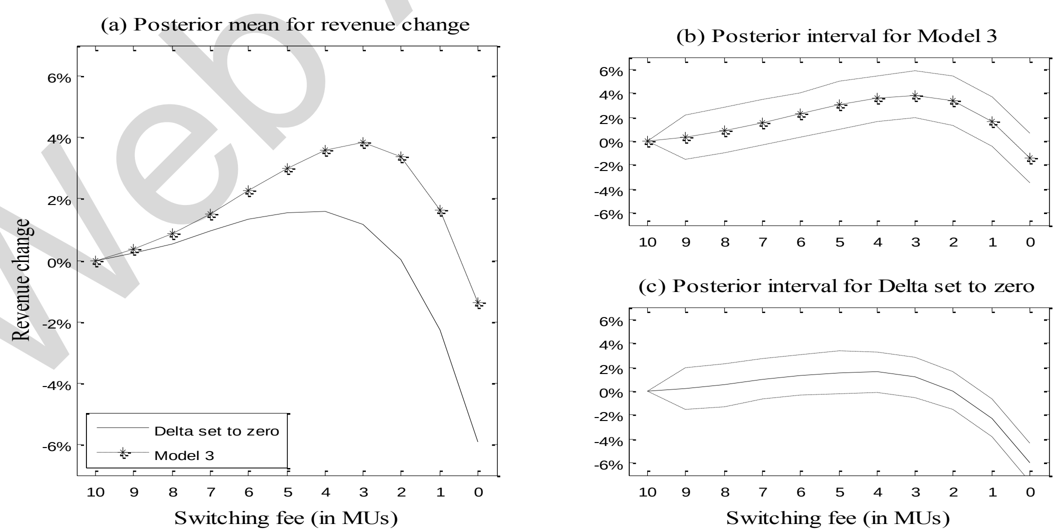
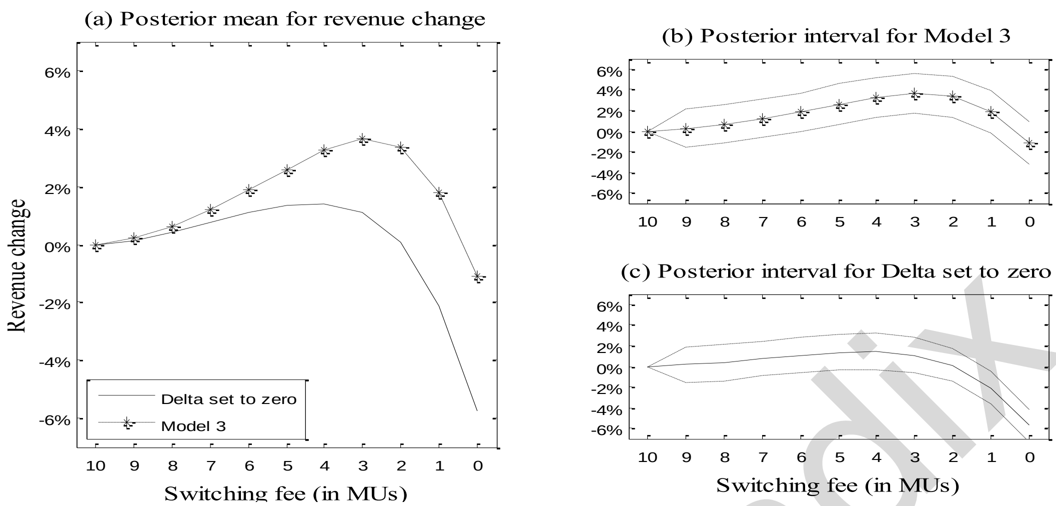


Figure A13: Change in revenue due to reduction of the switching fee if ρ_1 is reduced by 20%



Alternative model specification: Additional value for "free" minutes only

An alternative way to build our model would be to assume that three-part tariff customers assign greater value only to minutes strictly below the allowance, and not to all three-part tariff minutes. In our data, three-part tariff usage mostly lies beyond the allowance: 72% of three-part tariff observations exceed the usage allowance, by an average of 88.4%. As a consequence, a behavioral theory that limits the effect of free minutes to usage below the allowance seems, in principle, unable to explain the pattern in our data.

To further confirm this claim, we re-estimate Model 2 as presented in the main manuscript but allow the effect of free minutes, δ_i, to apply to minutes within the allowance only. We find that such a model does not reflect the phenomenon we observe well. First, the fit is worse than that of Models 2 and 3 (Model section of the main manuscript) that assume that the additional valuation δ_i applies to all three-part tariff minutes. The MSE of the alternative model is 55.08 versus a MSE of 46.33 in Model 2 and 45.54 in Model 3. In the alternative model we obtain a MAPE of 75.74 versus a MAPE of 72.4 in Model 2 and of 71.87 in Model 3. Second, we obtain a negative posterior mean of the variable relating to the value of free minutes. This estimate is negative because in our sample customers generally consume above the allowance. As a consequence, a model that only estimates δ_i from minutes within the allowance would overestimate the satiation level for customers who switch to a three-part tariff and often consume above the allowance (i.e., the majority of our three-part tariff customers). Then, in the periods in which these customers consume within the allowance, δ_i needs to be negative to compensate for the overestimation of their satiation level. A negative delta cannot explain the usage increase observed in the data and documented in the main manuscript, and it is not consistent with previous literature indicating that "free" would lead to increased valuation of the good. We conclude that this model specification is not a good representation of the phenomenon we observe.

Table A8: Posterior distribution of parameter estimates for model where δ_i applies to free minutes within the allowance only

	Model 2 "free" minutes only		
	Mean	95% Interval	
Demand intercept			
Mean, μ_η	5.520	5.501	5.538
Std. dev., σ_η	0.690	0.679	0.701
Demand slope, *b*	240.356	230.992	250.097
Variance of usage shock, *1/r*	0.216	0.214	0.218
Valuation of free units			
Mean, μ_δ	-0.301	-0.364	-0.234
Std. dev., σ_δ	0.478	0.433	0.536
Preferences in tariff choice, ζ_{ijt}			
SC bw. tariffs, ρ_1	-1.025	-1.035	-1.014
SC to other provider, ρ_2	-47.792	-48.164	-47.425
Preference for the three-part tariff			
Mean, μ_λ	-4.749	-5.226	-4.252
Std. dev., σ_λ	5.556	5.250	5.858
Log Marginal Density	-394749		-
MSE ('000)	55.08		-
MAPE	75.74		-

N=5,831 customers, 63,449 usage and 63,616 choice observations
Demographic shifters of the demand slope included but not reported for readability.

REFERENCES

Atchade, Yves F. (2006), "An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift," Methodology and Computing in Applied Probability, 8, 235-254.

Blundell, Richard, and Stephen Bond. (1998). "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," Journal of Econometrics 87, 115-143.

Roodman David (2009) "How to do xtabond2: An introduction to difference and system GMM in Stata," Stata J.,9, 86-136