

Determining Consumers' Discount Rates with Field Studies

Song Yao Carl F. Mela Jeongwen Chiang Yuxin Chen¹

June 15, 2011

¹Song Yao is an Assistant Professor of Marketing at the Kellogg School of Management, Northwestern University (email: s-yao@kellogg.northwestern.edu). Carl F. Mela is the T. Austin Finch Foundation Professor of Business Administration at the Fuqua School of Business, Duke University (email: mela@duke.edu). Jeongwen Chiang is a Professor of Marketing at China Europe International Business School (email: jwchiang@ceibs.edu). Yuxin Chen is the Polk Brothers Professor in Retailing and Professor of Marketing at the Kellogg School of Management, Northwestern University (email: yuxin-chen@kellogg.northwestern.edu). The authors would like to thank seminar participants at Northwestern University, 2010 China India Consumers Insights Conference, 2011 UTD Forms Marketing Conference, 2011 Marketing Science Conference, as well as Eric Anderson, Praveen Kopalle, Harikesh Nair, Aviv Nevo and Florian Zettelmeyer for their feedback and Hong Ke and Jason Roos for their invaluable research assistance. The authors are listed in reverse alphabetical order.

Abstract: Determining Consumers' Discount Rates With Field Studies

Because utility/profits, state transitions and discount rates are confounded in dynamic models, discount rates are typically fixed to estimate the other two factors. Yet these rate choices, if mis-specified, generate poor forecasts and policy prescriptions.

Using a field study wherein cellphone users transitioned from a linear to three-part-tariff pricing plan, we estimate a dynamic structural model of minute usage and obtain discount factors that would normally be unidentifiable. The identification rests upon imputing the utility under linear pricing plans without dynamic structure; then using these utilities to identify discount rates when consumers were switched to a three-part tariff where dynamics became material.

We find that the estimated segment-level weekly discount factors (0.86 and 0.91) are much lower than the value typically assumed in empirical research (0.995). When using a standard 0.995 discount factor, we find the price coefficients are underestimated by 23%. Moreover, the predicted intertemporal substitution pattern and demand elasticities are biased, leading to a 27% deterioration in model fit; and suboptimal pricing recommendations that would lower potential revenue gains by 74-88%.

Keywords: dynamic structural model, identification, forward-looking consumers, heterogeneous discount rates, nonlinear pricing.

1 Introduction

Individuals often face situations where they must choose between engaging in consumption in the present or waiting to consume at a future time. A rich stream of recent literature has adopted dynamic structural models to study intertemporal consumption, yielding deep insights into consumer behavior, such as Rust (1987), Erdem and Keane (1996), Hendel and Nevo (2006), and Sun (2005).

Albeit the increasingly ubiquitous use of structural models to study dynamic consumption behavior due to their appealing theoretical foundation, the identification of these models are problematic (Rust, 1994). In particular, to identify consumer utility functions, it is often necessary to i) assume or fix the discount factor at a given value (normally between 0.90 and 0.9999), and ii) assume that this rate is common across individuals. For example, an annual discount factor of 0.95 is often justified via the argument that this value is consistent with an annual interest rate of about 5%. While this rate might be suitable for analyzing the decisions of firms who are subject to capital constraints in the financial market, it is not clear whether this interest rate applies across consumption contexts where individuals have different degrees of access to capital and credit rates. Moreover, it would be desirable to relax the assumption of homogeneous discount rates. Those with low discount rates are more likely to defer consumption. As a result, targeting policies predicated on low discount rates, such as low introductory rates or trial promotions, might be misplaced if the future is not so material to less forward looking consumers.

Because the use of dynamic structural models in marketing is becoming more common and because the assumed discount rate affects inferences regarding agent behavior, optimal policy and forecast outcomes, we develop a dynamic structural model in order to identify and measure heterogeneous consumer discount rates using field data. More specifically, we estimate a dynamic structural model using customer cellphone minute usage data during a field study that involves switching pricing plans to consumers. In our data, customers were initially under a linear “pay-per-minute” plan. Later the cellphone service provider

implemented a field experiment and these customers were switched to a nonlinear three-part tariff plan.¹ The switch induced intertemporal substitution trade-offs in the later period of the data as a consumer's decision of minute usage early in the month had consequences for the rates she faced later in the month. In comparison, under the preceding linear pricing schedules, the consumer paid a constant marginal price for usage. Consequently, there was no intertemporal substitution of minute usage and customers made consumption decision statically. Relying on this field experiment, we first obtain customers' utilities of phone usage from the data of the linear plan. Conditional on these identified utilities, we then estimate heterogeneous discount factors of customers with the three-part tariff data.

As an ancillary benefit of the field study, we are able to explore the potential for hyperbolic discounting. Although there are studies show the existence of hyperbolic discounting (see the survey by Angeletos et al. (2001)), we find that no strong evidence of hyperbolic discounting in the focal context of monthly cellphone usage. Our result is consistent with the studies by Chevalier and Goolsbee (2005) and Dubé et al. (2010b), which also do not find strong support for hyperbolic discounting.

Our results indicate that customers demonstrate considerable heterogeneity in their preferences and discounting patterns. In particular, the customers have very high discount rates. The customers only has a weekly discount factor of around 0.86 to 0.91, much lower than the 0.995 weekly discount factor commonly assumed in the empirical dynamic model literature (e.g., Erdem and Keane (1996) and Sun (2005)). This traditional discount rate implies an equally priced minute at the beginning of the month to be worth the same as 1.02 minutes at the end of the month; instead we compute that customers value a minute now more closely to 1.5 to 1.9 minutes at the end of the month (depending on the consumer segment). Furthermore, setting the discount factor to 0.995 leads to an underestimation of price coefficients

¹A three-part tariff plan contains three components: an access fee, a certain amount of allowance minutes, and a marginal price if the customer's usage exceeds the allowance within a billing cycle. As a result, the customer may be subject to extra fees when the consumption exceed the allowance (overage) and overpays if the consumption falls below the allowance (underage). Due to the existence of the allowance and the high marginal price, a customer needs to decide how to allocate her consumption across time within a billing cycle, intending to avoid overage and underage so as to maximize her total utility.

of up to 23%. The intuition behind this result is that setting the discount rate too high implies customers would excessively tradeoff current within-allowance minute consumption with their future consumption, where they may have to pay the marginal price. Given that they do not actually do this in the data, it results in the model making an effort to lower cross-period substitution in demand; this is achieved by estimating a lower price sensitivity. Correspondingly, the estimated demand elasticities and predicted intertemporal usage patterns under the 0.995 discount factor are also biased, leading to worse fit. In the case of usage patterns, the traditional discount rate assumes individuals are more patient than they really are, understating the tendency to consume minutes early.

The model also enables us to investigate the impacts of the firm’s pricing strategy on its revenue and the attendant implications of discount rates for pricing strategy. We conclude that roughly 10% of the customers in our context have a greater baseline consumption rate and lower price sensitivity than the remaining customers, making them good candidates for customized pricing. Accordingly, we calculate the firm’s revenue and customers welfare under alternative pricing strategies. We find that the firm’s revenue under the alternative pricing plan may potentially increase by nearly 2% on those targeted customers, with little impact on their welfare as measured by the utilities of phone usage. In contrast, alternative strategies developed from a model with commonly used discount rates lead to sub-optimal outcomes that lower potential revenue gains by 36-78%, depending on the market segment considered.

The remainder of the paper is organized as follows. In section 2, we overview the relevant literature to differentiate our paper from past research. Next, to better illustrate our model and identification strategy, we introduce the unique aspects of our data. We then detail the model and estimation. Subsequently, we present and discuss the results and corresponding managerial implications. We conclude with some future research directions.

2 Past Literature

Given that discount rates are not typically identified, several approaches have emerged to contend with the problem, including i) assuming a fixed value for the discount rate, ii) functional identification via structural assumptions and/or estimation via exclusion restrictions, and iii) experimental approaches. Table 1 overviews a sample of these approaches and their resulting discount values converted to their weekly equivalents. Table 1 makes it apparent that discount rates vary considerably across studies. The mean weekly discount factor is 0.979 with a standard deviation of 0.034. The corresponding weekly discount rates average 2.25% with a large standard deviation of 3.76%. In short, there is no clear consensus regarding the value of discount factors, partially due to the fact that discount rates are typically not identified. This large variation is problematic in practice because, as we shall show, the optimal policy for the firm or consumer can vary substantially with the imputed or articulated discount rate.

[Insert Table 1 about here.]

First, most studies assume or fix the discount factors to certain values, typically between 0.995 to 1.0. For the purpose of identification, it is also a common practice to assume the discount factor is the same across individuals which might be, in some instances, a strong assumption (Frederick et al., 2002).

A second approach to the identification of discount rates includes the imposition of structure on the model such as assuming the distribution of the model errors, individuals knowing the state transition probability, and no unobserved heterogeneity (e.g., Hotz and Miller, 1993).² However, these structures for the purpose of identification may be difficult to sub-

²Besides dynamic structural models, it should be noted that there are studies estimating implied discount factors, using data of consumers trading off immediate return/cost and future flow of return/cost, e.g., current capital cost of a more energy efficient durable product and future operating cost of that product. The literature include Hausman (1979), Dubin and McFadden (1984), Warner and Pleeter (2001), Harrison et al. (2002), Allcott and Wozny (2011) and others. For a more thorough discussion, see Frederick et al. (2002). However, to identify the discount factor, it is still necessary to impose some structures, which are sometimes difficult to verify. For example, to quantify the future operating cost of a durable appliance, it is

stantiate in some contexts (e.g., homogeneous consumers, full information of state transition probability for new products or markets, etc.).

A third identification strategy is to rely on the exclusion restriction condition (Magnac and Thesmar, 2002). Exclusion restrictions involve specifying a set of exogenous variables that do not affect current utility but do affect state transitions. Accordingly, variation in these exogeneous variables affects future utilities through their impact on the state transition but do not have an effect on current utility. By exploring how choices are made in light of changes in future utility when current utility remains fixed, the utility and the discount factor can be simultaneously identified. Some studies implement exclusion restriction to identify not only the exponential discount factor but also the hyperbolic discount factor (e.g., Chung et al. 2010, Fang and Wang 2010). However, such exclusion restrictions are often unavailable in field data or are difficult to validate. Furthermore, though discount factors may vary across individuals (Frederick et al., 2002), it is not clear whether the identification of heterogeneous discount factors under the exclusion restriction is feasible.

To alleviate these concerns, recent work by Dubé et al. (2010b) use experimental conjoint analysis data to identify dynamic model in the context of durable goods adoption. In particular, Dubé et al. (2010b) manipulate consumers' beliefs about state transitions by informing them alternative future market situations in the experiments. As a result, they are able to identify utility and discount factors. This approach is most similar to ours in that it uses data rather than invoking assumptions to infer discount rates. Though an important step forward, it is often difficult to replicate dynamic choices in lab settings owing to demand artifacts and contracted durations; it would therefore be desirable to supplement this research using a field context with choices made in practice and over extended periods.³ Moreover, field data enables one to explore the potential revenue and utility consequences

necessary to assume the consumer has perfect foresight about future prices of electricity. As a result, there is no uncertainty about the future state transtion.

³For example, Dubé et al. (2010b) consider annual budget tradeoffs in a lab experiment that lasts one session.

of mis-specifying discount rates and also conduct appropriate policy simulations involving changes in marketing strategy in the context of dynamic choice.

Therefore we advance the research in dynamic structural models by identifying heterogeneous discount factors using field experiment data. Our identification strategy is to first identify consumers' heterogeneous utilities and the distribution of random consumption shocks using data that have no dynamics involved. Then we further recover their discount factors when the dynamic structure was exogeneously imposed.

Our contributions are fourfold. First, we identify and measure discount rates using field data. This is useful because the estimates are informative for setting discount rates in cases where exogenous variation in temporal decision making does not exist. Second, we consider the role of heterogeneity and the potential for hyperbolic discounting in the context of a field setting. Third, we explore the potential for biased parameter estimates as a result of mis-specifying discount rates as well as the potentially suboptimal marketing decision making. Fourth, our research also advances the empirical literature on the non-linear pricing of telephony or Internet services (e.g., Narayanan et al. 2007, Lambrecht 2006, Lambrecht et al. 2007, Iyengar et al. 2007). Most previous empirical studies are based on aggregate usage data, limiting their ability to investigate customers' intertemporal substitution in consumptions. Since our data are at the disaggregate level, we are able to evaluate the tradeoff of consumptions across time and the corresponding managerial implications for the firm's pricing strategy.

3 Data

3.1 Consumer Usage Data and Carrier Tariff Structure

The data for our analysis are supplied by a major mobile phone service provider in China, covering the period from September 2004 to January 2005. The data provider accounted for more than 70% of the market share of Chinese mobile phone service market during that

time. Initially, this firm used only linear pricing schedules, i.e., customers were billed on a pay-per-minute basis. In November 2004, on an experimental basis, the firm offered three-part tariff plans to a randomly selected set of customers. The firm divided these customers into multiple groups based on their past usage volumes. The firm then offered each group a respective three-part tariff plan. A customer could choose the three-part tariff plan or remain on the original pay-per-minute plan.

3.1.1 Tariff Structure

Table 2 depicts the pricing structure of the most popular three-part tariff plans, covering 90% of the customer base. Customer who enroll in one of the listed plans are allowed a fixed number of free calling minutes in a given calendar month by paying the monthly access fee.

[Insert Table 2 about here.]

When a given customer places or receives a call, the minutes of the phone call are deducted from the allowance and the customer does not need to pay for that usage. However, when the monthly allowance is exhausted, the customer is billed the marginal price for each minute of usage beyond the allowance. There is no “roll-over” for these plans, i.e., unused allowance minutes can not be carried over to next month. At the beginning of next month, the customer’s allowance of free minutes is replenished after paying the new month’s access fee. The customers have different linear rates before the switch. The mean linear rate is 0.27 with a standard deviation of 0.09.

3.1.2 Usage Data

For the first four months (from September 2004 to December 2004), we observe each customer’s aggregate monthly minute usage and expenditures. However, in the last month (January 2005), we observe call level customer records, including the starting time, duration, and expense of each phone call. The data also include some demographic information, including the age, gender, and zip code of each customer.

Table 3 summarizes the customers’ average usage levels (normalized by their allowance level) and demographic information. The average usage under both the linear and three-part tariff plans are close. However usage variation is considerable, suggesting heterogeneity in usage behavior is material.

[Insert Table 3 about here.]

3.2 Overage and Underage

Underage occurs when customers do not use all the allowance at the end of a month. In this case, customers are overpaying in the sense that they have been charged for minutes they do not use. In comparison, overage occurs when usage exceeds the allowance. In this case, customers again overpay inasmuch as a plan with more allowance minutes normally has a lower average price per allowance minute (Iyengar et al. (2007)). As a result, customers who strategically manage the minutes should evidence less underage or overage.

In Figure 1, we plot the histogram of the ratios of minutes used to minutes allowed for the last month of data. The average ratio is close to 1 (0.96) under the three-part tariff, suggesting that customers on average tend to avoid overage or underage. Yet this average behavior belies a large standard deviation (0.35). Hence, we next consider whether and how users manage their minutes over the month to comport with the allowance; to the extent this behavior changes as the allowance becomes more salient, evidence is afforded for the strategic use of minutes.

[Insert Figure 1 about here.]

3.3 Strategic Minute Usage within a Month

We consider some model free evidence that customers are strategic in their usage of allowance minutes. This evidence is predicated on the notion that minute consumption changes as the distance between minutes used and the allowance becomes small; in particular, consumers

start to conserve minutes as the number of free minutes dwindles and the overage potential increases.

Dividing the last month of the data into five 6-day periods, $t = 1, \dots, 5$,⁴ we compute the ratio of cumulative minutes used to the allowance for each customer at the end of each six-day period. Figure 2 portrays a scatter plot of this ratio and its lag value for each period t , $t = 2, \dots, 5$. The line in this figure depicts a nonparametric function relating the ratio and its lag and the gray band indicates the 95% confidence interval for this function.⁵ A key insight from this figure is that this function is concave when the cumulative usage is within quota (the ratio in period $t - 1$ is less than 1). In contrast, when usage exceeds quota (the ratio in period $t - 1$ is greater than 1), the line is almost linear. The concavity of the line pre-quota suggests that people decelerate usage as they approach the quota, that is, they start to ration their minutes to avoid overage. Moreover, those who are far from the quota appear to accelerate usage to avoid underage. In comparison, customers who have already exceeded the quota do not decelerate (or accelerate) their usage. Instead, they follow some relatively stable usage rates, as might be expected were they to no longer face an intertemporal tradeoff in usage. Misra and Nair (2009) and Chung et al. (2010) use similar methods to investigate the effect of quota on salesperson’s allocation of efforts across time. They found analogous patterns of dynamic effort allocation in salesforce due to the existence of quota.

[Insert Figure 2 about here.]

To further elaborate upon these insights arising from Figure 2, we consider how usage acceleration (deceleration) changes as individuals approach their allowance/quota. This acceleration can be summarized by the statistic (Usage during period t)/(Usage during period $t - 1$). This ratio is analogous to the slope of the line in Figure 2. When the ratio is one, consumers are neither decelerating or accelerating use. When the ratio is greater than one, usage is accelerating. When the ratio is less than one, usage is decelerating. We compute

⁴We use t to index periods within a month and τ to index months.

⁵We consider both spline and local regression methods. The results are similar. The figure presented shows the results from spline method.

this ratio for each person in each period and then, in Figure 3, present a histogram of this minute acceleration measure across persons and periods conditioned on users distance to quota. For example, the upper leftmost histogram shows the distribution of usage acceleration observations conditioned upon customer usage at time $t - 1$ being less than 20% of their allowance. Figure 3 indicates that customers usage decelerates as they approach their allowance. Further, when the customer reaches an overage situation (where there is no longer an intertemporal tradeoff in usage), the slopes average around 1, indicating a stable usage rate. These observations are consistent with Figure 2. Overall, we conclude that there exists some model free evidence of strategic behavior on the part of consumers.

[Insert Figure 3 about here.]

4 Model

4.1 Utility under the Linear Pricing Plan

In this section, we first specify the consumer utility for consumption under a linear pricing plan and derive the optimal level of consumption. We then extend the analysis to the case of the three-part tariff plan.

Similar to Lambrecht et al. (2007), we begin by assuming that customer i , in market segment g , derives utility from phone usages and the consumption of a composite outside product (numeraire). To be specific,

$$u_{it}(x_{it}, z_{it}) = \left(\frac{d_{it}x_{it}}{b_g} - \frac{x_{it}^2}{2b_g} \right) + z_{it}, \quad (1)$$

$$s.t. \quad z_{it} = y_i - p_{i0}x_{it}, \quad (2)$$

$$d_{it}, b_g > 0$$

where $t = 1, 2, \dots, T$ are the periods within a month; x_{it} is the minutes of phone usage during period t ; p_{i0} is the linear price rate of customer i before switching to the three-part tariff; z_{it} is

the consumption of the numeraire; y_i is the income; d_{it}/b_g is the main effect of minute usage; b_g is the price sensitivity; and d_{it} represents the baseline consumption level (cf. equation 5).⁶

Following Narayanan et al. (2007) and Lambrecht et al. (2007), we further allow baseline consumption, d_{it} , to be affected by time-variant consumer characteristics, D_{it} , and a random shock ν_{it} ,

$$d_{it} = \exp(D_{it}'\alpha_g) + \nu_{it} \quad (3)$$

where α_g is a vector of parameters and $\nu_{it} \sim N(0, \zeta_g^2)$ is exogeneously i.i.d. across customers and periods.⁷ Sources of the shock may include (1) technical problems with the customer's phoneset or coverage which limit the phone usage; (2) unexpected events that require extra communications with others, and so on. Though the random shocks are unknown to the researchers, the customer observes the shocks *before* deciding her usage levels accordingly. Given all customers in the dataset are experienced users, we assume the distribution of the shocks is known to the customer.⁸

Substituting the budget constraint into equation 1, the utility function can be rewritten as

$$u_{it}(x_{it}, z(x_{it})) = \frac{d_{it}x_{it}}{b_g} - \frac{x_{it}^2}{2b_g} + y_i - p_{i0}x_{it} \quad (4)$$

Customer i then chooses the optimal levels of phone usage x_{it} and numeraire consumption z_{it} so as to maximize her total utility subject to the budget constraint. Solving the maximization

⁶Note that the budget constraint normalizes the benefits of consuming one unit of the numeraire to 1. The purpose of this normalization is twofold. First, it transforms the utility up to a monetary scale, which makes any welfare interpretations more meaningful. Second, since we do not observe customer churning or variations in plan choices in the data, the identification of the benefits of consuming the numeraire is infeasible. Such a normalization treatment for the purpose of identification is similar to Narayanan et al. (2007) and Ascarza et al. (2009).

⁷The exponential function ensures that, on average $d_{it} > 0$. One related concern with the use of a normal distribution assumption for the random shocks is that the baseline demand, d_{it} , may become negative. One approach is to consider a truncated normal distribution. However this imposes a considerable computational challenge. Hence we instead assume that the magnitude of ν_{it} (standard deviation) is small compared to $\exp(D_{it}'\alpha_g)$ so a normal distribution is a good approximation of a truncated normal distribution. This assumption is confirmed to be sensible based on the estimation results.

⁸Though this assumption is not material for the static model because the error is revealed prior to the usage decision, the assumption becomes important under the context of the three-part tariff when future shocks become relevant to current period consumptions.

problem of equation 4 yields the optimal consumption

$$x_{it}^* = \begin{cases} 0, & \text{if } d_{it} - b_g p_{i0} < 0 \\ d_{it} - b_g p_{i0}, & \text{if } d_{it} - b_g p_{i0} \geq 0 \end{cases} \quad (5)$$

The foregoing equation clarifies the interpretation of (1) b_g as the price sensitivity and (2) d_{it} as the baseline consumption level under the linear pricing plan as it represents a fixed shift in the demand curve as well as the minute consumption level when $p_{i0} = 0$ (Lambrecht et al., 2007). Summing optimal period consumptions within the same month τ yields the optimal total minutes consumed within a month as

$$q_{i\tau} = \sum_{t'=1}^T x_{it'}^*$$

Finally, we presume customers are heterogeneous across segments but homogeneous within a segment.⁹ Accordingly, the conditional probability of customer i belonging to segment g is

$$f_{ig} = \exp(\lambda_{0g} + D_i' \lambda_g) / \sum_{g'} \exp(\lambda_{0g'} + D_i' \lambda_{g'}) \quad (6)$$

where D_i are time-invariant customer characteristics.

4.2 Utility under the Three-part Tariff Plan

The three-part tariff plan can be described as the triple $\{F, A, p\}$, where F is the fixed access fee, A is the allowance amount, and p is the marginal price after the customer exhausts the allowance.

⁹Although we can estimate a model with heterogeneity at the individual level, the scarcity of observations per customer renders a very noisy identification of the preferences and computational difficulties. Hence we use a latent class structure to capture preference heterogeneity in the spirit of Kamakura and Russell (1989).

4.2.1 Period Utility and Budget Constraint

At period t during a given month, customer i who belongs to segment g has a utility level

$$u_{it}(x_{it}, z(x_{it}); s_{it}, \nu_{it}) = \frac{d_{it}x_{it}}{b_g} - \frac{x_{it}^2}{2b_g} + z_{it}(x_{it}), \quad (7)$$

$$s.t. \ z_{it}(x_{it}) = y_i - C(x_{it})$$

$$C(x_{it}) = \left(\sum_{k=1}^{t-1} x_{ik} + x_{it} - A \right) p I_{\sum_{k=1}^{t-1} x_{ik} + x_{it} > A}$$

$$d_{it} = \exp(D'_{it}\alpha_g) + \nu_{it}$$

where s_{it} is a vector containing state variables at period t that include (1) cumulative usage up to period $t - 1$, $\sum_{k=1}^{t-1} x_{ik}$, and (2) period t (or the distance to the terminal period). Among these state variables, the cumulative usage is endogenous and the period t is exogenous. $I_{\sum_{k=1}^{t-1} x_{ik} + x_{it} > A}$ is an indicator, which takes the value of 1 if $\sum_{k=1}^{t-1} x_{ik} + x_{it} > A$ and 0 otherwise. Note that the fixed access fee F does not enter the period budget constraint since it is essentially a sunk cost. It does not affect the optimal decision at period t as long as the choice is not a corner solution.

Substituting the budget constraint to equation 7, we may rewrite the period utility as

$$u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) = \left(\frac{d_{it}x_{it}}{b_g} - \frac{x_{it}^2}{2b_g} \right) + y_i - C(x_{it}) \quad (8)$$

Similarly to the period utility under the linear pricing plan, we assume that d_{it} is affected by the random shock ν_{it} and $\nu_{it} \sim N(0, \zeta_g^2)$. ν_{it} is observed by customer i at the beginning of period t , before making the decision of minute consumption.

4.2.2 Total Discounted Utility

As a customer's current minute consumption may affect her future marginal price, the customer aims to maximize her total discounted utility by optimizing her consumption over

time. In particular, the total discounted utility at period $t \leq T - 1$ can be presented as

$$U_{it}(u_{it}, u_{i(t+1)}, \dots, u_{iT}) \equiv u_{it} + \sum_{k=1}^{T-t} \delta_g^k u_{i(t+k)}$$

where $\delta_g \in [0, 1]$, representing the discount factor.

We model the customer's minute usage decision as the dynamic optimization problem of a Markov Decision Process (MDP) such that the strategy of minute usage of period t only depends on the then-current state vector s_{it} (Rust (1994)) and the random shock ν_{it} . To facilitate the exposition, we first define $\sigma_{it} = \sigma_{it}(s_{it}, \nu_{it})$ as the strategy of customer i at period t , depending on the state variables s_{it} and random shock. We also define $\Sigma_{it} \equiv (\sigma_{it}, \sigma_{i(t+1)}, \dots, \sigma_{iT})$ as a strategy profile for this MDP from period t onwards; this profile includes a set of decision rules that dictate current and future consumptions. Also denote $V_{it}(s_{it}; \Sigma_{it})$ as the expected continuation utility at period t conditioned on s_{it} and Σ_{it} . Because of the finite horizon of this MDP, V_{it} can be defined recursively as follows:

The expected utility of the terminal period T for a given s_{iT} is

$$V_{iT}(s_{iT}; \Sigma_{iT}) \equiv \mathbf{E}u_{iT}(\sigma_{iT}, y_i - C(\sigma_{iT}); s_{iT}, \nu_{iT}) \quad (9)$$

$$\text{where } C(\sigma_{iT}) = \left(\sum_{k=1}^{T-1} x_{ik} + \sigma_{iT} - A \right) p I_{\sum_{k=1}^{T-1} x_{ik} + \sigma_{iT} > A} \quad (10)$$

where the expectation is taken over the random shock ν_{iT} .

Then the continuation utility function V_{it} at period $t < T$ can be written recursively as

$$V_{it}(s_{it}; \Sigma_{it}) = \mathbf{E}u_{it}(\sigma_{it}, y_i - C(\sigma_{it}); s_{it}, \nu_{it}) + \delta_g [V_{i(t+1)}(s_{i(t+1)}; \Sigma_{i(t+1)}) | s_{it}, \sigma_{it}] \quad (11)$$

$$\text{where } C(\sigma_{it}) = \left(\sum_{k=1}^{t-1} x_{ik} + \sigma_{it} - A \right) p I_{\sum_{k=1}^{t-1} x_{ik} + \sigma_{it} > A} \quad (12)$$

where the expectation is taken over the random shock ν_{it} . Further, given s_{it} , ν_{it} and $\sigma_{it}(s_{it}, \nu_{it})$, the state transition $\pi(s_{i(t+1)}|s_{it}, \sigma_{it})$ is deterministic such that $\sum_{k=1}^t x_{ik} = \sum_{k=1}^{t-1} x_{ik} + \sigma_{it}$, and the customer is one period closer to the terminal period T .

We further recursively define the optimal strategy profile $\Sigma_{it}^* \equiv (\sigma_{it}^*, \sigma_{i(t+1)}^*, \dots, \sigma_{iT}^*)$, starting with the terminal period:

$$\sigma_{iT}^* = \arg \max_{x_{iT}} u_{iT}(x_{iT}, y_i - C(x_{iT}); s_{iT}, \nu_{iT}) \quad (13)$$

and optimal strategy of period $t < T$ is defined recursively as

$$\sigma_{it}^* = \arg \max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \delta_g [V_{i(t+1)}(s_{i(t+1)}; \Sigma_{i(t+1)}^*) | s_{it}, x_{it}] \quad (14)$$

4.2.3 Hyperbolic Discounting

A complication embedded in the dynamic behavior of minute usage is potential time inconsistency among customers. As shown in the literature (cf. Angeletos et al. (2001)), customers may demonstrate time inconsistency in their inter-temporal preferences such that they have a keener preference for short-term return than long-term return (short-term impatience for receiving the return). Accordingly, we also consider an alternative specification to accommodate hyperbolic discounting, which captures the potential time-inconsistency in preference (Phelps and Pollak (1968); Laibson (1997); O'Donoghue and Rabin (1999)).

To be specific, the preference at period $t < T$ is presented by

$$U_{it}(u_{it}, u_{i(t+1)}, \dots, u_{iT}) \equiv u_{it} + \beta_g \sum_{k=1}^{T-t} \delta_g^k u_{i(t+k)}$$

where $\beta_g \in [0, 1]$, $\delta_g \in [0, 1]$. δ_g is the standard exponential discount factor that captures long-term, time-consistent discounting. β_g is the present-bias factor which represents short-term impatience. The commonly used exponential discounting specification is a special case where $\beta_g = 1$ (O'Donoghue and Rabin (1999)).

In comparison to Equation 14, the optimal strategy of period $t < T$ becomes

$$\sigma_{it}^* = \arg \max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \beta_g \delta_g [V_{i(t+1)}(s_{i(t+1)}; \Sigma_{i(t+1)}^*) | s_{it}, x_{it}] \quad (15)$$

5 Estimation and Identification

5.1 Minutes Usage under the Linear Pricing Plans

For a given month τ under the linear pricing plans, we observe customer i 's characteristics D_{it} ($t = 1, 2, \dots, T$). In our specific application, D_{it} is time-variant demographic information, including (1) age and (2) the customer's tenure with the firm.

We also observe individual customer's monthly aggregate usage $q_{i\tau} = \sum_t x_{it}^*$.¹⁰ As shown in Appendix, while there is no closed form for the distribution of $q_{i\tau}$, the distribution can be approximated by a truncated normal distribution. As a result, we can write down the likelihood function of customer i for the minutes usage under linear pricing plans.

$$L_{i,linear}^g = \prod_{\tau} \tilde{f}(q_{i\tau} | \Omega_g) \quad (16)$$

where $\tilde{f}(\cdot)$ is the approximation density function of $q_{i\tau}$ detailed in the Appendix; $\Omega_g \equiv \{\alpha_g, b_g, \zeta_g\}$, i.e., the utility parameters and the distribution of random shocks.

5.2 Minutes Usage in Terminal Period T under the Three-part Tariff

In terminal period T , the consumption becomes a static decision given the allowance will be reset next month. Hence we may solve the optimal minute consumption strategy σ_{iT}^* such

¹⁰Note that for the linear pricing plans, we only observe $q_{i\tau}$ but not the individual x_{it}^* 's.

that

$$\sigma_{iT}^* = \begin{cases} d_{iT} - b_g p, & \text{if } \sum_{t=1}^{T-1} x_{it} + d_{iT} - b_g p > A \\ d_{iT}, & \text{if } \sum_{t=1}^{T-1} x_{it} + d_{iT} < A \\ A - \sum_{t=1}^{T-1} x_{it}, & \text{if } \sum_{t=1}^{T-1} x_{it} + d_{iT} - b_g p \leq A \leq \sum_{t=1}^{T-1} x_{it} + d_{iT} \end{cases} \quad (17)$$

The first component of equation 17 accounts for the situation under which the customer faces a positive marginal price after her cumulative usage exceeds the allowance. The second component represents the situation when the customer's cumulative usage is less than the allowance and the marginal price is zero. The third component represents the situation when the cumulative usage under the optimal σ_{iT}^* exceeds the allowance at a zero marginal price but falls below the allowance with a positive marginal price. We follow Lambrecht et al. (2007) and set the optimal usage under such a situation at a mass point $\sigma_{iT}^* = A - \sum_{t=1}^{T-1} x_{it}$.

According to equation 17, the density of each observed consumption level x_{iT} conditioned on σ_{iT}^* can be written as

$$f_T(x_{iT}|\sigma_{iT}^*, \Omega_g) = \begin{cases} f(x_{iT} = d_{iT} - b_g p) & \text{if } \sum_{t=1}^T x_{it} > A \\ f(x_{iT} = d_{iT}) & \text{if } \sum_{t=1}^T x_{it} < A \\ \Pr(\sum_{t=1}^{T-1} x_{it} + d_{iT} - b_g p \leq A \leq \sum_{t=1}^{T-1} x_{it} + d_{iT}) & \text{if } \sum_{t=1}^T x_{it} = A \end{cases} \quad (18)$$

and the likelihood for the customer i in the terminal period T is

$$L_{iT}^g = f_T(x_{iT}|\sigma_{iT}^*, \Omega_g) \quad (19)$$

5.3 Minutes Usage in Period $t < T$ under the Three-part Tariff

For period $t < T$ of January 2005, we observe each customers period minute consumption x_{it} . However, since there is no closed form solution to the optimal strategy profile Σ_{it}^* , a likelihood function based on observed x_{it} and Σ_{it}^* becomes infeasible. Instead, we implement a numerical

approximation method to establish a simulated likelihood function for estimation. This approximation method contains two steps: (1) using Monte Carlo integration to simulate the value function V_{it} at a subset of state points and interpolating V_{it} at the remaining state points using regression; (2) simulating the density for each observed x_{it} using V_{it} from the previous step. We elaborate each step below.

5.3.1 Simulating and Interpolating $V_{it}(s_{it}; \Sigma_{it}^*)$

Using backward recursion and simulation, it is possible to numerically evaluate the value function under the optimal strategy Σ_{it}^* specified in equations 14, 15, and 17. To be specific, starting with the terminal period T

1. Conditioned on Ω_g , make $nr = 100$ draws from the distribution of random shocks ν_{iT} .
2. Make $ns = 250$ draws of state points $\sum_{k=1}^{T-1} x_{ik}$, i.e., the cumulative minute usage at the beginning of period T .
3. At each of the ns state points, calculate nr optimal minute consumption levels $x_{iT}^*(s_{iT}, \nu_{iT})$ using equation 17, one for each random shock draw ν_{iT} .
4. For each state point that we draw, the continuation value function $V_{iT}(s_{iT})$ can be approximated by $\tilde{V}_{iT}(s_{iT}) = \frac{1}{nr} \sum_{\nu_{iT}} u_{iT}(x_{iT}^*; s_{iT}, \nu_{iT})$.
5. For state points that are not drawn, based on the value functions obtained from step 4, we use a spline interpolation to approximate their values.

Then for period $t < T$, we have the following backward recursion steps:

6. Conditioned on Ω_g , make $nr = 100$ draws from the distribution of random shocks ν_{it} .
7. Make $ns = 250$ draws of state points $\sum_{k=1}^{t-1} x_{ik}$, i.e., the cumulative minute usage at the beginning of period t .

8. At each of the ns state points, conditioned on Ω_g , δ_g , β_g and the $\tilde{V}_{i(t+1)}$, calculate nr optimal minute consumption levels $x_{it}^*(s_{it}, \nu_{it})$ using the following equations (the first for exponential discounting and the second for hyperbolic discounting), one for each random shock draw ν_{it} .

$$x_{it}^*(s_{it}, \nu_{it}) = \arg \max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \delta_g \tilde{V}_{i(t+1)}(s_{i(t+1)} | s_{it}, x_{it})$$

or

$$x_{it}^*(s_{it}, \nu_{it}) = \arg \max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \beta_g \delta_g \tilde{V}_{i(t+1)}(s_{i(t+1)} | s_{it}, x_{it})$$

Note that $s_{i(t+1)}$ is deterministic given s_{it} and x_{it} .

9. For each state point that we draw, the continuation value function $V_{it}(s_{it})$ can be approximated by.

$$\tilde{V}_{it}(s_{it}) = \frac{1}{nr} \sum_{\nu_{it}} [u_{it}(x_{it}^*; s_{it}, \nu_{it}) + \delta_g \tilde{V}_{i(t+1)}(s_{i(t+1)} | s_{it}, x_{it}^*)]$$

or

$$\tilde{V}_{it}(s_{it}) = \frac{1}{nr} \sum_{\nu_{it}} [u_{it}(x_{it}^*; s_{it}, \nu_{it}) + \beta_g \delta_g \tilde{V}_{i(t+1)}(s_{i(t+1)} | s_{it}, x_{it}^*)]$$

10. For state points that are not drawn, based on the continuation functions obtained from step 9, we use a spline interpolation to approximate their values.

5.3.2 Simulating the Density of Observed x_{it} , $\tilde{f}_{it}(x_{it} | s_{it}, \Omega_g, \delta_g)$ or $\tilde{f}_{it}(x_{it} | s_{it}, \Omega_g, \delta_g, \beta_g)$

For each x_{it} observed in the data and its corresponding state point s_{it} , we use the following steps to simulate its density:

1. First draw $nr_{density} = 100$ random shocks ν_{it} ;
2. For each random draw of ν_{it} and the observed s_{it} , calculate the optimal minute consumption by solving the following equations (the first for exponential discounting and

the second for hyperbolic discounting).¹¹

$$x_{it}^*(s_{it}, \nu_{it}) = \arg \max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \delta_g \tilde{V}_{i(t+1)}(s_{i(t+1)} | s_{it}, x_{it})$$

or

$$x_{it}^*(s_{it}, \nu_{it}) = \arg \max_{x_{it}} u_{it}(x_{it}, y_i - C(x_{it}); s_{it}, \nu_{it}) + \beta_g \delta_g \tilde{V}_{i(t+1)}(s_{i(t+1)} | s_{it}, x_{it})$$

3. Using the calculated $nr_{density} = 100$ optimal $x_{it}^*(s_{it}, \nu_{it})$'s, simulate $\tilde{f}_{it}(\cdot)$, the density of the observed x_{it} , using Gaussian kernel density estimator.

With the simulated densities for all observed x_{it} , we are able to write a likelihood function for customer i such that

$$L_i^g = \prod_t \tilde{f}_{it}(x_{it} | s_{it}, \Omega_g, \delta_g)$$

or

$$L_i^g = \prod_t \tilde{f}_{it}(x_{it} | s_{it}, \Omega_g, \delta_g, \beta_g)$$

5.4 Heterogeneity

We use a finite mixture approach to capture heterogeneity because i) this approach invokes minimal structure on the distribution of preferences and ii) the limited number of observations per subject suggests person-specific effects would be weakly identified. The prior probability of customer i belonging to segment g is $f_{ig} = \exp(\lambda_{0g} + D_i' \lambda_g) / \sum_{g'} \exp(\lambda_{0g'} + D_i' \lambda_{g'})$, where D_i is time-invariant demographic information, including gender and rural residency

¹¹In step 7 and step 8 of section 5.3.1, one may choose to include the observed states in the set of $ns = 250$ draws of state points. Then there would be no need to recompute the optimal x_{it}^* 's in this current step. However, observed states may be sparse in some areas of the state space. As a result, the interpolated \tilde{V}_i 's may be inaccurate in those areas. So we choose not to use the observed states in section 5.3.1. Instead, we draw all of the 250 state points randomly so as to cover the state space as well as possible and then supplement these draws with the observed states.

status. Consequently, the unconditional likelihood of the whole data is

$$L = \prod_i \sum_g [f_{ig} L_{i,linear}^g L_{iT}^g]$$

We use MLE to estimate the parameters.

5.5 Identification

We provide an informal discussion of the identification of parameters. Since the identification of segment parameters λ_g follows classical argument (cf. McHugh, 1956), we will focus on the remaining parameters. Besides λ_g 's, the parameters that construct our model can be categorized into two sets. The first set of parameters appear under both the linear pricing plan and the three-part tariff plan, including $\Omega_g \equiv \{\alpha_g, b_g, \zeta_g\}$, i.e., the utility parameters and the distribution of random shocks. The second set of parameters only affect the demand under the three-part tariff plan, including the discount factors δ_g and β_g . In essence, the parameters $\Omega_g \equiv \{\alpha_g, b_g, \zeta_g\}$ are identified from choices under the linear plan and the terminal period of the three-part tariff, where there are no dynamics involved. Conditioned on Ω_g , we then recover the discount factors δ_g and β_g .

5.5.1 The Identification of Ω_g

The consumption decisions under the linear plans and the terminal period of the three-part tariff have no dynamics involved. Besides individual consumption across time, we further observe the following information under the linear plan and the terminal period of the three-part tariff.

- Different linear prices across individual customers.
- Depending on whether a customer has exhausted her allowance at the beginning of the terminal period, there are variations in marginal prices across individuals.

- Variation in demographic characteristics across individuals and time.

The variations in consumption across and within individuals over time, conditioned on the variations in prices and demographics, enable us to identify α_g and b_g . Together, α_g , b_g , prices and demographics determine the mean levels of consumptions of each individual over time. The observed deviations from such mean levels across individuals and time identify ζ_g , the standard deviation of the random shocks ν_{it} 's.

5.5.2 The Identification of Discount Factors

Myopic customers do not tradeoff consumption over time, meaning they are inclined to ignore potential overage charges later in the month and, as a result, consume many minutes earlier in the month. In contrast, fully forward-looking customers (no discounting of future utilities) consider overage and lower early minute usage accordingly. Hence, there is a difference in the distribution of minutes over time between the two types, with forward-looking consumers shifting a greater proportion of consumption to later periods. These difference in consumption patterns over time enable identification of the discount rate, conditioned on knowing the utility of consumption.

The fact that such intertemporal tradeoffs might be inconsistent between contiguous periods and discontinuous periods distinguish hyperbolic discount factor from exponential discount factor.

More formally, conditioned on the already identified Ω_g , the state variables, marginal prices and demographics at each period $t < T$, we can compute the static consumption levels if the customers were myopic (i.e., $\delta_g = 0$). Instead, if the data demonstrate inconsistency from those static consumption levels, the discount factors can be identified.

More important, when there are at least three periods of data within a billing cycle, exponential discount factor and hyperbolic discount factor can be separately identified (Fang and Silverman, 2009). For example, consider three periods of consumption data. For the case wherein there is only *exponential* discounting, during the first period, observed consumptions

would be consistent with a pattern of discounting the second period’s utility with a factor of δ_g and discounting the third period with a factor of δ_g^2 . In comparison, if there is hyperbolic discounting, the observed consumption should be consistent with discounting the second period by a factor of $\beta_g \delta_g$ and discounting the third period by a factor of $\beta_g \delta_g^2$. Unless $\beta_g = 1$, the aforementioned two data generating processes and hence the observed data are distinguished. Consequently, δ_g and β_g are separately identified.

6 Results

To conserve space, we report results for the 83 customers (15% of the observations) who select the three-part tariff plan with a monthly access fee of 168RMB (about \$20.30) an allowance of 800 minutes, and a marginal price of 0.4RMB. As a robustness check, extending the analysis to a second group of 284 customers (access fee 98RMB, allowance 450 minutes) indicates the results change little. The key difference stems from an increase in the standard errors on the order of 10%; this increase arises from computational considerations that require a smaller numbers of simulations than the focal group.¹²

6.1 Segmentation

We consider different potential degrees of unobserved heterogeneity based on the number of segments. Table 4 compares the BIC for alternative specifications of segment numbers and indicates that a two-segment model provides the best fit. Accordingly, our subsequent analyses are predicated upon the 2-segment specification.

[Insert Table 4 about here.]

Based on the magnitude of the segment parameters, the implied segment sizes are 88.8% and 11.2%, respectively. We find that urban male customers are more likely to be in segment

¹²As the second group is three times larger, we reduce the number of draws; instead of $nr = nr_{density} = 100$, $ns = 250$, we use $nr = nr_{density} = 50$, $ns = 100$.

2. Given the nature of the Chinese economy and society, such a cohort of customers are more likely to be white-collar or business people with a relatively higher income level.

6.2 Parameter Estimates

Table 5 reports the estimates of the two-segment model.

[Insert Table 5 about here.]

We find that forward-looking behavior exists for both segments. Segment 2 customers, whom we conjecture to be more likely business oriented, have a higher exponential discount factor, which implies that they are more patient than segment 1 customers (0.91 vs. 0.86).¹³ Of particular interest, discount factors in both segments are much smaller than those typically used in empirical studies (mean=0.979, see Table 1). Placing this result in perspective, a weekly discount factor of 0.995 implies that a consumer values a one-minute phone call at the beginning of the month to be worth about 1.02 minutes at the end of the month (under the assumption of a constant pricing rate). In contrast, our estimates indicate a minute phone call now is worth closer to 1.6 to 2.1 minutes at the end of the month (depending on the consumer segment).

Like Dubé et al. (2010b), we do not find strong support for the existence of hyperbolic discounting in our context. For both segments, the estimates of hyperbolic discount factors are approaching 1. Since the exponential discounting model is nested within the hyperbolic discounting model with $\beta_g = 1$, we also implement the nested Likelihood Ratio test for the two specifications. We can not reject the null hypothesis that the hyperbolic discount factors are not different from 1. Such a result is consistent with Chevalier and Goolsbee (2005) and Dubé et al. (2010b). Although time-inconsistent preferences and hence hyperbolic discounting exist (Angeletos et al., 2001), it may not be universal in all contexts.

¹³We reparametrize the discount factors as $\delta_g = \exp(\pi_g)/(1 + \exp(\pi_g))$ during estimation. The same treatment applies to β_g .

Between the two segments, segment 2 (11.2%) has a higher average consumption level, lower variance in usage, and is less price sensitive. This is also consistent with our intuition about the nature of more business oriented usage. Such a pattern makes segment 2 customers potentially better candidates for targeting with a higher price level. We will further explore the implication on firm pricing strategy in next section.

7 Managerial Implications

7.1 Usage Prediction and Intertemporal Substitution Pattern

7.1.1 Biased Price Effects

To assess the potential bias in model estimates arising from specifying the commonly employed discount factor of 0.995 rather than using the estimated heterogeneous discount factors, we re-estimate the model by fixing the discount factor to 0.995. While there is little impact on most estimates, we find the price coefficients to be underestimated (have smaller absolute magnitudes) by 23% in segment 1 and 15% in segment 2. The price coefficients of segment 1 and segment 2 become 2.31 and 2.25 (vs. 2.99 and 2.64 in Table 5), with the standard deviations as 0.03 and 0.04, respectively.¹⁴ Setting a higher discount factor such as 0.995 implies customers excessively substitute future consumption for current within allowance consumption. Given future over-allowance consumption is costly, the model compensates by lowering price sensitivity to generate the same level of overall utility for the future consumption occasion. Consequently, the smaller price coefficients imply that the overage charge has less impact on future utility; as a result there is no need for the customer to make the tradeoff between consumptions across time.

¹⁴The estimates of remaining parameters can be obtained from the authors.

7.1.2 Biased Forecasts

To ascertain how well the model fits the data and resulting intertemporal substitution pattern, we calculate the mean absolute percentage error (MAPE) and mean percentage error (MPE) across segments and time under both the estimated discount rates and under the assumed weekly discount factor of 0.995. The MAPE measures a model's overall accuracy of fitting the data while the MPE indicates bias in model predictions. Table 6 and Table 7 depict the results.

[Insert Table 6 about here.]

According to Table 6, the fit under the 0.995 discount factor is universally worse than the fit under the estimated coefficients across segments and across time. To develop a better sense of why the higher discount factor performs more poorly, we next turn to the MPE.

[Insert Table 7 about here.]

Based on Table 7, there is no obvious forecasting bias from using the higher discount rate when summing across all periods, yet aggregating across time obscures the patterns in intertemporal substitution. When setting the discount factor at 0.995, the demand in the first 4 periods is under-estimated; and the demand in the last period is over-estimated. This bias occurs because that customers are more impatient than what $\delta_g = 0.995$ implies. As a result, in the early periods, when the allowance has not been exhausted, impatient customers consume more than predicted under 0.995 discount factor. Further, customers are more price sensitive than the 0.995 discount factor case implies (recall that the price coefficients are underestimated under the 0.995 discount factor). As a result, customers in overage (roughly coincident with the last period) evidence lower consumption than predicted under the 0.995 discount factor.

7.2 Elasticities

To ascertain how customers' minutes usage varies under alternative allowance and the marginal price levels, we compute their respective monthly minutes demand changes across segments for both the estimated discount factors and 0.995. Table 8 presents the results.

[Insert Table 8 about here.]

The elasticities in Table 8 suggest that the 0.995 discount factor leads to an underestimation of the effect of allowances and price on usage (that is, users are not as price sensitive as it implies when the discount factor is set to 0.995). Were consumers to actually have a discount factor of 0.995, they would be more forward looking than they were under the actual discount rates we estimate. As a result, the more forward looking consumers implied by 0.995 should conserve minutes so as not to pay overage in later periods. Because they do not actually conserve minutes, the model with a 0.995 discount factor needs to rationalize the observed overage. It does so by estimating a relative lower sensitivity to price and allowance; a lower price and allowance sensitivity means that consumers do not mind paying overage as much and have lower elasticities.

7.3 Alternative Pricing Schedule

Based on our communication with the data provider, their process of picking the three-part tariffs in this field experiment is ad hoc. There was no optimization consideration during the design of the experiment. As a result, the focal three-part tariff is not likely to be optimal for the firm in terms of maximizing its revenue. To assess the potential for revenue improvement, we create a grid of alternative allowance and price levels for each segments. For

each combination of allowance and price, we calculate the percentage of revenue change.¹⁵ Table 9 reports the results.

[Insert Table 9 about here.]

Table 9 includes the current plan (allowance=800 minutes, price=RMB0.40). Surrounding the current plan, each column from left to right represents a 2-cent change in the marginal price for minutes in overage and each row from top to bottom stands for a 20-minute change in the allowance. As customers are heterogeneous, especially in their price sensitivity, the optimal price structure differs across the two segments. For both segments, a lower allowance enhances the possibility of overage; and a moderately decreased price tends to increase the consumption level under the overage situation. For segment 1, the customers are more impatient. Hence the optimal allowance level of segment 1 is relatively higher than that of segment 2 since the former are already more likely to be overage (780 vs. 760). Also, segment 1 are more price sensitive. So the corresponding optimal marginal price for segment 1 should be lower than that of segment 2 (0.36 vs. 0.38). The revenue of the firm would increase by 0.42% and 1.90% for segment 1 and segment 2, respectively. To the extent that similar exercises can be implemented across all groups of customers, the revenue increase would be considerable.

As shown earlier, under the discount factor of 0.995, the model may lead to biased estimates of coefficients and elasticities. To see whether such biases may lead to inaccurate policy recommendations, we re-create the same grid but calculate the revenue changes using the estimates under the 0.995 discount factor. Table 10 reports the results. As indicated in the Table, with the 0.995 discount factor, the model generates notably different pricing

¹⁵Note that we do not model plan choice since there are no plan choice data available. To ensure that the presented price/allowance changes do not lead to customer leaving the company or opt to a different plan, we calculate customer welfare for each point on the grid as measured by customers' total discounted utility. We then compare it with the welfare level under the original plan. None of the welfare changes is significantly different from zero, hence we do not believe that the recommended policies will result in substantial plan switching. Further, the company had a significant market share and there was no cellphone number portability in China until October, 2010 (ChinaTechNews.com, 2010). Consequently, we conclude (1) that the new price structures in the Table are unlikely causing large customer churning and plan switching, and (2) that the effect of competitive response is likely to be modest.

plan recommendations for both segments. Since the assumed discount factor is much higher, to enhance customers' likelihood of overage, the allowance levels would be much lower. As a result, under the 0.995 discount factor, the optimal allowance levels for segment 1 and segment 2 become 700 and 720 (instead of 780 and 760), respectively. Further, since the price sensitivities are underestimated, the optimal price would be higher. This effect manifests for segment 1, the optimal price changes from 0.36 to 0.38. In short, the firm sets its allowance too low and its marginal price somewhat high, thereby overcharging its customers when using the standard practice of setting discount rates.

[Insert Table 10 about here.]

The predicted revenue gains are also quite different between the two scenarios, ($\delta_1 = 0.86, \delta_2 = 0.91$) vs. ($\delta_1 = \delta_2 = 0.995$). To illustrate the difference, for each grid point of each segment, we calculate the predicted revenue difference between the two scenarios. As shown in the Figure 4, the percentage differences can be substantial. By implementing the pricing plans as suggested by the model with $\delta_1 = \delta_2 = 0.995$, the firm would foregone potential revenue gains. Take segment 1 as an example, the firm's revenue improvement would be 1.90% with an allowance of 760 and a price of 0.38 (where $\delta_1 = 0.86, \delta_2 = 0.91$). Instead, if the firm adopted the plan with an allowance of 720 and a price of 0.38 (where $\delta_1 = \delta_2 = 0.995$), the revenue improvement would only be 0.55%, which would reduce the potential gains by 74% relative to the optimal pricing level. As for segment 2, the corresponding loss would be even higher, at 88%. It is interesting to note that the potential bias is more substantial for the business oriented segment with its relatively lower level of price sensitivity.

[Insert Figure 4 about here.]

8 Conclusion

Owing to the ability to capture the trade-off between long-term and short-term goals, the application of dynamic structural models to consumer and firm decision making has become increasingly widespread. However, dynamic structural models face a fundamental identification problem, namely, the preference, the state transition, and the discount factor are confounded and become difficult to identify simultaneously. Should the rate be misspecified, inferences about agent behavior might be misleading and the implied policies for improving agent welfare might be suboptimal.

To address this problem, several solutions have been proposed. The most common approach has been to assume a fixed discount factor that is consistent with the market interest rate and is common across individuals. Given that consumers' discounting behavior may be inconsistent from the market interest rate and may vary across individuals (Frederick et al., 2002), it would be desirable to relax these assumptions.

A second identification strategy is to invoke the exclusion restriction condition (Magnac and Thesmar, 2002), where a set of exogenous variables affecting future state transition but not current utility. The variation in these exogenous variables then helps to separately identify the utility and discount factor. Though a promising approach, exogenous variables may not exist in some contexts and when they do, the exogeneity assumption may be hard to validate.

A third approach, pioneered by Dubé et al. (2010b), is to use experiment data that enable researchers to disentangle discount effects from changes in utility; and our research extends work in this vein. We augment this analysis by considering field studies, an approach that considers decisions made over a longer duration than regular lab settings to measure their trade-offs and involves monetary incentives on the scale of the choices.

Accordingly, we advance the literature on identifying heterogeneous discount factors by using field data to measure them. Specifically, we estimate a dynamic structural model using consumers' cellphone usage data. The data contain observations of consumers' cellphone

consumptions under both static setting and dynamic setting. Using the static data, we first identify consumers' heterogeneous utilities and the distribution of random consumption shocks. Conditioned on the identified utilities and random shocks, we then recover the heterogeneous discount factors using the dynamic data.

Findings suggest that discount rates in practice (0.86 and 0.91 across segments) are well below those commonly assumed in the literature (0.995). As a consequence, price effects are underestimated in our application. Moreover, the higher rate leads to a mistaken presumption that more minutes would be saved for later use, leading to a 27% increase in the mean absolute percentage error in model fit. The attendant consequences for pricing policy are notable, leading to pricing recommendations that are generally too high and would lower potential revenue gains by 74-88%.

The inherent complexity of dynamic structural models often requires simplifications that correspondingly represent future research opportunities. Our model is no exception. First, we note that risk aversion may play a role in consumers' dynamic decision making. While the quadratic (concave) form of utility we estimate under linear pricing does not explicitly capture risk aversion (there is no uncertainty so there can be no risk), this form of utility does have risk implications for forward looking consumers. Given the quadratic utility, an increase in the forecast demand variation will strictly lower utility over the case where future demand is less volatile. Hence, riskier decisions have lower utility in our model. That said, it would be desirable to allow the utility function to accommodate differences between satiation (as in the linear pricing case) and risk aversion (as in the three-part tariff); our analysis, like all those that proceed it (e.g., Erdem and Keane (1996)), do not make this distinction. Future research should therefore consider to collect data that enable researchers to disentangle these two effects.

Second, our study focuses on a specific consumption context with a small focal group of customers over a specific duration. Therefore, the results may not generalize to other contexts involving different consumers, decisions or decision durations. Hence, more research

is necessary to generalize the degree to which discount rates are an inherent trait or the degree to which they are context dependent. For example, it would be fruitful to consider how discount rates might vary in practice when one considers different contexts of intertemporal consumption; do consumers invoke the same level of patience when making choices over years as they do when making decisions over days?

Third, we do not explicitly consider learning, but rather control for it by incorporating a consumption effect for new users. While most users in our data are experienced, estimates from Table 5 indicate the few users who are new in segment 2 evidence lower consumption rates. Clearly, uncertainty plays a crucial role for new users in their plan choices and consumption levels (Lambrecht, 2006; Lambrecht et al., 2007; Iyengar et al., 2007). Accordingly, a more formal characterization of learning that also considers consumers' forward-looking behavior would provide more insights in contexts with a greater number of novice consumers.

Fourth, two potential sources of selection bias exist in our field study. The first arises from the firm's choices of customers to participate in the plans. Per our discussion with the firm's managers, customer selection was randomized so this form of selection bias is not germane. The second selection bias arises from the customer's decision of whether to adopt the three-part tariff plan that was offered. If the decision to adopt the plan is correlated with usage model error, then our estimates will be biased. This correlation could arise from unobserved person specific factors common to both choices, or omitted person time effects. Via inclusion of unobserved heterogeneity, we control for the former. Given plan choice is not a time varying decision over the duration of our data, the latter source of omitted factors are also not likely to be material. Regardless, one limit of our data is that we have no valid instruments to model plan choice, nor do we observe which agents declined the plan. Accordingly, to the extent selection does manifest, our analyses should be considered to be conditional on plan adoption. As a result, another area of interest is to extend our research into the domain of plan choice.

In sum, this paper is the first (to our knowledge) to provide field-study based evidence regarding the nature of discount rates that obviate the need for structural assumptions or exclusion restrictions to identify discount rates. Consistent with Dubé et al. (2010b) and Ishihara (2010), we find evidence that discount rates are substantially lower than those used in practice and that this difference is material from a policy perspective. Given the widespread use of dynamic models in marketing and economics, we hope our analysis will spark future work to pin down how such intertemporal trade-offs are made in practice.

References

- Akerberg, Daniel A. 2003. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review* **44**(3) 1007–1040.
- Allcott, Hunt, Nathan Wozny. 2011. Gasoline prices, fuel economy, and the energy paradox. *Working Paper* .
- Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, Stephen Weinberg. 2001. The hyperbolic consumption model: Calibration, simulation, and empirical evaluation. *The Journal of Economic Perspectives* **15**(3) pp. 47–68.
- Ascarza, Eva, Anja Lambrecht, Naufel Vilcassim. 2009. When talk is “free”: An analysis of subscriber behavior under two- and three-part tariffs. *Working Paper* .
- Chevalier, Judith, Austan Goolsbee. 2005. Are durable goods consumers forward-looking? evidence from college textbooks. *NBER Working Paper 11421* (4).
- ChinaTechNews.com. 2010. China to start trial of mobile phone number portability in october 2010 .
- Chung, Doug, Thomas Steenburgh, K. Sudhir. 2010. Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans. *Working Paper* .
- Dubé, Jean-Pierre, Gunter J. Hitsch, Pradeep Chintagunta. 2010a. Tipping and concentration in markets with indirect network effects. *Marketing Science* **29**(2) 216–249.
- Dubé, Jean-Pierre, Gunter J. Hitsch, Pranav Jindal. 2010b. Estimating durable goods adoption decisions from stated preference data. *Working Paper* .
- Dubin, Jeffrey A., Daniel L. McFadden. 1984. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* **52**(2) pp. 345–362.
- Erdem, Tulin, Michael P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science* **15**(1) 1–20.
- Fang, Hanming, Dan Silverman. 2009. Time-inconsistency and welfare program participation: Evidence from the NLSY. *International Economic Review* **50**(4) 1043–1077.
- Fang, Hanming, Yang Wang. 2010. Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *NBER Working Paper* .
- Frederick, Shane, George Loewenstein, Ted O’Donoghue. 2002. Time discounting and time preference: A critical review. *Journal of Economic Literature* **40**(2) pp. 351–401.
- Gordon, Brett R. 2009. A dynamic model of consumer replacement cycles in the pc processor industry. *Marketing Science* **28**(5) 846–867.
- Härdle, Wolfgang, Oliver Linton. 1994. Applied nonparametric methods. Robert Engle, Daniel McFadden, eds., *Handbook of Econometrics*, vol. 4, chap. 38. Amsterdam: Elsevier Science, 2295–2339.

- Harrison, Glenn W., Morten I. Lau, Melonie B. Williams. 2002. Estimating individual discount rates in denmark: A field experiment. *The American Economic Review* **92**(5) pp. 1606–1617.
- Hartmann, Wesley R. 2006. Intertemporal effects of consumption and their implications for demand elasticity estimates. *Quantitative Marketing and Economics* **4**(4) 325–349.
- Hartmann, Wesley R., Harikesh S. Nair. 2010. Retail competition and the dynamics of demand for tied goods. *Marketing Science* **29**(2) 366–386.
- Hausman, Jerry A. 1979. Individual discount rates and the purchase and utilization of energy-using durables. *The Bell Journal of Economics* **10**(1) pp. 33–54.
- Hendel, Igal, Aviv Nevo. 2006. Measuring the implications of sales and consumer inventory behavior. *Econometrica* **74**(6) 1637–1673.
- Hotz, V. Joseph, Robert A. Miller. 1993. Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies* **60**(3) 497–529.
- Ishihara, Masakazu. 2010. Dynamic demand for new and used durable goods without physical depreciation: The case of japanese video games. *Working Paper* .
- Iyengar, Raghuram, Asim Ansari, Sunil Gupta. 2007. A model of consumer learning for service quality and usage. *Journal of Marketing Research* **44**(4) 529–544.
- Kamakura, Wagner A., Gary J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* **26**(4) 379–390.
- Kim, Jun, Paulo Albuquerque, Bart J. Bronnenberg. 2009. Online demand under limited consumer search. *Working Paper* .
- Laibson, David. 1997. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* **112**(2) 443–477.
- Lambrecht, Anja, Katja Seim, Bernd Skiera. 2007. Does uncertainty matter? consumer behavior under three-part tariffs. *Marketing Science* **26**(5) 698–710.
- Lambrecht, Bernd Skiera, Anja. 2006. Paying too much and being happy about it: Existence, causes and consequences of tariff-choice biases. *Journal of Marketing Research* **43**(2) 212–223.
- Magnac, Thierry, David Thesmar. 2002. Identifying dynamic discrete decision processes. *Econometrica* **70**(2) 801–816.
- McHugh, Richard. 1956. Efficient estimation and local identification in latent class analysis. *Psychometrika* **21** 331–347, 10.1007/BF02296300.
- Misra, Sanjog, Harikesh Nair. 2009. A structural model of sales-force compensation dynamics: Estimation and field implementation. *Working Paper* .
- Narayanan, Sridhar, Pradeep Chintagunta, Eugenio Miravete. 2007. The role of self selection, usage uncertainty and learning in the demand for local telephone service. *Quantitative Marketing and Economics* **5**(1) 1–34.

- O'Donoghue, Ted, Matthew Rabin. 1999. Doing it now or later. *The American Economic Review* **89**(1) 103–124.
- Phelps, Edmund S., Robert A. Pollak. 1968. On second-best national saving and game-equilibrium growth. *The Review of Economic Studies* **35**(2) 185–199.
- Rust, John. 1987. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society* **55**(5) 999–1033.
- Rust, John. 1994. Structural estimation of markov decision processes. Robert F. Engle, Daniel L. McFadden, eds., *Handbook of Econometrics*, vol. IV. Amsterdam: Elsevier Science.
- Sun, Baohong. 2005. Promotion effect on endogenous consumption. *Marketing Science* **24**(3) 430–443.
- Warner, John T., Saul Pleeter. 2001. The personal discount rate: Evidence from military downsizing programs. *The American Economic Review* **91**(1) pp. 33–53.

Table 1: Example Discount Rates in Empirical Studies

Study	Choice Domain	Approach ¹	Discount factor	Period	Weekly Discount Factor	Implied Weekly Interest Rate (%)
Rust (1987)	Bus engine replacement	F	0.9999	month	→ 1	0.002
Hotz and Miller (1993)	Children/sterilization	E	0.65 (0.68)	year	0.992	0.83
Erdem and Keane (1996)	Laundry detergent	F	0.995	week	0.995	0.50
Ackerberg (2003)	Yogurt	E	0.98 (0.02)	week	0.98	2.04
Hartmann (2006)	Golf	F	0.99	day	0.932	7.29
Gordon (2009)	Personal computers	F	0.98	month	0.995	0.47
Kim et al. (2009)	Camcorders	F	1.0	minutes	1.0	0
Dubé et al. (2010a)	Video game consoles	F	0.9	month	0.976	2.49
Dubé et al. (2010b)	HD DVD players	P	0.7	year	0.993	0.69
Hartmann and Nair (2010)	Razors and blades	F	0.998	week	0.998	0.20
Chung et al. (2010) ²	Salesforce compensation	E	0.95	month	0.988	1.20
Fang and Wang (2010) ²	Mammography exams	E	0.72 (0.09) 0.80 (0.03)	two years	0.997 0.998	0.32 0.21
Ishihara (2010)	Video games	E	0.885 (0.006)	week	0.885	12.99
Mean					0.979	2.25
Std. Dev					0.034	3.76

Notes:

1. F in the “Approach” column indicates an assumed fixed value for the discount factor. The study labeled P estimates the discount rate using experimental data. E indicates the discount parameter is estimated by functional restrictions and/or the use of exclusion restrictions. The standard errors of the estimates are reported in the parentheses.

2. Chung et al. (2010) and Fang and Wang (2010) also consider hyperbolic discounting. We only report their results of exponential discount factors. Fang and Wang (2010) use two specifications in their estimation, hence we report two discount factors. Chung et al. (2010) obtain the discount rate via grid search so there is no sampling error to report. Note that the grid search approach yields estimated parameter distributions that are conditionally marginal with respect to the discount rate, which can lead to inefficient estimates.

Table 2: Three-part Tariff Plans

Access Fee (CNY)	Allowance (Minutes)	Marginal Price (CNY)	Number of Enrollees	Percentage
98	450	0.40	284	50.35
128	600	0.40	111	19.68
168	800	0.40	83	14.72
218	1100	0.36	50	5.92
288	1500	0.36	21	2.86
388	2500	0.30	15	2.31

Table 3: Summary Statistics

	Mean	Std. Dev.	Min.	Max.
Monthly Usage under Linear Plan/Allowance Level	0.92	0.46	0.02	2.22
Monthly Usage under Three-part Tariff/Allowance Level	0.96	0.35	0.01	1.63
Female	0.16	-	0	1
Rural Residency	0.41	-	0	1
Age (years)	36.18	7.45	19	58
New Customer (enrolled less than 12 months)	0.16	0.36	0	1

Table 4: Alternative Numbers of Latent Segments

	BIC
1 Segment	3411.11
2 Segments	3381.32
3 Segments	3405.01
4 Segments	3430.21

Note: Bold fonts indicate the best fit

Table 5: Estimates of Model Parameters

	Segment 1 (88.8%) Rural, Female	Segment 2 (11.2%) Urban, Male
	Estimate (S.E.)	Estimate (S.E.)
Satiation		
Constant	4.99 (0.21)	5.29 (0.55)
Price (cent)	2.99 (0.02)	2.64 (0.03)
New Customer	-0.06 (0.05)	-0.30 (0.05)
Age	0.03 (0.11)	0.29 (0.38)
Age ²	-0.02 (0.02)	-0.04 (0.06)
Std. Dev. of Shocks	8.99 (1.91)	8.75 (2.31)
Segment Parameters		
Constant	1.83 (0.31)	---
Rural Residency	0.39 (0.08)	---
Female	0.51 (0.12)	---
Discount Factors	0.86 (0.04)	0.91 (0.05)

Note: Bold fonts indicate the estimates being significant at 95% level.

Table 6: Mean Absolute Percentage Error (MAPE) Comparison

		Mean Absolute Percentage Error		
		Segment 1	Segment 2	Aggregate
Estimated Discount Factors	First 4 Periods	0.09	0.10	0.09
	Final Period	0.10	0.09	0.10
	Monthly Aggregate	0.10	0.10	0.10
0.995	First 4 Periods	0.13	0.12	0.13
	Final Period	0.13	0.15	0.14
	Monthly Aggregate	0.13	0.13	0.13

Table 7: Mean Percentage Error (MPE) Comparison

		Mean Percentage Error		
		Segment 1	Segment 2	Aggregate
Estimated Discount Factors	First 4 Periods	0.02	-0.02	0.01
	Final Period	$\rightarrow 0$	0.01	$\rightarrow 0$
	Monthly Aggregate	0.01	$\rightarrow 0$	0.01
0.995	First 4 Periods	-0.06	-0.05	-0.06
	Final Period	0.08	0.05	0.08
	Monthly Aggregate	0.02	$\rightarrow 0$	0.02

Table 8: Demand Elasticities

		Demand Elasticity w.r.t. Price		
		Segment 1	Segment 2	Aggregate
Estimated Discount Factors		0.10 (0.08, 0.11)	0.08 (0.07, 0.09)	0.09 (0.08, 0.11)
0.995		0.08 (0.07, 0.08)	0.05 (0.04, 0.07)	0.08 (0.06, 0.08)
		Demand Elasticity w.r.t. Allowance		
		Segment 1	Segment 2	Aggregate
Estimated Discount Factors		0.29 (0.27, 0.31)	0.08 (0.06, 0.10)	0.28 (0.27, 0.30)
0.995		0.25 (0.24, 0.26)	0.03 (0.02, 0.03)	0.24 (0.24, 0.25)

Note 1: 95% confidence intervals are in parentheses.

Note 2: Bold fonts indicate the biases are significant ($p < 0.05$).

Table 9: Revenue Percentage Change under Alternative Pricing Schedules ($\delta_1 = 0.86, \delta_2 = 0.91$)

Segment 1		Marginal Price					
Allowance (minutes)	0.34	0.36	0.38	0.40	0.42	0.44	0.46
680	-0.42	-0.34	-0.26	-0.30	-0.37	-0.39	-0.44
700	-0.01	0.02	0.05	0.09	0.07	0.01	0.002
720	0.16	0.20	0.24	0.29	0.27	0.22	0.14
740	0.23	0.24	0.25	0.21	0.17	0.16	0.12
760	0.33	0.37	0.34	0.32	0.30	0.29	0.25
780	0.37	0.42	0.40	0.39	0.37	0.34	0.31
800	0.04	0.05	0.02	0	-0.26	-0.56	-0.79
820	-0.23	-0.22	-0.26	-0.28	-0.85	-1.16	-1.37
840	-0.51	-0.46	-0.48	-0.53	-1.43	-1.74	-2.62

Segment 2		Marginal Price					
Allowance (minutes)	0.34	0.36	0.38	0.40	0.42	0.44	0.46
680	-0.72	-0.70	-0.67	-0.63	-0.60	-0.51	-0.55
700	-0.35	-0.22	-0.10	-0.07	-0.05	-0.09	-0.13
720	0.08	0.13	0.50	0.46	0.42	0.32	0.08
740	1.02	1.07	1.14	1.10	0.76	0.38	0.11
760	1.56	1.70	1.90	1.37	1.19	0.54	0.41
780	1.01	1.04	1.07	1.09	1.17	1.03	0.82
800	-0.11	-0.05	-0.02	0	0.21	0.48	0.29
820	-0.17	-0.14	-0.12	-0.10	-0.15	-0.55	-0.86
840	-0.25	-0.23	-0.21	-0.20	-0.32	-0.63	-0.91

Table 10: Revenue Percentage Change under Alternative Pricing Schedules ($\delta_1 = \delta_2 = 0.995$)

Segment 1		Marginal Price					
Allowance (minutes)	0.34	0.36	0.38	0.40	0.42	0.44	0.46
680	0.86	0.89	0.93	0.92	0.92	0.91	0.90
700	1.05	1.10	1.13	1.12	1.10	1.09	1.01
720	0.68	0.70	0.71	0.73	0.72	0.71	0.71
740	0.50	0.52	0.53	0.55	0.52	0.50	0.49
760	0.32	0.34	0.38	0.37	0.36	0.35	0.33
780	0.15	0.18	0.19	0.20	0.20	0.19	0.19
800	-0.02	-0.01	-0.002	0	-0.002	-0.01	-0.02
820	-0.25	-0.25	-0.25	-0.26	-0.27	-0.28	-0.30
840	-0.49	-0.49	-0.50	-0.51	-0.52	-0.54	-0.55

Segment 2		Marginal Price					
Allowance (minutes)	0.34	0.36	0.38	0.40	0.42	0.44	0.46
680	0.10	0.13	0.15	0.18	0.10	0.02	-0.04
700	0.16	0.19	0.22	0.24	0.17	0.09	0.01
720	0.28	0.32	0.55	0.38	0.30	0.28	0.15
740	0.22	0.25	0.31	0.28	0.24	0.16	0.08
760	0.04	0.17	0.21	0.22	0.14	0.06	0.01
780	-0.02	0.01	0.04	0.06	0.08	0.10	0.05
800	-0.07	-0.05	-0.02	0	0.02	0.04	0.01
820	-0.16	-0.14	-0.11	-0.08	-0.10	-0.17	-0.21
840	-0.25	-0.23	-0.21	-0.20	-0.19	-0.28	-0.37

Figure 1: Histogram of Total Usage vs Allowance

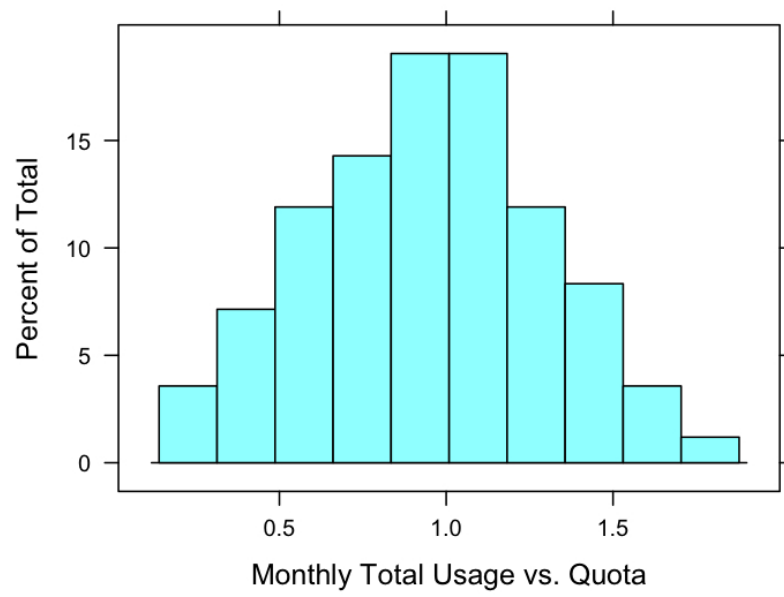


Figure 2: The Effect of Allowance on Minute Usage over Time

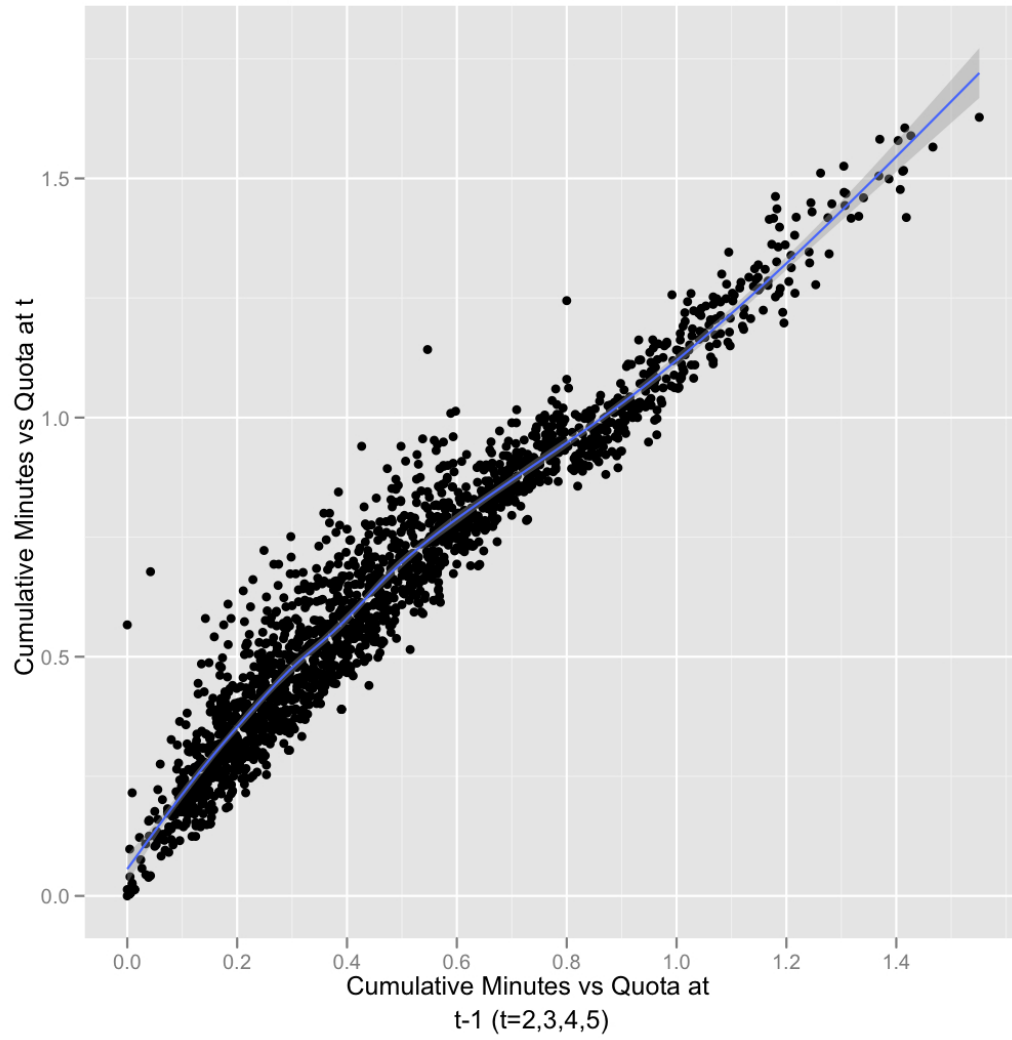


Figure 3: Usage Change over Time

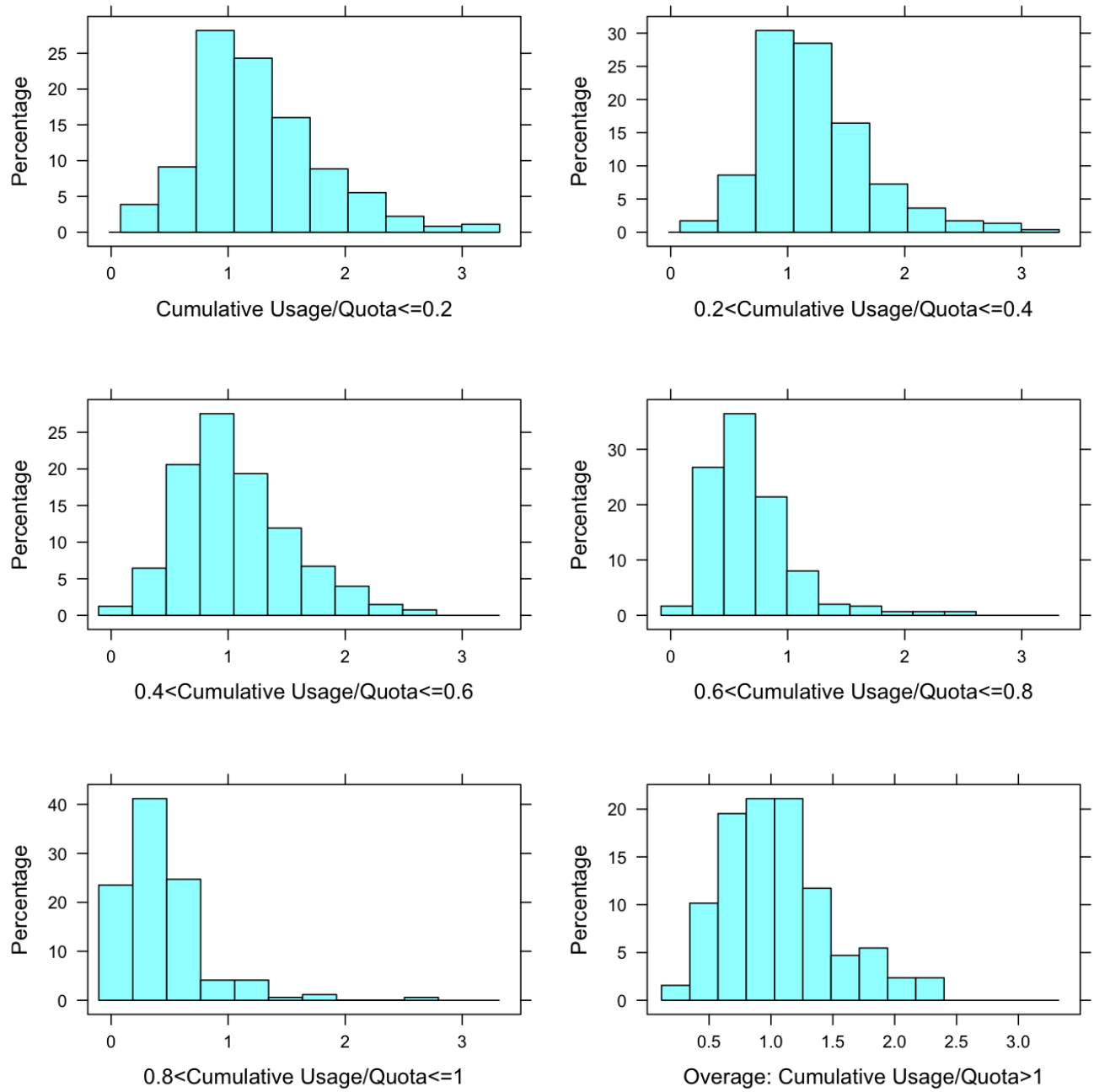
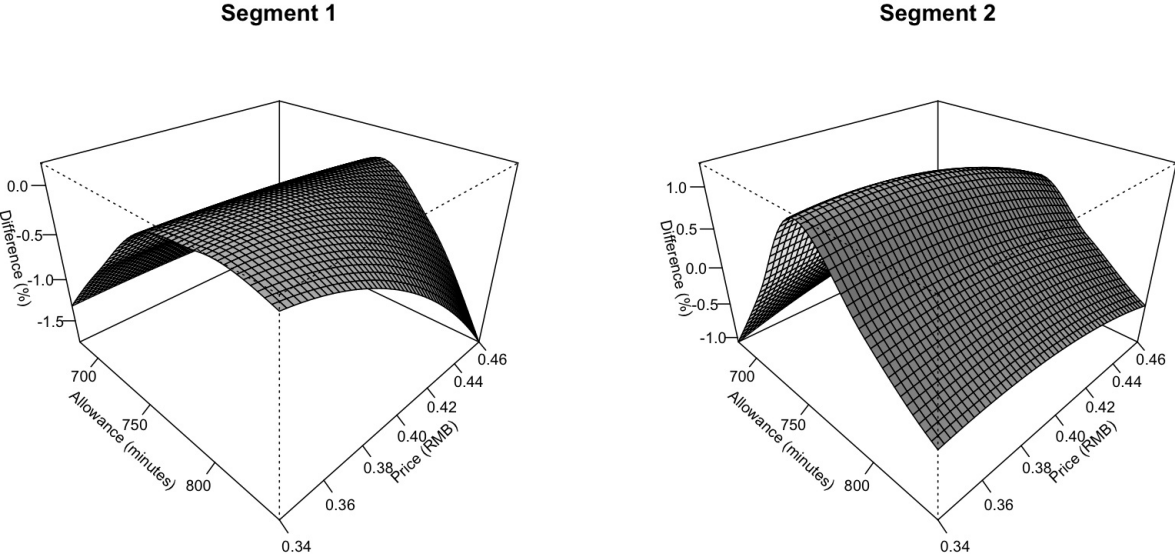


Figure 4: Revenue Prediction Differences: $(\delta_1 = 0.86, \delta_2 = 0.91)$ vs. $(\delta_1 = \delta_2 = 0.995)$



Appendix

A The Distribution of Monthly Minutes Usage $q_{i\tau}$ under Linear Pricing Plan

Since we only observe the monthly minute consumption $q_{i\tau} = \sum_t x_{it}^*$ but not each respective x_{it}^* , we have to find the likelihood of $q_{i\tau}$.

Suppose customer i is a member of preference segment g . As discussed in equation 5, the optimal minutes usage at period t , x_{it}^* , may take two values, 0 (if $d_{it} - b_g p_{i0} \leq 0$) and $d_{it} - b_g p_{i0}$ (if $d_{it} - b_g p_{i0} > 0$). Since $d_{it} = \exp(D'_{it}\alpha_g) + \nu_{it}$ and $\nu_{it} \sim N(0, \zeta_g^2)$, x_{it}^* follows a normal distribution $N(\exp(D'_{it}\alpha_g) - b_g p_{i0}, \zeta_g^2)$ that is truncated at zero. Thus the density of x_{it}^* is

$$f(x_{it}^*) = \frac{1}{\zeta_g} \phi\left(\frac{x_{it}^* - \mu_{it}}{\zeta_g}\right) / [1 - \Phi(-\frac{\mu_{it}}{\zeta_g})] \quad (\text{A1})$$

where $\mu_{it} = \exp(D'_{it}\alpha_g) - b_g p_{i0}$

The monthly minute consumption $q_{i\tau} = \sum_t x_{it}^*$ can then be written as the summation of a series truncated normal random variables with the same truncation at zero. Although there is no closed form for the distribution of $q_{i\tau}$, if the occurrence of zero minute consumption for any period is nearly zero ($\Pr(x_{it}^* > 0) \rightarrow 1, \forall t$), $q_{i\tau}$ can be approximated well by a normal density function that has the mean as $\sum_t \mu_{it}$ and the variance as $T\zeta_g^2$.¹⁶ Intuitively, although x_{it}^* is a truncated normal r.v., if $\Pr(x_{it}^* > 0)$ is nearly one, the truncation becomes moot and the distribution of x_{it}^* can be approximated well by a normal distribution. The summation of a series of i.i.d. normal r.v.'s is also normally distributed.

¹⁶The analytical proof of the validity of this approximation and Monte Carlo simulation results can be obtained from the authors.

We implement a Monte Carlo simulation using the approximation mentioned above. The parameters are recovered with reasonable accuracy. As a robustness check to this approximation, we also use a Kernel estimator to compute the density in the Monte Carlo simulation (Härdle and Linton (1994)). The results are similar to the ones using the approximation but the Kernel estimation is much more computationally demanding.