Identification of Consumer Switching Behavior with Market Level Data

Yong Hyeon Yang*

UCLA

November 18th, 2010

Abstract

Switching costs and the persistence of idiosyncratic preference influence switching behavior of consumers. This paper provides theorems identifying those switching factors using churn rates observed at the market level, which are defined as the rate at which users of a product switch out in the next period. Nonparametric methods are employed to investigate what can be identified about each switching factor from churn rates. Transition in idiosyncratic preference shocks is fully identified, and time-varying product-dependent switching costs are identified. Moreover, the underlying utility function is nonparametrically identified throughout the paper. These results imply that consistent estimation is possible without individual level data. I estimate demand for the U.S. mobile phone service, which illustrates that excluding substantial switching costs from a model may lead to inconsistent estimation.

*I am grateful to my advisors Daniel Ackerberg and Rosa Matzkin for their helpful comments and great inspiration. I have also benefited from conversations with Moshe Buchinsky, Jinyong Hahn, and Connan Snider as well as participants at the industrial organization and econometrics proseminars. All remaining errors are my own. Correspondence: econyang@ucla.edu

1 Introduction

Consumers often purchase the same product repeatedly, either because they have a persistent preference or because switching between products is costly. Discovering the factors that affect consumer switching behavior is important for firms and governments, because misunderstanding them may lead to incorrect policies.[1] Though some of those switching factors have been empirically studied, they have not been formally identified. Intuitively individual level data would identify switching factors, but such data is not readily available.[2] This paper therefore provides theorems identifying switching factors using readily observable market level information on switching patterns.

I show identification using a churn rate, defined as the rate at which consumers who purchased a product in the last period switch out from it in this period. Low churn rates imply that many consumers stay with the same product and therefore signal high switching costs or highly persistent idiosyncratic preferences. Switching costs are the costs incurred by consumers who change products, including explicit fees as well as implicit transaction costs. Consumers who would otherwise switch out of a product may not do so to avoid paying such costs.[3] On the other hand, consumers may buy the same product because they still prefer it most. Heterogeneity across consumers which makes them choose different products can be highly persistent over time, in which case a small number of consumers switch out.

I extend a discrete choice random utility model in two different ways to include switching factors. In each extension, I consider only one switching factor and show what is the most that can be identified about it.[4] The first is the case where unobserved heterogeneity changes over time in a flexible way. Transition in the unobserved heterogeneity is nonparametrically identified with churn rates. It is shown that we can fully recover the transition under some assumptions. I also provide a theorem on its

[1] See Farrell and Klemperer (2007) for a theoretical review.

[2] I do not know any work which has proven identification of switching factors with individual level data. However, the theorem in this paper implies such a case, since market level data can always be constructed using individual level data with loss of information.

[3] If a switching cost is negative, consumers would switch more often. Although negative switching costs are unusual, this paper allows for such a case.

[4] Simultaneously identifying the two factors is not feasible without restrictive assumptions. A discussion on the model with the two factors is provided at the end of Section 4.

partial identification with weaker assumptions. In the literature, unobserved heterogeneity is usually assumed to be either fixed or completely independent over time for an individual, but this paper allows it to evolve in a flexible transition pattern.[5]

A model with switching costs is considered next. Time-varying product-dependent switching costs are identified under the assumption that they enter the utility function in an additively separable way. While there are many types of costs related to switching, I include only quitting costs in the model. This is mainly because modeling quitting costs utilizes the churn rates data most efficiently.[6] The model with quitting costs can approximate those with other switching costs to some extent, in which sense the identification theorem in this paper has implications on those models. It is easily shown that quitting costs are obtained through a contraction mapping when the idiosyncratic preference includes a random variable of the logit distribution.

Switching costs have been modeled frequently in the literature. Erdem, Keane, Öncü, and Strebel (2005) explicitly model how consumers pay different amounts to get different levels of information. Uncertainty and learning costs play the role of switching costs in this case. Osborne (2007) and Cullen and Shcherbakov (2010) count on individual level data to measure switching costs, while Kim (2006) and Shcherbakov (2009) estimate switching costs using market level data. However, these papers do not provide a formal identification result, as is presented here.[7] Shy (2002) illustrates identification of switching costs, but uses a model with two firms producing homogeneous goods. The model presented in this paper allows for more than two differentiated products.

This paper also contributes to the literature on identification of demand in a discrete choice model. A utility function is nonparametrically specified throughout this

[5]Chintagunta, Jain, and Vilcassim (1991) investigate brand switching behaviors with fixed effects in their model, and Gönül and Srinivasan (1993) study brand switching with a random coefficients logit model, where the random coefficients are fixed over time for a household. Sun, Neslin, and Srinivasan (2003) examine switching behaviors of forward looking consumers with fixed effects. These papers use individual level data.

[6]If the switching rate for every pair of products is observed, other types of switching costs can be identified in addition to quitting costs.

[7]Some papers give an intuitive explanation of what features in data identify the factors of interest. See Cullen and Shcherbakov (2010) for example.

paper. In addition, a contemporaneous distribution of unobserved heterogeneity is also assumed to be nonparametric. It is shown that the utility function and the heterogeneity distribution are identified with market level data. Since Hausman and Wise (1978) proposed an estimation method for a parametric demand model with individual level data, many papers have investigated the identification of demand. Ichimura and Thomson (1998), and Bajari, Fox, Kim, and Ryan (2009) show identification of the distribution of random coefficients in a linear utility model. Berry (1994) and Berry, Levinsohn, and Pakes (1995) discuss identification of a parametric demand model using market level data. This paper provides a nonparametric identification result with market level data which is not or only partially considered by the above papers.

Nonparametric identification of demand with market level data has been tried recently. Berry and Haile (2009) nonparametrically identify a demand function and the underlying distribution of the unobserved heterogeneity. However, no switching factor is considered in their model. In addition, they do not fully identify the underlying utility function.[8] Identifying the utility function enables us to do a counterfactual analysis, which would otherwise be limited by partial knowledge of the utility function. I do not allow price to be correlated with unobserved heterogeneity, while Berry and Haile (2009) do. However, the model in this paper is general enough, since the distribution of the unobserved heterogeneity is nonparametrically specified and identified. Hence this paper generalizes the multinomial probit assumption, which generates flexible substitution patterns across products in a market.

Chiappori and Komunjer (2009) show identification of a utility function with combined datasets of market-level and individual-level variables. However, they assume that the unobserved characteristic of products is additively separable in the utility function. I show that a utility function is identified with market level data, even when the utility function is not separable in the unobserved characteristic of products. This approach requires an assumption that the measurement unit of the unobserved characteristic be specified. I follow Matzkin (2010a) and Berry and Haile (2009) for such an assumption. Matzkin (2010b) studies identification of a utility function with individual level data, allowing for flexible extensions of the model, including one to the

[8]In particular, utility is not identified in terms of price, although the distribution of unobserved heterogeneity partially captures it.

discrete choice case. However, she does not consider identification with market level data, which this paper does.

Using data from the mobile phone service industry in the United States, I show that including switching costs changes the estimation result. The empirical model in this paper allows for fully flexible quitting costs, different across products and over time. A parametric utility function is employed due to the small sample size. The assumption of zero switching costs fails to predict churn rates in the data, which suggests that switching costs need to be considered. Including switching costs reduces the estimated price coefficient by 39 percent. This result implies that a model without switching costs yields biased estimates when there are substantial switching costs. Although the model is estimated under a simple logit assumption for the idiosyncratic preference, the implications apply to other specifications as well.

Churn rates are easier to obtain than individual level data as they are free from issues of private information and confidentiality for individual level data. Churn rates are available in industries where a subscription is required to purchase and use a commodity, so that switching patterns are kept track of over time. Examples include mobile phone service, banking, insurance, and cable television and internet providers.[9] Even in other industries without such features, churn rates can be provided whenever purchase histories of consumers are recorded.

The model developed in this paper can be extended in several ways. Knowledge of the evolution of unobserved heterogeneity enables firms to tailor promotional strategies to different types of consumers depending on the persistence of idiosyncratic preference. Introducing type-dependent promotion coupled with persistent idiosyncratic preferences has a dynamic impact on consumer behaviors. Identification of transition in unobserved heterogeneity shown in this paper has implications on such an extension. Another way of extension is modeling transition and switching costs together. Formal identification of such a model is not considered in this paper for its intractability, but its estimation is interesting because it tells us whether switching behaviors originate from state dependence or unobserved heterogeneity.[10] A model of learning with transition

[9]Some of those industries may not provide such information publicly, but it is potentially available.

[10]See Chamberlain (1984) for details on separate identification of state dependence and fixed effects. See also Honorè and Kyriazidou (2000) and Wooldridge (2005) for extensions.

in unobserved heterogeneity is also an interesting extension.

The paper is organized as follows. Section 2 describes the model and introduces necessary assumptions on an idiosyncratic preference. Section 3 discusses nonparametric identification of the transition probabilities of unobserved idiosyncratic preference shocks, and Section 4 focuses on identification of unobserved switching costs. Both sections allow for a nonparametric utility function. In Section 5, I estimate an empirical model of the U.S. mobile phone service industry. Section 6 concludes.

2 Model of Random Utility

A discrete choice model with random utility is considered in this paper. Identification of switching factors in such a model can be applied to many cases because a discrete choice model is widely used in the recent industrial organization literature to analyze industries where consumers buy at most one good in a given period with a very high probability.[11] If the industry is known for high or low frequency of consumer switching patterns, this paper suggests consideration of switching factors in the model.

In each period, consumers choose among J products, and buy good j if the utility from j is greater than that from any other good. The utility individual i gets from commodity j is given by

$$u_{ijt} = u(p_{jt}, z_{jt}, \xi_{jt}, \varepsilon_{ijt}) \tag{1}$$

where t denotes geographic or temporal markets. $p_{jt} \in \mathbb{R}_+$, $z_{jt} \in \mathbb{R}^G$, and $\xi_{jt} \in \mathbb{R}$ stand for price, observed characteristics, and the unobserved characteristic of product j in market t, respectively. There are several ways of interpreting the unobserved characteristic of products in the literature. It can be the quality of the product,[12] a single index of characteristics that are hard to measure, or market level demand shocks to the product.[13] ε_{ijt} captures any residual effect not explained by the above

[11] For example, Berry, Levinsohn, and Pakes (1995) use a discrete choice model in analyzing the automobile market with annual car sales data, since it is very likely that a person buys a car or none in a given year. Hendel and Nevo (2006) examine at the laundry detergent market, where consumers buy more often but the data are observed more frequently as well.

[12] See Bresnahan (1987) and Berry (1994) for example.

[13] See Nevo (2001) for example.

measures. It may be interpreted as individual preference shocks to consumer i for product j in market t, including income level, product loyalty, and individual need for the product. In general ε_{ijt} may be a multi-dimensional variable, but I restrict it to a single dimensional one, that is $\varepsilon_{ijt} \in \mathbb{R}$.

It is common in the literature to define the outside good. In principle it must refer to buying nothing, but sometimes it can be interpreted as buying minor products whose information is not available. Usually it is assumed that[14]

$$u_{i0t} = 0 \tag{2}$$

In the linear utility model, this assumption is equivalent to regarding the outside good as a product with no characteristic and zero price. This plays the role of location normalization since consumers make a pairwise comparison of utilities from all the products in a discrete choice model. I now impose some restrictions on ε_{ijt}.

Assumption 1 (Additive Preference Shock) *The preference shock ε_{ijt} enters the utility function additively.*

$$u_{ijt} = \delta(p_{jt}, z_{jt}, \xi_{jt}) + \varepsilon_{ijt} \tag{3}$$

Let $\varepsilon_{it} = (\varepsilon_{i1t}, \cdots, \varepsilon_{iJt})'$ and denote p_t, z_t and ξ_t similarly.

Assumption 2 (Independence) *The series $\{\varepsilon_{it}\}_{t=1}^{\infty}$ are independent of the series $\{p_t, z_t, \xi_t\}_{t=1}^{\infty}$.*

Assumption 3 (Distribution) *ε_{it} is independent across i, and is identically distributed over i and t. It is absolutely continuously distributed, and the support of the density function is connected. The joint distribution of ε_{it} is denoted by F_{ε}.*

Note that Assumption 1 alone does not impose a restriction on the utility function. For example, the random coefficient model is compatible with Assumption 1. Recall

[14] The utility of the outside good may be different from 0 depending on the model specification. For example, Gowrisankaran and Rysman (2009) let it be the utility of the good bought in the last period.

that the random coefficient utility can be written in the following way:

$$u_{ijt} = p'_{jt}\beta^p + z'_{jt}\beta^z + \xi_{jt} + \underbrace{p'_{jt}\beta_i^p + z'_{jt}\beta_i^z + e_{ijt}}_{\varepsilon_{ijt}}$$

Any general utility function can be rewritten as in (3) as long as ε_{ijt} is dependent on p_{jt}, z_{jt} and ξ_{jt}. It is the assumption of independence that makes Assumption 1 substantial. So Assumption 2 restricts the model in two ways, directly through the distribution of ε_{it} as well as indirectly through Assumption 1.

Although the independence assumption imposes a restriction on the model and is not compatible with the random coefficients model,[15] I still allow for flexible substitution patterns among products, because I do not assume independence of ε_{ijt} across j.[16] Indeed the random coefficients model is often used to comply with flexible substitution patterns across products by modeling unobserved heterogeneity on the coefficient parameters.[17] Another way to allow for such a feature is suggested by Hausman and Wise (1978) who introduce the conditional probit model. Each model has its own advantage. For example, the conditional probit model generates more flexible substitution patterns at the cost of a large number of parameters. Since I study a nonparametric identification of the distribution of ε_{it}, including its transition, I pursue and generalize the conditional probit model.

3 Identification of transition in preference shocks

We restrict our attention to temporal markets, $t = 1, \cdots, T$, with T potentially growing.[18] We are interested in identifying the mean utility function δ and the conditional

[15]The assumptions still allow for random coefficients on the characteristics that do not change across markets, for example, dummies on the product categories.

[16]It is well known that substitution patterns implied by a model are very limited when the model is specified with additive preference shocks independent across products. See Berry (1994) for a discussion.

[17]See Quandt (1966) or McFadden (1976) for a detailed discussion of the random coefficients model.

[18]Alternatively there may be multiple geographic markets $m = 1, \cdots, M$, where M potentially grows. In such a case, we only need $T = 2$ periods for each market. One may count on the theorems in this paper for such a model, since the theorems can be easily converted.

distribution of ε_{it} given $\varepsilon_{i,t-1}$. This section contributes to the identification literature by adding identification of the transition of idiosyncratic preference shocks. Previous works focus on identification of a utility function and the contemporaneous distribution of idiosyncratic preference shocks. Hausman and Wise (1978) propose an estimation method for a parametric utility function in the discrete choice model. Ichimura and Thomson (1998) offer a theorem on identification of a linear utility function with random coefficients in the binary choice model, and Bajari, Fox, Kim, and Ryan (2009) show that the distribution of random coefficients are nonparametrically identified in the multinomial logit model with a linear utility function. While the above three papers consider the model with individual purchase data, Berry (1994) and Berry, Levinsohn, and Pakes (1995) propose an estimation method for a model with aggregated market level data. Berry and Haile (2009) extend the model to one with a nonparametric utility function and discuss its identification.

I show that transition in idiosyncratic preference shocks is identified with market level data, if a market level variable on switching rates is available in addition to market shares and observed characteristics of products. It is relatively easy to see that we can reveal how idiosyncratic preference shocks evolve, given data on individual purchase history. Once the utility function is identified, changes in the purchase pattern show us how the unobserved preference shock evolves over time for an individual, examining purchase patterns for all individuals allows us to trace out the conditional distribution of ε_{it} and $\varepsilon_{i,t-1}$. On the other hand, market level data may not capture the behavior of the same individual over time. Without any information on links between intertemporal purchases, it would be impossible to identify any parameters governing the evolution of the preference shocks.

If we observe a switching rate, however, it tells us how many people switch from one good to other goods. This is obviously a market level variable, but there are several levels of data available for the switching rate. The most detailed would be the complete matrix of switching rates between all products. We may instead observe the matrix of switching rates between groups of products, which is less detailed. An intermediate level is the churn rate for each product. The churn rate tells us how many people who purchased a good in the last period did not purchase the same good in the current period. I count on the churn rate to identify the evolution of idiosyncratic shocks.

Section 3.1 below introduces the identification equations. Section 3.2 addresses non-identifiability and imposes some restrictions on the model. In Sections 3.3 and 3.4, I propose two methods to identify the model.

3.1 Identification equations

In this subsection, I derive two identification equations and make an additional assumption necessary for identification of transition in unobserved heterogeneity. The identification equations define market shares and churn rates. Let us begin with the market share equation. Denote market shares by $s_t = (s_{0t}, \cdots, s_{Jt})'$. Note that $s_t \in \Delta^J \subset \mathbb{R}^{J+1}$, where Δ^J is the simplex of dimension J defined in the $J+1$ dimensional Euclid space. The model described by Assumptions 1, 2, and 3 in the previous section implies that for $j = 1, \cdots, J$,

$$\begin{aligned} s_{jt} &= P(u_{ijt} \geq u_{ikt} \; \forall k) \\ &= P\Big(\delta(p_{jt}, z_{jt}, \xi_{jt}) + \varepsilon_{ijt} \geq \max(\delta(p_{kt}, z_{kt}, \xi_{kt}) + \varepsilon_{ikt}, u_{i0t}) \; \forall k \geq 1\Big) \\ &= P(A_{jt}) \end{aligned} \tag{4}$$

where P is a probability measure of idiosyncratic preference shocks, and $A_{jt} = \{\varepsilon_{it} : \delta(p_{jt}, z_{jt}, \xi_{jt}) + \varepsilon_{ijt} \geq u_{i0t}$ and $\delta(p_{jt}, z_{jt}, \xi_{jt}) + \varepsilon_{ijt} \geq \delta(p_{kt}, z_{kt}, \xi_{kt}) + \varepsilon_{ikt}$ for all $k \geq 1\}$ represents the event that product j is preferred to all the other products in market t. Note that A_{jt} depends on (p_t, z_t, ξ_t) only through $\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt})$. Assumptions 2 and 3 imply that P is independent of (p_t, z_t, ξ_t) and that P is identical across t. Therefore $P(A_{jt})$ is a function of $\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt})$, u_{i0t} and F_ε only. Define the function $\sigma_j : \mathbb{R}^J \to [0, 1]$ so that

$$s_{jt} = P(A_{jt}) =: \sigma_j\Big(\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt}); u_{i0t}, F_\varepsilon\Big)$$

Stack the equations and rewrite as

$$s_t = \sigma\Big(\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt}); u_{i0t}, F_\varepsilon\Big) \tag{5}$$

where $\sigma : \mathbb{R}^J \to \Delta^J$. This equation is used to identify the utility function δ and the distribution F_ε of ε_{it}. Denote $\delta_{jt} = \delta(p_{jt}, z_{jt}, \xi_{jt})$ for simplicity, and $\delta_t = (\delta_{1t}, \cdots, \delta_{Jt})'$; then we can write $s_t = \sigma(\delta_t; u_{i0t}, F_\varepsilon)$. It is obvious from (4) that

$$\sigma(\delta_t + c; u_{i0t} + c, F_\varepsilon) = \sigma(\delta_t; u_{i0t}, F_\varepsilon)$$

for any $c \in \mathbb{R}$, and thus I normalize at $u_{i0t} = 0$ as in (2). u_{i0t} and F_ε are included in the arguments to explicitly show dependence of σ on those, but I will sometimes omit either of them or both for simplicity.

Note that there is a one-to-one relationship between F_ε and σ. It implies that once we identify σ, we can find the distribution F_ε of ε_{it} whether δ is identified or not. It is obvious from (4) and (5) that σ is uniquely determined given F_ε. The following lemma shows that the converse is also true.[19]

Lemma 1 *Every σ is associated with only one F_ε.*

Proof. We need to show that the function F_ε is uniquely determined given σ. Take any $e \in \mathbb{R}^J$. Let $\delta = -e$. Now observe that

$$\begin{aligned} F_\varepsilon(e) &= P(\varepsilon \leq e) \\ &= P(\varepsilon_1 \leq -\delta_1, \cdots, \varepsilon_J \leq -\delta_J) \\ &= P(\delta_j + \varepsilon_j \leq 0 \; \forall j \geq 1) \\ &= \sigma_0(\delta_1, \cdots, \delta_J) \\ &= \sigma_0(-e) \end{aligned}$$

where the second last equality holds since $P(\delta_j + \varepsilon_j \leq 0 \; \forall j \geq 1)$ is the probability that the utility of buying nothing is greater than that of all the other goods, so must be equal to the share of the outside good when the mean utilities are given by $\delta_1, \cdots, \delta_J$, which is $\sigma_0(\delta_1, \cdots, \delta_J)$. Thus when we know the function σ completely, we can trace out F_ε. ∎

The next identification equation uses the churn rate. Let h_{jt} be the rate at which people who bought product j at time $t-1$ switch out of product j at time t. It is therefore defined as follows:

$$h_{jt} = P(\exists k \neq j \text{ such that } u_{ijt} < u_{ikt} \mid u_{ij,t-1} \geq u_{ik,t-1} \; \forall k)$$

This could be rewritten as

$$1 - h_{jt} = P(u_{ijt} \geq u_{ikt} \; \forall k \mid u_{ij,t-1} \geq u_{ik,t-1} \; \forall k) \tag{6}$$

[19]Berry and Haile (2009) employ the same idea in the proof of their Theorem 2.

where P is now a probability measure of idiosyncratic preference shocks over two periods. In order to have enough observations for identification I assume that P does not vary over time. Denote the joint distribution of ε_{it} and $\varepsilon_{i,t-1}$ by $F_\varepsilon^{t,t-1}$.

Assumption 4 (Invariance) *$F_\varepsilon^{t,t-1}$ is the same over t.*

Under Assumption 4, P is invariant over time, and thus (6) can be used to identify P, or equivalently $F_\varepsilon^{t,t-1}$. Assumption 4 also implies that the marginal distribution F_ε^t of ε_{it} is the same with F_ε^{t-1} of $\varepsilon_{i,t-1}$, since when $F_\varepsilon^{t,t-1}$ is the same with $F_\varepsilon^{t+1,t}$, their marginal distributions also have to coincide. Therefore this assumption is consistent with Assumption 3.

3.2 Non-identification and normalization

The model is not identified using the identification equations and the assumptions described so far. This subsection shows that some normalization assumptions are necessary for identification. It helps identify the model to define a more general form of the share function as follows:

$$S(p_t, z_t, \xi_t; \delta, F_\varepsilon) := \sigma(\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt}); F_\varepsilon) \tag{7}$$

Note that $S : \mathbb{R}_+^J \times \mathbb{R}^{J(G+1)} \to \Delta^J$ implicitly depends on δ and F_ε and that S is uniquely determined given δ and F_ε. We may write $s_t = S(p_t, z_t, \xi_t)$ for simplicity. Note also that there is no one-to-one relationship between S and F_ε as there is between σ and F_ε. This becomes apparent after we prove non-identification of the model (5) below.

In (5), the observables are $\{s_t, p_t, z_t\}_{t=1}^T$, while $\{\xi_t\}_{t=1}^T$ and the functions δ, σ and F_ε are unobserved. The model is not identified. To see this, simply fix the series $\{\xi_t\}_{t=1}^T$, and consider any strictly increasing and invertible $H : \mathbb{R} \to \mathbb{R}$. Let

$$\tilde{\delta} = H \circ \delta$$
$$\tilde{\sigma}(\delta_t) = \sigma(H^{-1}(\delta_{1t}), \cdots, H^{-1}(\delta_{Jt}))$$

and let $F_{\tilde{\varepsilon}}$ defined correspondingly by Lemma 1. Then

$$\begin{aligned}\tilde{\sigma}(\tilde{\delta}(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \tilde{\delta}(p_{Jt}, z_{Jt}, \xi_{Jt}))& \\ =&\sigma(H^{-1}(\tilde{\delta}(p_{1t}, z_{1t}, \xi_{1t})), \cdots, H^{-1}(\tilde{\delta}(p_{Jt}, z_{Jt}, \xi_{Jt}))) \\ =&\sigma(\delta(p_{1t}, z_{1t}, \xi_{1t})), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt}))\end{aligned}$$

implies that both $(\delta, \sigma, F_\varepsilon)$ and $(\tilde{\delta}, \tilde{\sigma}, F_{\tilde{\varepsilon}})$ satisfy (5), and thus they are observationally equivalent.[20] To put a different way, $\tilde{\delta}$ and $\tilde{\varepsilon}$ with a distribution function $F_{\tilde{\varepsilon}}$ generate the observationally equivalent equation. To find a way to normalize the functions, I derive an equation that shows the relationship between F_ε and $F_{\tilde{\varepsilon}}$ as follows. The proof of Lemma 1 establishes that $F_\varepsilon = \sigma_0(-e)$ for any $e \in \mathbb{R}^J$, which implies that

$$\begin{aligned}P(\tilde{\varepsilon} \leq e) &= F_{\tilde{\varepsilon}}(e) \\ &= \tilde{\sigma}_0(-e) \\ &= \sigma_0(H^{-1}(-e_1), \cdots, H^{-1}(-e_J)) \\ &= F_\varepsilon(-H^{-1}(-e_1), \cdots, -H^{-1}(-e_J)) \\ &= P(\varepsilon_1 \leq -H^{-1}(-e_1), \cdots, \varepsilon_J \leq -H^{-1}(-e_J)) \\ &= P(-\varepsilon_1 \geq H^{-1}(-e_1), \cdots, -\varepsilon_J \geq H^{-1}(-e_J)) \\ &= P(H(-\varepsilon_1) \geq -e_1, \cdots, H(-\varepsilon_J) \geq -e_J) \\ &= P(-H(-\varepsilon_1) \leq e_1, \cdots, -H(-\varepsilon_J) \leq e_J)\end{aligned}$$

This equation says that if $\tilde{\varepsilon}_{it}$ is defined by $\tilde{\varepsilon}_{ijt} := -H(-\varepsilon_{ijt})$, and the utility function is given by

$$\tilde{u}_{ijt} = \tilde{\delta}(p_{jt}, z_{jt}, \xi_{jt}) + \tilde{\varepsilon}_{ijt} = H(\delta(p_{jt}, z_{jt}, \xi_{jt})) - H(-\varepsilon_{ijt})$$

the observed market shares would be

$$\begin{aligned}s_t &= \tilde{\sigma}(\tilde{\delta}(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \tilde{\delta}(p_{Jt}, z_{Jt}, \xi_{Jt})) \\ &= \sigma(\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt}))\end{aligned}$$

which is the same as those when the utility function is given by

$$u_{ijt} = \delta(p_{jt}, z_{jt}, \xi_{jt}) + \varepsilon_{ijt}$$

[20] This example shows why there is no one-to-one relationship between S and F_ε. Both F_ε and $F_{\tilde{\varepsilon}}$ are compatible with the same function S to generate observationally equivalent equation $s_t = S(p_t, z_t, \xi_t)$.

The observational equivalence between (δ, ε) and $(\tilde{\delta}, \tilde{\varepsilon})$ leads us to make a normalization on either of the two. I make the following normalization on the distribution function of ε_{it}.

Assumption 5 (Normalization) *The joint distribution F_ε is such that the marginal distribution of ε_{i1t} is known.*

Remark 1 *We can also make a normalization assumption on the functional form of δ, by placing a restriction on δ so that any $\tilde{\delta} = H(\delta)$ with a strictly increasing H violates it. A sufficient condition under which $(\delta, \sigma, F_\varepsilon)$ is uniquely identified given $\{\xi_t\}_{t=1}^T$ is that there exists a set $B \subset \mathbb{R}^G \times \mathbb{R}_+ \times \mathbb{R}$ satisfying the following two conditions. $\delta(z, p, \xi)$ is known for all $(z, p, \xi) \in B$, and the image of B is the same as the image of the whole domain, or in other words, $\delta(B) = \delta(\mathbb{R}^G \times \mathbb{R}_+ \times \mathbb{R})$. Then choosing B is a normalization assumption. If we assume $\delta(\overline{p}, \overline{z}, \xi) = \xi$ for some $(\overline{p}, \overline{z})$, for example, $B := \{(\overline{p}, \overline{z}, \xi) : \xi \in \mathbb{R}\}$ would satisfy the above condition and uniquely identify $(\delta, \sigma, F_\varepsilon)$. This is a strong assumption to make when we want to know what kind of assumption is sufficient to uniquely identify $(\delta, \sigma, F_\varepsilon)$. On the other hand, making as weak an assumption on δ as possible leaves the assumption unintuitive or too complicated.*

Given the above five assumptions on idiosyncratic preference shocks, we are able to identify $(\delta, \sigma, F_\varepsilon)$ from (5) given $\{\xi_t\}_{t=1}^T$. But the model is still not identified without information on $\{\xi_t\}_{t=1}^T$. There are two issues regarding the unobserved characteristics. One relates to their unit of measurement. Given $(\delta, \sigma, F_\varepsilon, \{\xi_t\}_{t=1}^T)$, we can always find some $\tilde{\sigma}$ such that for some strictly increasing and invertible $H : \mathbb{R} \to \mathbb{R}$,

$$\begin{aligned}&\tilde{\sigma}(\delta(p_{1t}, z_{1t}, H(\xi_{1t})), \cdots, \delta(p_{Jt}, z_{Jt}, H(\xi_{Jt})))\\&\qquad = \sigma(\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt}))\end{aligned}$$

Define $\tilde{\xi}_{jt} := H(\xi_{jt})$ and let $F_{\tilde{\varepsilon}}$ defined by $\tilde{\sigma}$ from Lemma 1, then $(\delta, \sigma, F_\varepsilon, \{\xi_t\}_{t=1}^T)$ is observationally equivalent to $(\delta, \tilde{\sigma}, F_{\tilde{\varepsilon}}, \{\tilde{\xi}_t\}_{t=1}^T)$.

The other issue arises from the endogeneity of ξ_t in the model (5). Since we interpret ξ_t as a characteristic of products, it is likely that they are correlated with product prices. I will take two different approaches to correct for endogeneity in the following two subsections. One approach is to use instrumental variables. Its application is easy, but it is difficult to find instruments satisfying the identification assumption in a nonparametric model. When the utility function is specified parametrically, the condition for identification can usually be written in terms of the rank of a matrix. It is thus easy to check validity of instruments. In a nonparametric model, however, the identification assumption with instrumental variables is not easy to check.

The second approach is to add a set of equations from which prices are determined. This is more complete than the first approach, and more efficient if the additional equations are specified correctly. Thus there is no reason why we do not add equations if we can. It has some flaws on the other hand. If the additional equations are misspecified, the estimates might be inconsistent. Moreover, this approach may require some additional variables used for the new set of equations. When we do not have enough number of exogenous variables which directly enter the new set of equations,[21] the former method would be useful.

3.3 Using demand model only

In this subsection, I use only the models (5) and (6) and take an instrumental variables approach to address the endogeneity between price and the unobserved characteristic of products. Assume $z_{jt} \in \mathbb{R}$ for simplicity. For the general case $z_{jt} \in \mathbb{R}^G$, we only need to assume that the additional $G-1$ variables are independent of the unobserved characteristics across all products within a market. As mentioned in Section 3.2, the measurement unit of the unobserved characteristics has to be specified. I follow Berry and Haile (2009) for such an assumption.

Assumption 6 (Measure of Unobserved Characteristic)

$$\delta(p_{jt}, z_{jt}, \xi_{jt}) = \delta(p_{jt}, z_{jt} + \xi_{jt}) \tag{8}$$

and δ is strictly increasing in its second argument.

[21] We need at least the same number of exogenous variables as that of unobserved endogenous variables. As adding the new set of equations often increases the number of unobserved endogenous variables, we may need more exogenous variables to do so.

This assumption specifies the measurement unit of ξ_{jt} but also restricts the functional form of δ. Although it is not the weakest assumption required to identify ξ_{jt}, it has a clearer interpretation. I abuse a notation and use δ for the new mean utility function again. Since σ is invertible, we can write the equation (5) as

$$\sigma^{-1}(s_t) = \Big(\delta(p_{1t}, z_{1t} + \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt} + \xi_{Jt})\Big)'$$

Taking element-wise inversion with respect to the second argument of δ yields

$$z_{jt} + \xi_{jt} = \delta^{-1}\Big([\sigma^{-1}(s_t)]_j, p_{jt}\Big)$$

Defining $\gamma_j(s_t, p_{jt}) := \delta^{-1}([\sigma^{-1}(s_t)]_j, p_{jt})$, we have

$$z_{jt} + \xi_{jt} = \gamma_j(s_t, p_{jt}) \tag{9}$$

The series ξ_{jt} and γ are unknown in the above equation. Since p_{jt} is endogenous with the unobservable ξ_{jt}, instrumental variables are necessary to identify γ. Let w_{jt} denote the set of variables including all the exogenous and the instrumental variables. Importantly, z_{jt} is included in w_{jt}. Then I make the following identification assumptions.

Assumption 7 (Strong Instrument) *There exists w_{jt} which includes z_{jt} in it such that for any function $B : \Delta^J \times \mathbb{R}_+ \to \mathbb{R}$ with a finite expectation, $E[B(s_t, p_{jt})|w_{jt}] = 0$ a.e. implies $B(s_t, p_{jt}) = 0$ a.e.*

Assumption 8 (Mean Independence) $E[\xi_{jt}|w_{jt}] = 0$

Assumptions 7 and 8 specify the conditions which must be satisfied by the instrumental variables. Assumption 7 is made in accordance with Newey and Powell (2003), and Berry and Haile (2009). It is strong because it requires that we need instrumental variables with nearly perfect information on the movement of the price as well as the share. Intuitively, given some instrumental variables, we still have to observe a lot of joint variation in the price and the share on a large support. In addition, changes in instrumental variables must be associated with changes in the distribution of the price and the share in a nontrivial way. Its counterpart in a linear model is that instruments must be non-orthogonal to endogenous variables.

Newey and Powell (2003) provide some theorems by which this condition holds. For example, if the price, in our context, is distributed normally conditional on some instruments, the identification assumption can be converted into a rank condition. Another example is the case where both the price and the instruments have finite support and satisfy some rank condition. We can see a similar but weaker assumption in Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007), and Berry and Haile (2010). The last assumption plays the role of location normalization of ξ_{jt} in addition. In conjunction with Assumption 6, it restricts the way unobserved characteristics are distributed.

Let E_j be the support of the marginal distribution of ε_{ijt}. From the identification equations and the assumptions described so far, we have the following theorem.

Theorem 1 *Suppose Assumptions 1-8 hold. Assume further that the data* (s_t, h_t, p_t, z_t) *are available for all* $s_t \in \Delta^J$*. Then the unobserved characteristic* $\{\xi_t\}_{t=1}^T$ *is identified, and the utility function* $\delta(p_{jt}, z_{jt} + \xi_{jt})$ *is nonparametrically identified on the inverse image of* $-E_1$*. Also the joint density function* $f_\varepsilon^{t,t-1}$ *of* ε_{it} *and* $\varepsilon_{i,t-1}$ *is identified on* E_1^{2J}*. Especially,* $F_\varepsilon^{t,t-1}$ *is fully identified if* E_1 *contains* E_j *for all* j*.*

Proof. Recall Equation (9).

$$z_{jt} + \xi_{jt} = \gamma_j(s_t, p_{jt})$$

By taking a conditional expectation on both sides we get

$$E[z_{jt} + \xi_{jt}|w_{jt}] = E[\gamma_j(s_t, p_{jt})|w_{jt}]$$

$E[z_{jt}|w_{jt}] = z_{jt}$ since w_{jt} includes z_{jt}. Then it follows from Assumption 8 that

$$z_{jt} = E[\gamma_j(s_t, p_{jt})|w_{jt}]$$

Assumption 7 implies that $\gamma_j(s_t, p_{jt})$ is identified from the above equation.[22] Once γ_j is identified for all j, (9) identifies ξ_t. So we can identify S from $s_t = S(p_t, z_t, \xi_t)$ for all (p_t, z_t, ξ_t). Since the conditions of Lemma 2 are satisfied, the utility function $\delta(p_{jt}, z_{jt} + \xi_{jt})$ is identified on the inverse image of $-E_1$, and the joint distribution F_ε is identified on E_1^J. Also since the conditions of Lemma 3 are satisfied, it follows that the

[22] See Berry and Haile (2009) for the proof of identification using this assumption.

joint density function $f_\varepsilon^{t,t-1}$ of ε_{it} and $\varepsilon_{i,t-1}$ is identified on E_1^{2J}. If E_1 contains E_j for all j, $f_\varepsilon^{t,t-1}$ is fully identified since E_1^{2J} contains the whole support of $f_\varepsilon^{t,t-1}$. Therefore, $F_\varepsilon^{t,t-1}(e)$ can be obtained by integrating $f_\varepsilon^{t,t-1}(e)$ up to e. ∎

While the formal statements of Lemmas 2 and 3 and their proofs are given in the Appendix, an illustrative example is offered here. Suppose we have only one good with characteristics (p_t, z_t, ξ_t). Consumers would buy it if $\delta(p_t, z_t + \xi_t) + \varepsilon_{it} \geq 0$, and not buy otherwise. The share would be determined by

$$s_t = \Pr(\delta(p_t, z_t + \xi_t) + \varepsilon_{it} \geq 0) = 1 - F_\varepsilon(-\delta(p_t, z_t + \xi_t))$$

Since $F_\varepsilon(\cdot)$ is known, we have

$$-F_\varepsilon^{-1}(1 - s_t) = \delta(p_t, z_t + \xi_t) \tag{10}$$

If p_t were independent of ξ_t, we would have been able to identify δ and ξ_t without Assumption 7 using the above equation. To account for endogeneity of ξ_t we must take the following steps. Since δ is strictly increasing in its second argument, we can take an inversion of δ with its second argument and write

$$z_t + \xi_t = \delta^{-1}(-F_\varepsilon^{-1}(1 - s_t), p_t) =: \gamma(s_t, p_t) \tag{11}$$

Taking a conditional expectation and appealing to Assumption 7 leads to identification of γ. The series ξ_t can be identified from Equation (11), and δ from Equation (10). Now the joint distribution of ε_{it} and $\varepsilon_{i,t-1}$ is identified from the churn rate.

$$h_t = P(\delta(p_t, z_t + \xi_t) + \varepsilon_{it} \leq 0 | \delta(p_{t-1}, z_{t-1} + \xi_{t-1}) + \varepsilon_{i,t-1} \geq 0)$$

Arranging the equation yields

$$h_t s_{t-1} = P(\varepsilon_{it} \leq -\delta(p_t, z_t + \xi_t) \text{ and } -\varepsilon_{i,t-1} \leq \delta(p_{t-1}, z_{t-1} + \xi_{t-1}))$$

Since δ and ξ_t are already identified, the above equation tells us the joint distribution of ε_{it} and $\varepsilon_{i,t-1}$.

The above example shows which set the utility function δ is identified on. The left hand side of (10) ranges on $-E$, the mirror image of the support of ε_{it}. So the utility function δ is identified only at the points which are mapped into $-E$ by δ, or simply on the inverse image of $-E$. Extending such an idea to a general case of $J \geq 2$, δ is identified on the inverse image of $-E_1$ from the known marginal distribution of ε_{i1t}. When E_j is the same across j, this result leads to full identification of the distribution of ε_{it} as well as its transition. They might be partially identified if E_j is different across j, but there is a quick remedy to such a case as follows. If the mean of ε_{ijt} is different across j, we can make the means identical by including additive product-specific dummy variables in the utility function. If the spread of ε_{ijt} is different across j, we can take the product with the largest support and rename it to product 1. Such a product can be found by observing the distribution of shares since its share would be more dispersed than that of other products. Then E_1 contains E_j for all j now, and Theorem 1 guarantees that the distribution of ε_{it} and its transition are fully identified.

3.4 Adding supply side model

This subsection uses a simultaneous equations model to address the endogeneity issue and identify the model. Recall that Assumptions 1-3 imply the following equations on the demand side.

$$s_t = S(p_t, z_t, \xi_t) = \sigma(\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt})) \tag{12}$$

The market shares are determined given p_t, z_t and ξ_t. But the prices are set by firms considering all the characteristics of the products on the market, so we are concerned that p_t must be correlated with ξ_t. I add equations specifying the price decision, and formulate a simultaneous equations model. The additional equations give us full information about how the price is correlated with the unobserved characteristics of products, and thus eliminate implicitness of endogeneity. If the model is correctly specified, this method would be more efficient than the previous one-side model. I still assume that $z_{jt} \in \mathbb{R}$. The equilibrium prices can be determined in general by

$$p_t = \rho(z_t, \xi_t, w_t, \eta_t) \tag{13}$$

where w_t and η_t are vectors of observed and unobserved cost factors, respectively. I assume $w_{jt} \in \mathbb{R}$ and $\eta_{jt} \in \mathbb{R}$, but the model can be easily extended to a more general case with $w_{jt} \in \mathbb{R}^K$.

There are a couple of remarks that need to be mentioned about the specification in (13). It is less structural than that on the demand side. For a more structural way, we may define a cost function and specify a form of the game firms are playing.[23] However I focus on the identification of consumer's switching behavior on the demand side, rather than on that of cost functions. So I employ a more flexible specification to comply with a broad set of cost functions and forms of the competition.

By the above specification, I assume that there exists a set of equilibrium prices given z_t, ξ_t, w_t, and η_t. Each firm has its own equation for price decision to maximize its profit. We have J best response equations and J prices, so the assumption of the existence of equilibrium prices is reasonable. Note that the equilibrium price p_{jt} depends on the characteristics and cost factors of other products as well as its own. When setting price, a firm considers the characteristics and price of other products because they will affect demand for its own product through (12). Therefore every best response equation is written in terms of all the prices $p_{1t}, \cdots, p_{Jt}$, and the solution p_t to the system of equations depends on the characteristics and cost factors of all products, not just its own.

A unique equilibrium is not guaranteed, but is also not required in this analysis. It suffices that there exists a function satisfying (13). However we need a condition that the same equilibrium price function is used over time. This does not only require that the same equilibrium prices are chosen when the firms face the same z_t, ξ_t, w_t, and η_t, but also that the type of the game firms are playing does not vary over time. For example, suppose firms chose their price and quantity via the Cournot Nash game in the beginning of the industry. If one of the firms became larger than the others for some reason, and they switched to the Stackelberg leader game at some point, the condition of an invariant form of competition fails and (13) may not be used. In such

[23]For example, assume that a cost function is given by $C(s_{jt}, z_{jt}, \xi_{jt}, w_{jt}, \eta_{jt})$ and that firms compete by price. A first order condition is

$$s_{jt} + [p_{jt} - C'(s_{jt}, z_{jt}, \xi_{jt}, w_{jt}, \eta_{jt})] \frac{\partial s_{jt}}{\partial p_{jt}} = 0$$

Since s_{jt} depends on p_t, z_t and ξ_t, a profit maximizing price is $p_{jt} = \psi(p_{-jt}, z_t, \xi_t, w_{jt}, \eta_{jt})$. If this function is used instead of (13), a less restrictive assumption on ψ is required for identification than is imposed on ρ later. Apparently, one needs to choose between making assumptions on the structure or on the resulting function.

a case, if we have information as to when the competition scheme changes, we may use a more structural model with a specific form of the competition scheme.

Following the approach in Matzkin (2008, 2010a), we can express the unobserved terms as a function of observed variables by inverting Equations (12) and (13). Some assumptions on the functions are required for invertibility. Abuse a notation and let δ be the vector argument of the function σ, so δ is a $J \times 1$ vector when it is written without the arguments $(p_{jt}, z_{jt}, \xi_{jt})$. Let σ_j denote the j-th coordinate of the function σ, and δ_j denote the j-th argument of the function σ.

Assumption 9 *The utility function $\delta(p_{jt}, z_{jt}, \xi_{jt})$ is differentiable and such that*

$$\frac{\partial \delta}{\partial z_{jt}}(p_{jt}, z_{jt}, \xi_{jt}) = \frac{\partial \delta}{\partial \xi_{jt}}(p_{jt}, z_{jt}, \xi_{jt}) > 0$$

for all $(p_{jt}, z_{jt}, \xi_{jt})$ and all j, and the matrix

$$\frac{\partial \sigma}{\partial \delta'}(\delta) := \begin{pmatrix} \frac{\partial \sigma_1}{\partial \delta_1}(\delta) & \cdots & \frac{\partial \sigma_1}{\partial \delta_J}(\delta) \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_J}{\partial \delta_1}(\delta) & \cdots & \frac{\partial \sigma_J}{\partial \delta_J}(\delta) \end{pmatrix}$$

is positive definite for all $\delta \in \mathbb{R}^J$.

The differentiability of the share function $\sigma(\delta_t)$ is guaranteed by Assumption 3 that the density of ε_{it} is absolutely continuous. The first part of Assumption 9 specifies the measure of the unobserved characteristic as Assumption 6 does. The second part requires that changes in the characteristics and price of a product must have a higher impact on its own share, than on the share of the other products. To see this, consider the 2-good example. The matrix becomes

$$\begin{pmatrix} \frac{\partial \sigma_1}{\partial \delta_1}(\delta) & \frac{\partial \sigma_1}{\partial \delta_2}(\delta) \\ \frac{\partial \sigma_2}{\partial \delta_1}(\delta) & \frac{\partial \sigma_2}{\partial \delta_2}(\delta) \end{pmatrix}$$

For this to be a positive definite matrix, it has to be the case that

$$\frac{\partial \sigma_1}{\partial \delta_1}(\delta) > 0$$

$$\frac{\partial \sigma_2}{\partial \delta_2}(\delta) > 0$$

$$\frac{\partial \sigma_1}{\partial \delta_1}(\delta)\frac{\partial \sigma_2}{\partial \delta_2}(\delta) - \frac{\partial \sigma_1}{\partial \delta_2}(\delta)\frac{\partial \sigma_2}{\partial \delta_1}(\delta) > 0$$

The first and second lines follow immediately from Assumption 3 and Equations (4) and (5), which also imply that $\frac{\partial \sigma_2}{\partial \delta_1}(\delta) < 0$ and $\frac{\partial \sigma_1}{\partial \delta_2}(\delta) < 0$. The last line holds if $\frac{\partial \sigma_1}{\partial \delta_1}(\delta) > -\frac{\partial \sigma_2}{\partial \delta_1}(\delta)$ and $\frac{\partial \sigma_2}{\partial \delta_2}(\delta) > -\frac{\partial \sigma_1}{\partial \delta_2}(\delta)$, which means that changes in the characteristics and the price of good 1 move its share more than the share of good 2, and vice versa. This holds trivially when there are two products, since the share of the outside good would decrease when there is an increase in δ_1 with δ_2 unchanged.

As the number of products increases, we need more inequalities to hold, but the logic is similar. When there are many products, it is likely that there is less competition between any two products, and thus changes in the characteristics of a product may have less impact on the share of each of the other products. A different substitution pattern may appear if $\frac{\partial \sigma_j}{\partial \delta_j}(\delta) < -\frac{\partial \sigma_k}{\partial \delta_j}(\delta)$. Assumption 9 allows for such a possibility as long as it is not different too much from an usual pattern in the sense that $\frac{\partial \sigma_j}{\partial \delta_j}(\delta)\frac{\partial \sigma_k}{\partial \delta_k}(\delta) - \frac{\partial \sigma_j}{\partial \delta_k}(\delta)\frac{\partial \sigma_k}{\partial \delta_j}(\delta) > 0$ still holds. This ensures that every principal minor of size 2 has a positive determinant. Conditions for a bigger size of principal minors are more complicated, but are not strong since the diagonal term is likely bigger in the absolute value than the off-diagonal terms.

Given the second part of Assumption 9, we can invert the system of equations (12) in terms of ξ_t. To see this, note

$$\frac{\partial S}{\partial \xi_t'} = \begin{pmatrix} \frac{\partial \sigma_1(\delta)}{\partial \delta_1}\frac{\partial \delta(p_{1t},z_{1t},\xi_{1t})}{\partial \xi_{1t}} & \cdots & \frac{\partial \sigma_1(\delta)}{\partial \delta_J}\frac{\partial \delta(p_{Jt},z_{Jt},\xi_{Jt})}{\partial \xi_{Jt}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_J(\delta)}{\partial \delta_1}\frac{\partial \delta(p_{1t},z_{1t},\xi_{1t})}{\partial \xi_{1t}} & \cdots & \frac{\partial \sigma_J(\delta)}{\partial \delta_J}\frac{\partial \delta(p_{Jt},z_{Jt},\xi_{Jt})}{\partial \xi_{Jt}} \end{pmatrix}$$

and thus

$$\left|\frac{\partial S}{\partial \xi_t'}\right| = \begin{vmatrix} \frac{\partial \sigma_1}{\partial \delta_1}(\delta) & \cdots & \frac{\partial \sigma_1}{\partial \delta_J}(\delta) \\ \vdots & \ddots & \vdots \\ \frac{\partial \sigma_J}{\partial \delta_1}(\delta) & \cdots & \frac{\partial \sigma_J}{\partial \delta_J}(\delta) \end{vmatrix} \prod_{j=1}^{J} \frac{\partial \delta(p_{jt}, z_{jt}, \xi_{jt})}{\partial \xi_{jt}}$$

which is positive for all (p_t, z_t, ξ_t) by Assumption 9. Therefore S is invertible in ξ_t and we can write

$$\xi_t = \xi(s_t, p_t, z_t) \tag{14}$$

Moreover, it follows that $\frac{\partial \xi}{\partial s_t'}$ is also a positive definite matrix, which implies $\frac{\partial \xi_{jt}}{\partial s_{jt}} > 0$ for all j.

The first part of Assumption 9 imposes a restriction on the structural inverse function ξ defined in (14) as follows. Consider the identity

$$s_t = S(p_t, z_t, \xi(s_t, p_t, z_t))$$

Take a derivative of the above equation with respect to z_t at a given (s_t, p_t, z_t),

$$0 = \frac{\partial S}{\partial z_t'}(p_t, z_t, \xi_t) + \frac{\partial S}{\partial \xi_t'}(p_t, z_t, \xi_t)\frac{\partial \xi}{\partial z_t'}(s_t, p_t, z_t)$$

and thus

$$\frac{\partial \xi}{\partial z_t'}(s_t, p_t, z_t) = -\left[\frac{\partial S}{\partial \xi_t'}(p_t, z_t, \xi_t)\right]^{-1}\frac{\partial S}{\partial z_t'}(p_t, z_t, \xi_t)$$

Assumption 9 implies that

$$\frac{\partial S}{\partial z_{jt}}(p_t, z_t, \xi_t) = \frac{\partial \sigma(\delta)}{\partial \delta_j}\frac{\partial \delta(p_{jt}, z_{jt}, \xi_{jt})}{\partial z_{jt}} = \frac{\partial \sigma(\delta)}{\partial \delta_j}\frac{\partial \delta(p_{jt}, z_{jt}, \xi_{jt})}{\partial \xi_{jt}} = \frac{\partial S}{\partial \xi_{jt}}(p_t, z_t, \xi_t)$$

and thus that $\frac{\partial S}{\partial z_t'} = \frac{\partial S}{\partial \xi_t'}$. Therefore we have

$$\frac{\partial \xi}{\partial z_t'}(s_t, p_t, z_t) = -I \tag{15}$$

and this holds for all (s_t, p_t, z_t). We can derive similar results by imposing similar assumptions on the supply side.

Assumption 10 *The equilibrium price function ρ is such that*

$$\begin{aligned}\frac{\partial \rho_k}{\partial z_{jt}}(z_t, \xi_t, w_t, \eta_t) &= \frac{\partial \rho_k}{\partial \xi_{jt}}(z_t, \xi_t, w_t, \eta_t)\\ \frac{\partial \rho_k}{\partial w_{jt}}(z_t, \xi_t, w_t, \eta_t) &= \frac{\partial \rho_k}{\partial \eta_{jt}}(z_t, \xi_t, w_t, \eta_t) > 0\end{aligned}$$

for all $(z_t, \xi_t, w_t, \eta_t)$ and for all j and k, and the matrix

$$\frac{\partial \rho}{\partial \eta_t'}(z_t, \xi_t, w_t, \eta_t) := \begin{pmatrix} \frac{\partial \rho_1}{\partial \eta_1}(z_t, \xi_t, w_t, \eta_t) & \cdots & \frac{\partial \rho_1}{\partial \eta_J}(z_t, \xi_t, w_t, \eta_t) \\ \vdots & \ddots & \vdots \\ \frac{\partial \rho_J}{\partial \eta_1}(z_t, \xi_t, w_t, \eta_t) & \cdots & \frac{\partial \rho_J}{\partial \eta_J}(z_t, \xi_t, w_t, \eta_t) \end{pmatrix}$$

is positive definite for all $(z_t, \xi_t, w_t, \eta_t)$.

By the second part of Assumption 10, we can write

$$\eta_t = \phi(p_t, z_t, \xi_t, w_t).$$

Substitute $\xi_t = \xi(s_t, p_t, z_t)$ into the above equation, and obtain

$$\eta_t = \phi(p_t, z_t, \xi(s_t, p_t, z_t), w_t) =: \eta(s_t, p_t, z_t, w_t). \tag{16}$$

Note that s_t and p_t are endogenous observed variables and z_t and w_t are exogenous observed variables. I first show that η does not depend on z_t so we can use z_t as instruments for η. Taking a derivative of the identity

$$p_t = \rho(z_t, \xi(s_t, p_t, z_t), w_t, \eta(s_t, p_t, z_t, w_t))$$

with respect to z_t at a given (s_t, p_t, z_t, w_t) we get

$$\begin{aligned} 0 &= \frac{\partial \rho}{\partial z_t'}(z_t, \xi_t, w_t, \eta_t) + \frac{\partial \rho}{\partial \xi_t'}(z_t, \xi_t, w_t, \eta_t)\frac{\partial \xi}{\partial z_t'}(s_t, p_t, z_t) \\ &\qquad + \frac{\partial \rho}{\partial \eta_t'}(z_t, \xi_t, w_t, \eta_t)\frac{\partial \eta}{\partial z_t'}(s_t, p_t, z_t, w_t) \\ &= \frac{\partial \rho}{\partial z_t'}(z_t, \xi_t, w_t, \eta_t) - \frac{\partial \rho}{\partial \xi_t'}(z_t, \xi_t, w_t, \eta_t) + \frac{\partial \rho}{\partial \eta_t'}(z_t, \xi_t, w_t, \eta_t)\frac{\partial \eta}{\partial z_t'}(s_t, p_t, z_t, w_t) \\ &= \frac{\partial \rho}{\partial \eta_t'}(z_t, \xi_t, w_t, \eta_t)\frac{\partial \eta}{\partial z_t'}(s_t, p_t, z_t, w_t) \end{aligned}$$

where the second equality follows from (15), and the last follows by Assumption 10. Since $\frac{\partial \rho}{\partial \eta_t'}(z_t, \xi_t, w_t, \eta_t)$ is invertible, we have

$$\frac{\partial \eta}{\partial z_t'}(s_t, p_t, z_t, w_t) = 0 \tag{17}$$

for all (s_t, p_t, z_t, w_t). Taking a derivative of the above identity now with respect to w_t yields

$$0 = \frac{\partial \rho}{\partial w_t'}(z_t, \xi_t, w_t, \eta_t) + \frac{\partial \rho}{\partial \eta_t'}(z_t, \xi_t, w_t, \eta_t)\frac{\partial \eta}{\partial w_t'}(s_t, p_t, z_t, w_t)$$

The first part of Assumption 10 implies that

$$\frac{\partial \eta}{\partial w_t'}(s_t, p_t, z_t, w_t) = -I \tag{18}$$

Denote the endogenous variables by $y_t = (s_t, p_t)$ and the exogenous variables by $x_t = (z_t, w_t)$. Combine (14) and (16) to write

$$\begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix} = \begin{pmatrix} \xi(s_t, p_t, z_t) \\ \eta(s_t, p_t, z_t, w_t) \end{pmatrix} =: r(s_t, p_t, z_t, w_t) = r(y_t, x_t).$$

Note that $\xi(s_t, p_t, z_t)$ is not a function of w_t, so

$$\frac{\partial \xi}{\partial w_t'}(s_t, p_t, z_t) = 0.$$

This together with Equations (15), (17), and (18) yields

$$\frac{\partial r}{\partial x_t'}(y_t, x_t) = \begin{pmatrix} \frac{\partial \xi}{\partial z_t'}(s_t, p_t, z_t) & \frac{\partial \xi}{\partial w_t'}(s_t, p_t, z_t) \\ \frac{\partial \eta}{\partial z_t'}(s_t, p_t, z_t, w_t) & \frac{\partial \eta}{\partial w_t'}(s_t, p_t, z_t, w_t) \end{pmatrix} = -I.$$

Again, this condition is derived from assumptions on the measurement unit of unobserved terms ξ_t and η_t. By linking the values of unobservables to those of observables, we can fix the scale of the function we are interested in identifying.[24] Given the above condition, we can apply Theorem 3.1 in Matzkin (2010a) to identify the structural inverse functions ξ and η. We need some regularity assumptions.

Assumption 11 (ξ_t, η_t) *is independent of* (z_t, w_t)*, and the density function* $f_{(\xi,\eta)}$ *of* (ξ_t, η_t) *is everywhere positive and continuously differentiable.*

Assumption 12 *The density function* $f_{(z,w)}$ *of* (z_t, w_t) *is continuously differentiable and for any* $(\xi, \eta) \in \mathbb{R}^{2J}$ *and any* $(s, p) \in \Delta^J \times \mathbb{R}_+^J$*, there exists* (z, w) *such that* $(\xi, \eta) = r(s, p, z, w)$.

Assumption 13 *Conditional on* (z_t, w_t)*, the function* r *is twice continuously differentiable, 1-1, onto their support. Also there is a set of values* $(\overline{\xi}, \overline{\eta}, \overline{s}, \overline{p}, \overline{z}, \overline{w})$ *such that* $(\overline{\xi}, \overline{\eta}) = r(\overline{s}, \overline{p}, \overline{z}, \overline{w})$.

The following is the identification assumption.

[24] There are some other ways in which the measurement unit of unobserved variables are specified. See Matzkin (2003, 2007).

Assumption 14 *The density $f_{(\xi,\eta)}$ of (ξ_t, η_t) is such that*

- *for some value (ξ^*, η^*),*

$$\frac{\partial \log f_{(\xi,\eta)}}{\partial \xi_t}(\xi^*, \eta^*) = \frac{\partial \log f_{(\xi,\eta)}}{\partial \eta_t}(\xi^*, \eta^*) = 0$$

- *for each $j = 1, \cdots, J$, there exist values (ξ^{*j}, η^{*j}) and (ξ^{**j}, η^{**j}) such that*

$$\frac{\partial \log f_{(\xi,\eta)}}{\partial \xi_{jt}}(\xi^{*j}\eta^{*j}) \neq 0, \text{ and } \frac{\partial \log f_{(\xi,\eta)}}{\partial \eta_{jt}}(\xi^{**j}, \eta^{**j}) \neq 0$$

but

$$\frac{\partial \log f_{(\xi,\eta)}}{\partial \xi_{kt}}(\xi^{*j}, \eta^{*j}) = \frac{\partial \log f_{(\xi,\eta)}}{\partial \eta_{kt}}(\xi^{**j}, \eta^{**j}) = 0$$

for all $k \neq j$ and

$$\frac{\partial \log f_{(\xi,\eta)}}{\partial \eta_t}(\xi^{*j}, \eta^{*j}) = \frac{\partial \log f_{(\xi,\eta)}}{\partial \xi_t}(\xi^{**j}, \eta^{**j}) = 0$$

While Assumption 14 is long, it is quite mild. It holds if there is a maximum of the joint density function, and if, for any coordinate of unobserved variables, we can find a value for which the joint density is maximized or minimized with respect to other coordinates, while not with respect to it. For instance, there is a unique maximum for the joint normal distribution. Also for any value of some coordinate, unless it is the mean of the normal distribution, we can find a point where the joint density is maximized with respect to other coordinates but not with respect to the coordinate. If density functions are continuous on a closed support and vanish on the boundary, this assumption is trivially satisfied. Exceptions include the uniform distribution and the exponential distribution.

Let E_j be the support of the marginal distribution of ε_{ijt}. I state the main theorem as follows.

Theorem 2 *Suppose Assumptions 1-5 and 9-14 hold. Assume further that the data $(s_t, h_t, p_t, z_t, w_t)$ are available for all $s_t \in \Delta^J$. Then the unobserved characteristics $\{\xi_t\}_{t=1}^T$, and the unobserved cost factors $\{\eta_t\}_{t=1}^T$ are identified, and the utility function $\delta(p_{jt}, z_{jt}, \xi_{jt})$ is nonparametrically identified on the inverse image of $-E_1$. Also the joint density function $f_\varepsilon^{t,t-1}$ of ε_{it} and $\varepsilon_{i,t-1}$ is identified on E_1^{2J}. Especially, $F_\varepsilon^{t,t-1}$ is fully identified if E_1 contains E_j for all j.*

Proof. It is shown above that under Assumptions 1, 2, 3, 9, and 10, we can write the simultaneous equations model

$$\begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix} = r(y_t, x_t) = \begin{pmatrix} \xi(s_t, p_t, z_t) \\ \eta(s_t, p_t, z_t, w_t) \end{pmatrix}$$

where $y_t = (s_t, p_t)$ and $x_t = (z_t, w_t)$. It is verified that in the above model

$$\frac{\partial r}{\partial x}(y_t, x_t) = -I$$

Under Assumptions 11-14, it follows from the proof of Theorem 3.1 in Matzkin (2010a) that r and $f_{(\xi,\eta)}$ are identified. $\{\xi_t\}_{t=1}^T$ and $\{\eta_t\}_{t=1}^T$ are revealed from r, and thus inverting ξ identifies the function $S(p_t, z_t, \xi_t)$. Applying Lemma 2 under Assumptions 1, 2, 3, and 5, and the given condition on the support of s_t, the utility function $\delta(p_{jt}, z_{jt}, \xi_{jt})$ is identified on the inverse image of $-E_1$, and the joint distribution F_ε is identified on E_1^J. Now under Assumptions 1, 2, 3, and 4, Lemma 3 identifies the joint density function $f_\varepsilon^{t,t-1}$ on E_1^{2J}. If E_1 contains E_j for all j, $f_\varepsilon^{t,t-1}$ is fully identified since E_1^{2J} contains the whole support of $f_\varepsilon^{t,t-1}$. Therefore, $F_\varepsilon^{t,t-1}(e)$ can be obtained by integrating $f_\varepsilon^{t,t-1}(e)$ up to e. ∎

Although we are not interested in the equilibrium price function ρ, it is identified as well. Once we obtain $\{\xi_t\}_{t=1}^T$ and $\{\eta_t\}_{t=1}^T$, all the variables in Equation (13) are known. Identification and estimation of the function ρ immediately follows. As Theorem 3.2 in Matzkin (2010a) provides a way to estimate the structural inverse function $r(s_t, p_t, z_t, w_t)$, the functions S and ρ are easily estimated. Estimation of the utility function δ and the distribution function $F_\varepsilon^{t,t-1}$ using r is left for a future project.

4 Identification of switching costs

In this section, I present the model where there is a switching cost when consumers change products. For example, consumers may incur learning costs when they investigate new products.[25] When a subscription is required, there could be an initial

[25] Although we usually assume consumers have full information about all products, in reality they may not have full information about products they have not purchased recently. Modeling incomplete information is technically difficult, but we may approximate such a model by introducing a switching cost, where part of the switching cost is incurred from uncertainty about other products.

installment fee, a quitting fee, or both. Non-monetary costs from beginning a subscription or quitting it are also included. If there is a long-term contract with a cancellation fee, such a fee is also considered a switching cost. In some industries, using a service requires buying a product-specific equipment. Then switching consumers need to buy a new equipment, which is an additional switching cost.

Selecting which type of costs are included in switching costs has implications for my model, because different types of switching costs affect consumers' choices in different ways. Modeling all possible switching costs is not desirable considering the cost of increasing complexity in computation. In this paper, I assume that there exists a simple switching cost which is incurred whenever a consumer discontinues a subscription to some service. The other type of switching costs are not considered in the model. There are several reasons why I choose this model.

Since we use the churn rate, which is the proportion of people who quit from a service, modeling the quitting cost directly utilizes such data most efficiently. Although various type of switching costs affect consumers' decisions, the model with a quitting cost can approximate the effect of those costs to some extent. For example, If most people who quit from a service subscribe to another, the quitting cost would capture a high portion of the initial installment fee and the cost of buying new equipment.[26] Also those costs are observed relatively well so their identification is not so much an issue as that of unobserved switching costs. If we want to find an effect of those observed costs, we can simply run a reduced form regression. On the other hand, the quitting cost is pretty unknown, which we suspect has a hidden effect on estimation.[27]

Let c_{kt} denote a quitting cost that consumers pay when discontinuing a subscription to product k at time t. Note again that although we refer to c_{kt} as a quitting cost, it includes other costs of switching as well. If switching costs are interpreted

[26]The cancellation fee is different from other fees in that it affects long-term decisions. Consumers who engaged in a long-term contract are those who think that making a long-term contract with a less initial setup cost is expected to yield higher utility than paying a high initial setup cost. For those people, the cancellation fee would have a less impact on their choice on average than the initial cost or the quitting cost.

[27]A model with an unknown initial setup cost is indeed as computationally difficult as one with an unknown quitting cost. But as mentioned, most of such costs are known, so the model with an unknown quitting cost is conceptually more reasonable.

monetarily, a consumer gets a utility $\delta(p_{jt} - c_{kt}, z_{jt}, \xi_{jt}) + \varepsilon_{ijt}$ when switching from k to j. Even this model is not easy to identify due to nonlinearity of the utility function. Thus we need to impose more assumptions. Rather than appealing to unintuitive assumptions on the shape of the utility function, I assume that the switching cost enters the utility function in an additively separable way. The following assumption replaces Assumption 1.

Assumption 15 (Additive Switching Cost) *The utility function is given by*

$$u_{ijt} = \begin{cases} \delta(p_{jt}, z_{jt}, \xi_{jt}) - c_{kt} + \varepsilon_{ijt} & \text{when } i \text{ purchased } k \neq j, 0 \text{ at } t-1 \\ \delta(p_{jt}, z_{jt}, \xi_{jt}) + \varepsilon_{ijt} & \text{otherwise} \end{cases}$$

We may simply write $\delta_{jt} := \delta(p_{jt}, z_{jt}, \xi_{jt})$ as in Section 3. Then, the utility is $u_{ijt} = \delta_{jt} - c_{kt} + \varepsilon_{ijt}$ if i bought $k \neq j, 0$ in the previous period, and is $u_{ijt} = \delta_{jt} + \varepsilon_{ijt}$ otherwise. The equation specifying the shares of products is different from (5), since the shares at t now depend on the shares at $t-1$. Without switching costs, all consumers have the same choice probabilities ex ante, regardless of which product they purchased in the previous period. But when there are switching costs, two consumers may choose a different product depending on their purchase history even though they draw the same idiosyncratic preference shocks.

Ideally this calls for modeling a dynamic decision of consumers since they would consider the effect of their current decision on their future utility. However we restrict our attention to a static model for tractability. A dynamic model requires an analysis of value functions. While it is estimable, its identification is mathematically more demanding, and therefore I consider a static decision of switching in this paper. There are a couple of ways to relate a static model to a dynamic model. A static model is a dynamic model with impatient or myopic consumers. In another viewpoint, it may be regarded as an approximation to the dynamic model with patient consumers by assuming that the static utility function captures a part of expected future utility.[28]

[28] Berry, Levinsohn, and Pakes (1995) use such an approximation. They use a static utility model for an analysis of the automobile market. Although consumers pay today but can use a car for several years, or sell their old car and buy a new one, the expected utility of all such behaviors is captured in the static utility function in their model. See Schiraldi (2008) for a dynamic treatment with the automobile industry.

In this sense, the static decision model presented in this paper has an implication, although imperfect and limited, on the importance of switching costs.

Consider the consumer who purchased product k in the last period. She would have the set of potential utilities

$$(\delta_{1t}-c_{kt}+\varepsilon_{i1t},\cdots,\delta_{k-1,t}-c_{kt}+\varepsilon_{i,k-1,t},\delta_{kt}+\varepsilon_{ikt},\delta_{k+1,t}-c_{kt}+\varepsilon_{i,k+1,t},\cdots,\delta_{Jt}-c_{kt}+\varepsilon_{iJt})$$

with the utility for the outside good $u_{i0t} = -c_{kt}$. Assuming that her ε_{it} is independent of $\varepsilon_{i,t-1}$, the fact that she bought k in the last period does not provide any information on the distribution of ε_{it}. Therefore, the ex ante probability that she buys products $1, \cdots, J$ is given by

$$\sigma(\delta_{1t} - c_{kt}, \cdots, \delta_{k-1,t} - c_{kt}, \delta_{kt}, \delta_{k+1,t} - c_{kt}, \cdots, \delta_{Jt} - c_{kt} | u_{i0t} = -c_{kt})$$

or simply

$$\sigma(\delta_{kt}, \delta_{qt} - c_{kt} \ \forall q \neq k | u_{i0t} = -c_{kt})$$

which is a vector-valued function. Its j-th coordinate is the probability that she buys product j at t given that she bought product k at $t-1$. Thus it is a rate of switching to j for consumers who bought k at $t-1$. We can derive similar expressions for consumers who purchased products $1, \cdots, J$ in the last period. In addition, the choice probabilities for consumers who purchased nothing in the last period is simply $\sigma(\delta_t)$, since there is no initial setup cost in this model. Using the shares at $t-1$ with the conditional purchase probabilities derived above, we have the following equation for the share at t.

$$s_{jt} = \sum_{k=1}^{J} \sigma_j(\delta_{kt}, \delta_{qt} - c_{kt} \ \forall q \neq k | u_{i0t} = -c_{kt}) s_{k,t-1} + \sigma_j(\delta_t) s_{0,t-1} \tag{19}$$

The conditional purchase probability implies the equation for the churn rate as well. Since the churn rate h_{jt} is the rate at which people who bought product j at time $t-1$ switch out of product j at time t,

$$h_{jt} = 1 - \sigma_j(\delta_{jt}, \delta_{qt} - c_{jt} \ \forall q \neq j | u_{i0t} = -c_{jt})$$

or equivalently,

$$1 - h_{jt} = \sigma_j(\delta_{jt}, \delta_{qt} - c_{jt} \ \forall q \neq j | u_{i0t} = -c_{jt}) \tag{20}$$

Equations (19) and (20) for $j = 1, \cdots, J$ are the identification equations. I present a theorem on identification of switching costs using the following assumptions.

Assumption 16 (Distribution) *ε_{it} is independent and identically distributed over i and t. It is absolutely continuously distributed, and the support of the density function is connected. The joint distribution of ε_{it} is known, and denoted by F_ε.*

Assumption 17 *There exists w_{jt} which includes z_{jt} in it such that for any function $B : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ with a finite expectation, $E[B(\delta_{jt}, p_{jt})|w_{jt}] = 0$ a.e. implies $B(\delta_{jt}, p_{jt}) = 0$ a.e.*

Assumption 16 replaces Assumptions 3 and 5. The new assumption is stronger, but if the independence assumption does not hold, the equations for the market shares and the churn rates become intractable. I will discuss such a case later in this section. Assumption 17 replaces Assumption 7. This is a weaker assumption. Note that if the utility function is specified parametrically, this condition can be written in terms of the rank of a matrix.

Theorem 3 *Suppose Assumptions 2, 15, 16 hold. Then the switching costs $\{c_t\}_{t=2}^T$ are identified. If Assumptions 6, 8, and 17 hold in addition, the unobserved characteristics $\{\xi_t\}_{t=2}^T$, and the utility function $\delta(p_{jt}, z_{jt} + \xi_{jt})$ are identified.*

Proof. It follows from Assumptions 2, 15, and 16 and Lemma 4 that there exists a unique set of $\{(\delta_t, c_t)\}_{t=2}^T$ that satisfies (19) and (20). Once $\{(\delta_t, c_t)\}_{t=2}^T$ is identified, we observe $\{(\delta_t, p_t, z_t)\}_{t=2}^T$ from the set of equations

$$\delta_{jt} = \delta(p_{jt}, z_{jt}, \xi_{jt}) = \delta(p_{jt}, z_{jt} + \xi_{jt})$$

for all j and $t = 2, \cdots, T$. We can identify the utility function δ and ξ_{jt} following the same procedure used in the proof of Theorem 1. In other words, we invert the function with respect to its second argument and take a conditional expectation, then by Assumption 8,

$$E[\delta^{-1}(\delta_{jt}, p_{jt})|w_{jt}] = E[z_{jt} + \xi_{jt}|w_{jt}] = z_{jt}$$

Assumption 17 identifies the function δ^{-1}, and thus ξ_{jt} is identified from

$$\xi_{jt} = \delta^{-1}(\delta_{jt}, p_{jt}) - z_{jt}$$

for all j and $t = 2, \cdots, T$. Now we have $\{(\delta_t, p_t, z_t, \xi_t)\}_{t=1}^T$, so the utility function δ is obtained easily. ∎

Two results immediately follow. The first is that a model with a uniform switching cost across products is trivially identified given Theorem 3 since it is a special case of the above model where the switching cost is different across products. The model with product-specific switching costs is more useful if we are interested in a firm's incentive to raise the quitting cost for consumers, while the model with a uniform switching cost might suffice to examine the welfare effect of economic policies for quitting costs. An extreme case would be a model with a constant switching cost over time, which is also identified. The second is that the parametric utility function is identified. Given that $\{\delta_t\}_{t=2}^T$ are revealed, and that there exist a set of appropriate instrumental variables, using the generalized method of moments identifies the parameters in the utility function. This approach is applied to the estimation of a linear utility function using the U.S. mobile phone service industry data in Section 5.

With a known distribution of idiosyncratic preference shocks, mean utilities $\{\delta_t\}_{t=2}^T$ and switching costs $\{c_t\}_{t=2}^T$ are easily identified. For example, when the preference shocks are given by logit errors, the mean utilities and the switching costs are obtained from the contraction mapping. Berry, Levinsohn, and Pakes (1995) show a contraction mapping identifies $\{\delta_t\}_{t=1}^T$ in the random coefficients logit model without switching costs. A simple application of their theorem shows that given $\{\delta_t\}_{t=2}^T$, the switching costs $\{c_t\}_{t=2}^T$ are identified using the same sort of contraction mapping. This implies that we can use a double-layered contraction mapping to obtain $\{\delta_t\}_{t=2}^T$ and $\{c_t\}_{t=2}^T$. Unfortunately, it is not proven that the mapping on $\{\delta_t\}_{t=2}^T$ is also a contraction of modulus less than one. This is because the switching costs are implicitly a function of mean utilities, so the second part of Condition (1) in their theorem is not easily verified. However, in the estimation with the U.S. mobile phone service industry data, the contraction mapping always converges to a point and the difference between the mean utilities over iterations decreases as the mapping is applied repeatedly.

Finally I make technical remarks on identification of the model with serial correlation of idiosyncratic preference shocks in addition to switching costs. With switching costs, consumers with the same preference shocks ε_{it} make a different choice depending on what they purchased at $t-1$. This implies that s_{jt} cannot be characterized by the probability space generated by ε_{it}. This is why we need to write the share equation in the following way.

$$s_{jt} = P(i \text{ buys } j \text{ at } t) = \sum_{k=0}^{J} P(i \text{ buys } j \text{ at } t | \text{bought } k \text{ at } t-1) s_{j,t-1}$$

Without independence, $P(i \text{ buys } j \text{ at } t | \text{bought } k \text{ at } t-1)$ cannot be simplified.[29] In fact, it is not characterized by the probability space generated by ε_{it} and $\varepsilon_{i,t-1}$ for the same reason as above. We need information from $t-2$, which in turn requires information from $t-3$. The backward reference continues infinitely, and we end up with an equation with an infinite sum. Hence the mathematical identification of such a model seems infeasible.

5 Application to the U.S. mobile service industry

I show in this section that inclusion of switching costs is necessary for consistent estimation by applying the model in Section 4 to empirical estimation of the utility function for a mobile phone service in the United States. The data come from two *Global Wireless Matrix* reports on the mobile phone service industry by Merrill Lynch. One contains data from 1999 Q1 to 2004 Q2, and the other from 2005 Q1 to 2007 Q4 allowing the analysis to span 34 quarters over 9 years. Although the reports are global, in this study I focus on the United States. This is because demand functions may vary across countries, and because different regulation policies and the resulting competition patterns may potentially affect the estimation. The different policies may also generate different distributions of the unobserved characteristics that the estimation procedure crucially depends on.

The first report provides data on the six leading firms, while the second report contains data on the four leading firms. This is due to two big mergers and acquisitions among the first six leading firms which occurred between late 2004 and early 2005.[30]

[29] An AR(1) assumption fails to simplify the expression.

[30] The U.S. government approved the acquisition of AT&T Wireless by Cingular Wireless on Oct. 26th, 2004. Sprint and Nextel merged to form Sprint Nextel which was approved by The Department of Justice on Aug. 3rd, 2005.

There are more than a hundred mobile phone service providers in the United States, but the top six firms accounted for 72% of the total share in 1999, and the top four firms accounted for 80% of the total share in 2007. Although there were more mergers during the data period, those did not significantly increase the share of a particular firm.

I use 1999 Q4 as the first period since churn rates become available in the following period.[31] This results in 162 observations from six firms in the first 19 quarters, and four firms in the last 12 quarters. Due to the small sample size, a simple parametric utility function is employed as follows.

$$u_{ijt} = \alpha p_{jt} + \beta t + \gamma + \xi_{jt} + \varepsilon_{ijt}$$

where t denotes the quarter, and γ a constant. Inclusion of t is prompted by the fact that minutes of use increases over time for all services. This partially captures network externalities in demand since more people use a mobile phone service over time. Average revenue per user is denoted p_{jt}. Switching costs enter additively when i switches from $k \neq j, 0$ as described in Assumption 15 in Section 4. Following the specification used in identification of switching costs in Section 4, I assume that consumers are myopic and do not consider future costs on choosing a product.

Two instrumental variables are used for p_{jt}. As we have an EBITDA margin, we can use it to recover variable costs. In particular, average variable costs are calculated by subtracting average EBITDA from average revenue.[32] I use the firm's own average variable costs and an average of the other firms' average variable costs as instrumental variables for service price. When we interpret the unobserved characteristic ξ_{jt} as service quality it is a reasonable assumption that ξ_{jt} is uncorrelated with variable costs, since the quality of a mobile phone service is likely determined by long term investments rather than by short term spending. If the unobserved characteristic captures market

[31] Although churn rates are also unavailable in 1999 Q4, we need the shares in that period to calculate shares in 2000 Q1 with the model equations.

[32] EBITDA stands for the earnings before interest, taxes, depreciation, and amortization. An EBITDA margin is defined as the ratio of EBITDA to total revenue. EBITDA is not exactly economic profit, and thus subtracting it from total revenue does not coincide exactly with economic cost. However, since accounting costs excluding interest, taxes, depreciation, and amortization are related to production cost, and especially to variable costs, so they can be used as a proxy for economic cost.

specific demand shocks, it would not be correlated with variable costs as long as firms do not observe the shocks before making a decision on short term spending but only before setting price. Such an argument may fail if the short term spending includes expenses on promotion that boosts demand. But if this is the case, the estimate from instrumental variables approach would be the most conservative bound for the parameter.

Using the dataset I show that exclusion of switching costs from the model has a significant impact on the estimation results. Figure 1 illustrates how distorted the implied churn rates are when switching costs are assumed to be zero. Churn rates do not exceed 16% in the data, but the simulation with zero switching costs generates churn rates greater than 75% in general. Looking at (19) and (20), it is obvious to see why such distortion happens. A churn rate from a service must coincide with one minus its market share in the model with $c_{jt} = 0$ for all j and t, since every consumer has the same conditional choice probability no matter what service she used in the last period. As the largest market share is approximately 23% in the data, the above argument predicts that the model with zero switching costs would yield churn rates greater than 77%, which is consistent with the simulation result.

Therefore switching costs must be included in the model to appropriately account for churn rates in the data. I first use a simple logit model allowing for time-varying product-specific quitting costs. As shown in Section 4, switching costs can be identified using a contraction mapping in the logit model. A double contraction mapping is used to find mean utilities and switching costs. In the inner loop switching costs are obtained by matching churn rates in the data to those implied by the model given the guess on mean utilities. Mean utilities converge to a point in the outer loop through a mapping of which the only fixed point is the true mean utilities. The identified mean utility $\widehat{\delta}_{jt}$ is then used as a dependent variable in the following regression.

$$\widehat{\delta}_{jt} = \alpha p_{jt} + \beta t + \gamma + \xi_{jt}$$

The estimation results are presented in Table 1. The first three columns show the estimates with the assumption of zero switching costs, and the next three are the results with switching costs considered. As expected, an OLS regression of the implied mean utilities yields too low a price coefficient. Using instrumental variables yields a higher price coefficient, which implies that the instrumental variables seem to correct a bias.

Comparing (vi) to (iii) in Table 1, inclusion of switching costs reduces the price coefficient by 39%. This may look unintuitive at a first glance, if one thinks that the model without switching costs underestimates the price effect. Such a conclusion might hold if consumers who do not switch in fact due to high switching costs are misunderstood to be price insensitive in the model without switching costs. However, the effect of misspecification on the price coefficient is ambiguous. There is no one-to-one relationship between the current shares and characteristics of products. Current shares are determined by previous shares, the current characteristics of products, and switching costs. When switching costs are assumed away, previous shares do not matter, and thus the implied price coefficient captures the unconditional correlation between current share and price. On the other hand, the real price coefficient captures the correlation between the two variables conditional on previous shares and switching costs. Since the naïve reasoning made above ignores such dynamics, it may not lead to a true conclusion.

The coefficient on time is positive in all of the specifications. This is reasonable since the market share of each service is increasing over time during the data period. As mentioned above, this partially reflects the network externalities in demand. Mobile phones become more useful as there are more people using them. Another possibility would be an improvement in functions of mobile phones. Especially toward the end of the data period, smart phones become more widespread, which dramatically increased the utility from using a mobile phone. For the same reasons, we may suspect that the unobserved characteristic ξ_{jt} might be correlated over time for a service. Thus I also provide heteroskedasticity and autocorrelation consistent (HAC) standard errors.[33]

Note that the coefficient on time is less in (vi) than in (iii). This has the following implication: without switching costs in the model, we may incorrectly conclude that the share increased more quickly due to the reasons mentioned in the above paragraph. While such effects exist to some extent, consumers may remain subscribing in order to avoid switching costs. Comparing (vi) to (iv), the coefficient on time does not change much when instrumental variables are used. This is reasonable since time is exogenous.

[33] Newey and West (1987) suggest a simple way to estimate a HAC covariance matrix. For details see Andrews (1991).

There is a relatively large difference in the coefficient on time between (i) and (iii). It is probably because the model is incorrectly specified, in which case the estimates may be more sensitive to the use of instrumental variables.

For the second specification for ε_{ijt}, I estimate the nested logit model. Let all the service providers in a group, and the outside option in the other group. Let $\sigma \in [0, 1]$ be the portion of ε_{ijt} explained by a group-wise logit error. If switching costs do not exist, Berry (1994) shows that the following holds.

$$\log \frac{s_{jt}}{s_{0t}} = \sigma \log \frac{s_{jt}}{1 - s_{0t}} + \alpha p_{jt} + \beta t + \gamma + \xi_{jt}$$

First I simply mimic such a method and estimate the following regression model.

$$\widehat{\delta}_{jt} = \sigma \log \frac{s_{jt}}{1 - s_{0t}} + \alpha p_{jt} + \beta t + \gamma + \xi_{jt}$$

where $\widehat{\delta}_{jt}$ is the mean utility identified before. Table 2 (i) and (ii) present the results from the above two regressions. The estimates are not precise,[34] and moreover the estimate for σ does not fall in $[0, 1]$.

Indeed, mimicking Berry's equation is a naïve approach when there are switching costs. Shares and churn rates are a function of true mean utilities, switching costs, and the distribution of idiosyncratic shocks. Since the distribution of idiosyncratic shocks depends on σ, the mean utility matching the data would depend on σ as well. Therefore we need to find the mean utility $\widetilde{\delta}_{jt}$ considering the effect of σ on shares and churn rates. Once such $\widetilde{\delta}_{jt}$ is obtained, I estimate the following equation.

$$\widetilde{\delta}_{jt} = \alpha p_{jt} + \beta t + \gamma + \xi_{jt}$$

The outer loop searches for σ, while the inner loop identifies $\widetilde{\delta}_{jt}$, estimates the above regression, and calculates the objective function from the moment equations given σ. The results are given in Table 2 (iii) and (iv). They indicate that $\sigma = 0$ explains the data best, and that the estimates of utility parameters coincide with those in Table 1 (v) and (vi).

The last column (v) in Table 2 shows the result from the model where a random coefficient is given to the group dummy, which is similar to the nested logit model.

[34] One of the reasons is that the 2SLS estimator does not have a finite first moment when the model is exactly identified. See Kinal (1980).

In this case $\sigma \in \mathbb{R}_+$ represents the standard deviation of the random coefficient, and is estimated at 0. The utility parameter estimates are again the same with those in Table 1 (vi). Finally, I employ a model where there is unobserved heterogeneity in price sensitivity across consumers. Even when the switching costs are not considered, the parameter which controls the variance of the price coefficient is not precisely estimated. This is also the case when I try several models with different types of switching costs.

6 Conclusion

It is well known in the industrial organization literature that data on market shares and characteristics of products can reveal the utility function or the demand function. Using the nonparametric identification technique recently developed in econometrics, I show in this paper that additional data on churn rates from products can fully identify the transition of idiosyncratic preference shocks. This is useful when we want to have information on which type of consumers are likely to purchase the same product, and which type are likely to switch. This information can be used to change regulation or to improve advertisements. I also show that switching costs are identified from churn rates. As the empirical estimation of the United States mobile phone service illustrates, exclusion of substantial switching costs from a model may lead to inconsistent estimation. Including switching costs also increases the fit of the model in predicting churn rates.

This paper has some limitations. I assume that idiosyncratic preference shocks are independent of products characteristics. Identifying how idiosyncratic shocks evolve by modeling their dependence on observed characteristics as in the random coefficients model would be an interesting exercise. Nonparametric dependence seems very difficult to identify, but employing parametric dependence offers a way to comply with nonparametric identification of idiosyncratic errors. Considering endogenous switching costs is also an interesting extension to a different direction. When firms can control switching costs to some extent, consumers would optimally respond to firms' incentive to change switching costs. Such a model would involve more serious dynamic decisions. A nonparametric estimation of a utility function is not proposed in this paper, and left for a future project.

References

[1] Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

[2] Bajari, P., J. Fox, K. Kim, and S. Ryan (2009), "The Random Coefficients Logit Model Is Identified," NBER working paper.

[3] Berry, S. (1994), "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics* 25, 242-262.

[4] Berry, S., and P. Haile (2009), "Identification of Discrete Choice Demand from Market Level Data," working paper, Yale University.

[5] Berry, S., and P. Haile (2010), "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," working paper.

[6] Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica* 63, 841-890.

[7] Bresnahan, T. (1987), "Competition and Collusion in the American Automobile Industry: The 1955 Price War," *Journal of Industrial Economics* 35, 457-482.

[8] Chamberlain, G. (1984), "Panel Data," Chapter 22 in *Handbook of Econometrics*, Vol. 2, edited by Z. Griliches and M.D. Intriligator, Elsevier B.V., 1247-1318.

[9] Chernozhukov, V., and C. Hansen (2005), "An IV Model of Quantile Treatment Effects," *Econometrica* 73, 245-261.

[10] Chernozhukov, V., G. Imbens, and W. Newey (2007), "Instrumental variable estimation of nonseparable models," *Journal of Econometrics* 139, 4-14.

[11] Chiappori, P., and I. Komunjer (2009), "On the Nonparametric Identification of Multiple Choice Models," discussion paper, UCSD.

[12] Chintagunta, P.K., D.C. Jain, and N.J. Vilcassim (1991), "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data," *Journal of Marketing Research* 28, 417-428.

[13] Cullen, J., and O. Shcherbakov (2010), "Measuring Consumer Switching Costs in the Wireless Industry," working paper.

[14] Erdem, T., M. Keane, T. Öncü, and J. Strebel (2005), "Learning About Computers: An Analysis of Information Search and Technology Choice," *Quantitative Marketing and Economics* 3, 207-246.

[15] Gönül, F., and K. Srinivasan (1993), "Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues," *Marketing Science* 12, 213-229.

[16] Gowrisankaran, G., and M. Rysman (2009), "Dynamics of Consumer Demand for New Durable Goods," working paper.

[17] Hausman, J.A., and D.A. Wise (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica* 46, 403-426.

[18] Hendel, I., and A. Nevo (2006), "Measuring the Implications of Sales and Consumer Inventory Behavior," *Econometrica* 74, 1637-1673.

[19] Honorè, B.E., and E. Kyriazidou (2000), "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica* 68, 839-874.

[20] Ichimura, H., and T.S. Thompson (1998), "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution," *Journal of Econometrics* 86, 269-295.

[21] Kim, J., (2006), "Consumers' Dynamic Switching Decisions in the Cellular Service Industry," working paper.

[22] Kinal, T.W. (1980), "The Existence of Moments of k-class Estimators," *Econometrica* 48, 241-249.

[23] Matzkin, R.L. (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339-1375.

[24] Matzkin, R.L. (2007), "Nonparametric Identification," Chapter 73 in *Handbook of Econometrics*, Vol. 6b, edited by J.J. Heckman and E.E. Leamer, Elsevier B.V., 5307-5368.

[25] Matzkin, R.L. (2008), "Identification in Nonparametric Simultaneous Equations, *Econometrica* 76, 945-978.

[26] Matzkin, R.L. (2010a), "Estimation of Nonparametric Models with Simultaneity," working paper.

[27] Matzkin, R.L. (2010b), "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity," mimeo, UCLA.

[28] McFadden, D.L. (1976), "Quantal Choice Analysis: A Survey," *Annals of Economic and Social Measurement* 5, 363-390.

[29] Nevo, A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica* 69, 307-342.

[30] Newey, W.K., and J.L. Powell (2003), "Instrumental Variable Estimation of Nonparametric Models," *Econometrica* 71, 1565-1578.

[31] Newey, W.K., and K.D. West (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

[32] Osborne, M. (2007), "Consumer Learning, Switching Costs, and Heterogeneity: A Structural Examination," working paper.

[33] Quandt, R. (1966), "Probabilistic Theory of Consumer Behavior," *Quarterly Journal of Economics* 70, 507-536.

[34] Schiraldi, P. (2008), "Automobile Replacement: a Dynamic Structural Approach," working paper.

[35] Shcherbakov, O. (2009), "The Effect of Consumer Switching Costs on Market Power of Cable Television Providers," working paper.

[36] Shy, O. (2002), "A Quick-and-easy Method for Estimating Switching Costs," *International Journal of Industrial Organization* 20, 71-87.

[37] Sun, B., and S.A. Neslin (2003), "Measuring the Impact of Promotions on Brand Switching When Consumers Are Forward Looking," *Journal of Marketing Research* 40, 389-405.

[38] Wooldridge, J.M. (2005), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics* 20, 39-54.

Appendices

A. Identification without full support of shares

The assumption on the support of shares s_t required for Theorems 1 and 2 is quite strong. In particular, for some of the distributions F_ε and the utility function, we may not observe the shares on the boundary of Δ^J.[35] However, the proof of Lemma 2 exploits the full support assumption intensively. It is essential for nonparametric identification of the utility function δ and the joint distribution F_ε of the idiosyncratic preference shocks.

In this appendix, I provide another version of identification theorems which do not require the full support of shares. Note that Assumption 5 specifies the marginal distribution of F_ε and thus leads us to use a share on the boundary of the full support. This is why Lemma 2 and thus Theorems 1 and 2 heavily rely on the full support of shares. I introduce a different normalization assumption than Assumption 5. Let $\mathcal{S}$ be the joint support of the distribution of (p_t, z_t, ξ_t). Although ξ_t is unobserved, $\mathcal{S}$ will be known after the series $\{\xi_t\}_{t=1}^T$ are identified.

Assumption 18 (Normalization) *The support $\mathcal{S}$ is convex and symmetric. Pick $(\overline{p}, \overline{z}, \overline{\xi})$ so that $(\overline{p}, \overline{z}, \overline{\xi})^J \in \mathcal{S}$. Then*

(i) The utility function $\delta(p_{jt}, z_{jt}, \xi_{jt})$ is known at a point $(\overline{p}, \overline{z}, \overline{\xi})$. Let $\overline{\delta} := \delta(\overline{p}, \overline{z}, \overline{\xi})$.

(ii) The share function $\sigma(\delta_t)$ is such that $\sigma_1(\delta_1, \overline{\delta}, \cdots, \overline{\delta})$ is known for all $\delta_1 \in \mathbb{R}$.

The second part of the above assumption is not restrictive. Note first that the range of $\sigma_1(\delta_1, \overline{\delta}, \cdots, \overline{\delta})$ can always be $[0, 1]$. For any distribution function F_ε, it is clear from (4) that $\sigma_1(\delta_1, \overline{\delta}, \cdots, \overline{\delta}) \to 0$ as $\delta_1 \to -\infty$, that $\sigma_1(\delta_1, \overline{\delta}, \cdots, \overline{\delta}) \to 1$ as $\delta_1 \to \infty$, and that $\sigma_1(\delta_1, \overline{\delta}, \cdots, \overline{\delta})$ is increasing in δ_1. Now recall from Lemma 1 that

[35] With the logit error distribution or the normal distribution which is defined on $\mathbb{R}^J$, every product earns a non-zero share and cannot have an entire share, so a share on the boundary of Δ^J does not appear. Since we often use such distributions under whose support is a whole Euclidian space $\mathbb{R}^J$, application of Theorems 1 and 2 is limited for these kind of models.

σ and F_ε have a one-to-one relationship. Hence $\sigma_1(\delta_1, \overline{\delta}, \cdots, \overline{\delta})$ can be regarded as a quasi-conditional distribution of δ_1 given $(\delta_2, \cdots, \delta_J) = (\overline{\delta}, \cdots, \overline{\delta})$. An assumption that $\sigma_1(\delta_1, \overline{\delta}, \cdots, \overline{\delta})$ is known is similar in its spirit, although not the same, with Assumption 5 that the marginal distribution of ε_{i1t} is known. On the other hand, the first part of Assumption 18 puts some restriction on the utility function δ. However, this can be thought of the location normalization of δ. Let $\mathcal{S}_1 := \{(p_{1t}, z_{1t}, \xi_{1t}) : (p_t, z_t, \xi_t) \in \mathcal{S}\}$, then $\mathcal{S} = \mathcal{S}_1^J$ since $\mathcal{S}$ is symmetric.

Theorem 4 *Let $\mathcal{S} = \mathcal{P}^J \times \mathcal{Z}^J \times \Xi^J \subset \mathbb{R}_+^J \times \mathbb{R}^{2J}$ be the support of (p_t, z_t, ξ_t). Suppose Assumptions 1-4, 6-8, and 18 hold. Then the unobserved characteristic $\{\xi_t\}_{t=1}^T$ is identified, and the utility function $\delta(p_{jt}, z_{jt} + \xi_{jt})$ is nonparametrically identified on $\mathcal{P} \times \mathcal{Z} \times \Xi$. The joint distribution F_ε is identified on $-\delta(\mathcal{P} \times \mathcal{Z} \times \Xi)^J$, and the joint density function $f_\varepsilon^{t,t-1}$ of ε_{it} and $\varepsilon_{i,t-1}$ is identified on $-\delta(\mathcal{P} \times \mathcal{Z} \times \Xi)^{2J}$.*

Proof. Following the proof of Theorem 1, it is clear that Assumptions 1-3 and 6-8 identify $\{\xi_t\}_{t=1}^T$ and thus $S(p_t, z_t, \xi_t)$. Then Assumptions 1-3 and 18, and Lemma 5 imply that the utility function $\delta(p_{jt}, z_{jt} + \xi_{jt})$ is identified on $\mathcal{P} \times \mathcal{Z} \times \Xi$, and that the joint distribution F_ε is identified on $-\delta(\mathcal{P} \times \mathcal{Z} \times \Xi)^J$. Now Assumptions 1-4, and Lemma 3 imply that the joint density function $f_\varepsilon^{t,t-1}$ of ε_{it} and $\varepsilon_{i,t-1}$ is identified on $-\delta(\mathcal{P} \times \mathcal{Z} \times \Xi)^{2J}$. ∎

Theorem 5 *Let $\mathcal{S} = \mathcal{P}^J \times \mathcal{Z}^J \times \Xi^J \subset \mathbb{R}_+^J \times \mathbb{R}^{2J}$ be the support of (p_t, z_t, ξ_t). Suppose Assumptions 1-4, 9-14, and 18 hold. Then the unobserved characteristics $\{\xi_t\}_{t=1}^T$ and the unobserved cost factors $\{\eta_t\}_{t=1}^T$ are identified, and the utility function $\delta(p_{jt}, z_{jt}, \xi_{jt})$ is nonparametrically identified on $\mathcal{P} \times \mathcal{Z} \times \Xi$. The joint distribution F_ε is identified on $-\delta(\mathcal{P} \times \mathcal{Z} \times \Xi)^J$, and the joint density function $f_\varepsilon^{t,t-1}$ of ε_{it} and $\varepsilon_{i,t-1}$ is identified on $-\delta(\mathcal{P} \times \mathcal{Z} \times \Xi)^{2J}$.*

Proof. Following the proof of Theorem 2, it is clear that Assumptions 1-3 and 9-14 identify $\{\xi_t\}_{t=1}^T$, $\{\eta_t\}_{t=1}^T$, and thus $S(p_t, z_t, \xi_t)$. The rest of the proof is the same with that of Theorem 4. ∎

Note a couple of comments. Even if the support $\mathcal{S}$ is not symmetric, we may find its symmetric subset and apply the above theorems. Alternatively, we may extend the theorems in the way that the utility function is identified on $\overline{\mathcal{S}}_1 := \{(p_{1t}, z_{1t}, \xi_{1t}) : (p_t, z_t, \xi_t) \in \overline{\mathcal{S}}\}$ where $\overline{\mathcal{S}} := \{(p_t, z_t, \xi_t) \in \mathcal{S} : (p_{jt}, z_{jt}, \xi_{jt}) = (\overline{p}, \overline{z}, \overline{\xi})\ \forall j \neq 1\}$. In such a case, the identified set depends on the product we choose for Assumption 18, so we may want to choose the one that makes the identified set as large as possible. Second, we may specify the utility function semiparametrically as in Newey and Powell (2003) so that it is nonparametric on a compact subset of the full support, but parametric on the rest of the support. Then Lemma 5 shows that we are able to globally identify the utility function while allowing it to be partially nonparametric.

Appendix B. Lemmas

As in Section 3, let E_j be the support of the marginal distribution of ε_{ijt}.

Lemma 2 *Suppose Assumptions 1-3 and 5 hold. Assume further that the data* (s_t, p_t, z_t) *are available for all* $s_t \in \Delta^J$ *and that* $\{\xi_t\}_{t=1}^T$ *are given. If the share function* $S(p_t, z_t, \xi_t)$ *defined by (7) is known, the utility function* $\delta(p_{jt}, z_{jt}, \xi_{jt})$ *is identified on the inverse image of* $-E_1$*. Also the joint distribution* F_ε *is identified on* E_1^J*. Especially,* F_ε *is fully identified if* E_1 *contains* E_j *for all* j*.*

Proof. Define the set S_a as the collection of (p_t, z_t, ξ_t) generating the share $a \in \Delta^J$ so that

$$S_a = \{(p_t, z_t, \xi_t) | S(p_t, z_t, \xi_t) = a\}$$

Pick $a = (a_1, 0, \cdots, 0) \in \Delta^J$. For all $(p_t, z_t, \xi_t) \in S_a$, we have

$$\sigma(\delta(p_{1t}, z_{1t}, \xi_{1t}), \cdots, \delta(p_{Jt}, x_{Jt}, \xi_{Jt})) = (a_1, 0, \cdots, 0)$$

Equation (4) implies

$$\begin{aligned} a_1 =& P\big(\varepsilon_{ikt} - \varepsilon_{i1t} \leq \delta(p_{1t}, z_{1t}, \xi_{1t}) - \delta(p_{kt}, z_{kt}, \xi_{kt}) \ \forall k \text{ and } -\varepsilon_{i1t} \leq \delta(p_{1t}, z_{1t}, \xi_{1t})\big) \\ 1 - a_1 =& P\big(\varepsilon_{ikt} \leq -\delta(p_{kt}, z_{kt}, \xi_{kt}) \ \forall k\big) \end{aligned}$$

and it follows that $P\big(-\varepsilon_{i1t} \leq \delta(p_{1t}, z_{1t}, \xi_{1t})\big) = a_1$.[36] Since the marginal distribution of ε_{i1t} is known, we have

$$\begin{aligned} P\big(\varepsilon_{i1t} \geq -\delta(p_{1t}, z_{1t}, \xi_{1t})\big) &= a_1 \\ 1 - F_{\varepsilon_{i1t}}(-\delta(p_{1t}, z_{1t}, \xi_{1t})) &= a_1 \\ F_{\varepsilon_{i1t}}(-\delta(p_{1t}, z_{1t}, \xi_{1t})) &= 1 - a_1 \\ \delta(p_{1t}, z_{1t}, \xi_{1t}) &= -F_{\varepsilon_{i1t}}^{-1}(1 - a_1) \end{aligned}$$

for all $(p_t, z_t, \xi_t) \in S_a$. Repeat this for all $a_1 \in [0, 1]$ to identify δ at all points on the inverse image of $-E_1$. Once δ is identified, it is easy to identify σ from δ and S. Pick

[36] Note that $P(A) \geq P(A \cap B)$ and $1 - P(A) = P(A^c) \geq P(A^c \cap C)$. If $P(A \cap B) = a$ and $P(A^c \cap C) = 1 - a$ for some B and C, then it must be that $P(A) = a$.

any $\delta_t \in -E_1^J$, and find $(p_t, z_t, \xi_t) \in \mathbb{R}_+^J \times \mathbb{R}^{2J}$ such that $\delta(p_{jt}, z_{jt}, \xi_{jt}) = \delta_{jt}$ for all j. Then $\sigma(\delta_t) = S(p_t, z_t, \xi_t)$ identifies σ on $-E_1^J$. Now F_ε is identified on E_1^J from σ by Lemma 1. ∎

Lemma 3 *Suppose Assumptions 1-4 hold. If* $\{\xi_t\}_{t=1}^T$ *are given and the utility function* $\delta(p_t, z_t, \xi_t)$ *is known on* $D \subset \mathbb{R}_+ \times \mathbb{R}^2$*, the joint density function* $f_\varepsilon^{t,t-1}$ *is identified on* $-\delta(D)^{2J}$*, where* $\delta(D)$ *stands for the image of* δ *on the domain* D*.*

Proof. The churn rate is defined by

$$h_{jt} = P(j \text{ is not chosen at } t | j \text{ was chosen at } t-1)$$

so

$$\begin{aligned} 1 - h_{jt} =& P(j \text{ is chosen at } t | j \text{ was chosen at } t-1) \\ =& P(u_{ijt} \geq u_{ikt} \ \forall k | u_{ij,t-1} \geq u_{ik,t-1} \ \forall k) \\ =& P(\varepsilon_{ikt} - \varepsilon_{ijt} \leq \delta(p_{jt}, z_{jt}, \xi_{jt}) - \delta(p_{kt}, z_{kt}, \xi_{kt}) \ \forall k \text{ and } -\varepsilon_{ijt} \leq \delta(p_{jt}, z_{jt}, \xi_{jt}) | \\ & \quad \varepsilon_{ik,t-1} - \varepsilon_{ij,t-1} \leq \delta(p_{j,t-1}, z_{j,t-1}, \xi_{j,t-1}) - \delta(p_{k,t-1}, z_{k,t-1}, \xi_{k,t-1}) \ \forall k \\ & \quad \text{and } -\varepsilon_{ij,t-1} \leq \delta(p_{j,t-1}, z_{j,t-1}, \xi_{j,t-1})) \end{aligned}$$

Pick $e = (e_t', e_{t-1}')' \in -\delta(D)^{2J}$ and find $(p_t, z_t, \xi_t, p_{t-1}, z_{t-1}, \xi_{t-1})$ such that

$$\begin{aligned} -e_{1t} =& \delta(p_{1t}, z_{1t}, \xi_{1t}) \\ & \vdots \\ -e_{Jt} =& \delta(p_{Jt}, z_{Jt}, \xi_{Jt}) \\ -e_{1,t-1} =& \delta(p_{1,t-1}, z_{1,t-1}, \xi_{1,t-1}) \\ & \vdots \\ -e_{J,t-1} =& \delta(p_{J,t-1}, z_{J,t-1}, \xi_{J,t-1}) \end{aligned}$$

Using h_{1t} and $s_{1,t-1}$ which correspond to $(p_t, z_t, \xi_t, p_{t-1}, z_{t-1}, \xi_{t-1})$, we can write

$$\begin{aligned} (1 - h_{1t}) s_{1,t-1} =& P(\delta(p_{1\tau}, z_{1\tau}, \xi_{1\tau}) + \varepsilon_{i1\tau} \geq 0 \text{ and} \\ & \quad \delta(p_{1\tau}, z_{1\tau}, \xi_{1\tau}) + \varepsilon_{i1\tau} \geq \delta(p_{k\tau}, z_{k\tau}, \xi_{k\tau}) + \varepsilon_{ik\tau} \ \forall k, \tau = t, t-1) \\ =& P(-e_{1\tau} + \varepsilon_{i1\tau} \geq 0, -e_{1\tau} + \varepsilon_{i1\tau} \geq -e_{k\tau} + \varepsilon_{ik\tau} \ \forall k \geq 2, \tau = t, t-1) \\ =& P(-\varepsilon_{i1\tau} \leq -e_{1\tau}, \varepsilon_{ik\tau} - \varepsilon_{i1\tau} \leq e_{k\tau} - e_{1\tau} \ \forall k \geq 2, \tau = t, t-1) \end{aligned} \tag{21}$$

Define

$$b := (-\varepsilon_{i1t}, \varepsilon_{i2t} - \varepsilon_{i1t}, \cdots, \varepsilon_{iJt} - \varepsilon_{i1t}, -\varepsilon_{i1,t-1}, \varepsilon_{i2,t-1} - \varepsilon_{i1,t-1}, \cdots, \varepsilon_{iJ,t-1} - \varepsilon_{i1,t-1})$$

then we have

$$b = \begin{pmatrix} -1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \cdots & 0 & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \ddots & \vdots \\ -1 & 0 & \cdots & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & -1 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_{i1t} \\ \varepsilon_{i2t} \\ \vdots \\ \varepsilon_{iJt} \\ \varepsilon_{i1,t-1} \\ \varepsilon_{i2,t-1} \\ \vdots \\ \varepsilon_{iJ,t-1} \end{pmatrix} =: B \begin{pmatrix} \varepsilon_{it} \\ \varepsilon_{i,t-1} \end{pmatrix} \qquad (22)$$

Let F_b be the joint distribution of b and f_b be its density function, and we can rewrite (21) using F_b and B as follows.

$$(1 - h_{1t})s_{1,t-1} = P(b \leq Be) = F_b(Be)$$

$|B| = 1$ implies that the relationship (22) between b and $(\varepsilon_{it}, \varepsilon_{i,t-1})$ is globally one-to-one. Therefore we have by the transformation of variables formula that

$$f_\varepsilon^{t,t-1}(e) = f_b(Be)|B| = f_b(Be)$$

Since $e \in -\delta(D)^{2J}$ was arbitrarily chosen, the above equation identifies $f_\varepsilon^{t,t-1}$ on $-\delta(D)^{2J}$. ∎

Lemma 4 *Suppose Assumptions 2, 15 and 16 hold so that the equations for the market shares and the churn rates are given by* (19) *and* (20)*, which are for* $j = 1, \cdots, J$,

$$s_{jt} = \sum_{k=1}^{J} \sigma_j(\delta_{kt}, \delta_{qt} - c_{kt}\ \forall q \neq k | u_{i0t} = -c_{kt}) s_{k,t-1} + \sigma_j(\delta_t) s_{0,t-1}$$

$$1 - h_{jt} = \sigma_j(\delta_{jt}, \delta_{qt} - c_{jt}\ \forall q \neq j | u_{i0t} = -c_{jt})$$

There is a unique set of $\{(\delta_t, c_t)\}_{t=2}^{T}$ *satisfying the above equations.*

Proof. Suppose that two different sets of $\{(\delta_t, c_t)\}_{t=2}^T$ and $\{(\widetilde{\delta}_t, \widetilde{c}_t)\}_{t=2}^T$ satisfy equations (19) and (20). The proof consists of two steps. In the first step, I show it cannot be the case that $c_t \neq \widetilde{c}_t$ for some t. To see this, suppose by way of contradiction that $c_t \neq \widetilde{c}_t$ for some t. There are three subcases.

The first one is when $c_{jt} - \widetilde{c}_{jt}$ is positive for all j or negative for all j. Let $c_{jt} > \widetilde{c}_{jt}$ for all j without loss of generality and $\delta_{lt} - \widetilde{\delta}_{lt} =: a$ be the largest value among $\{\delta_{qt} - \widetilde{\delta}_{qt} : q = 1, \cdots, J\}$. (20) yields

$$\begin{aligned}
\sigma_l(\delta_{lt} + c_{lt}, \delta_{qt}\ \forall q \neq l | u_{i0t} = 0) &= \sigma_l(\delta_{lt}, \delta_{qt} - c_{lt}\ \forall q \neq l | u_{i0t} = -c_{lt}) \\
&= 1 - h_{lt} \\
&= \sigma_l(\widetilde{\delta}_{lt}, \widetilde{\delta}_{qt} - \widetilde{c}_{lt}\ \forall q \neq l | u_{i0t} = -\widetilde{c}_{lt}) \\
&= \sigma_l(\widetilde{\delta}_{lt} + a + \widetilde{c}_{lt}, \widetilde{\delta}_{qt} + a\ \forall q \neq l | u_{i0t} = a) \\
&= \sigma_l(\delta_{lt} + \widetilde{c}_{lt}, \widetilde{\delta}_{qt} + a\ \forall q \neq l | u_{i0t} = a)
\end{aligned}$$

Since $c_{lt} > \widetilde{c}_{lt}$ and $\delta_{qt} - \widetilde{\delta}_{qt} \leq a$, or equivalently $\delta_{qt} \leq \widetilde{\delta}_{qt} + a$ for all q, the strict monotonicity of σ_l implies that $a < 0$. Now take the minimum $\delta_{mt} - \widetilde{\delta}_{mt} =: b$ among $\{\delta_{qt} - \widetilde{\delta}_{qt} : q = 1, \cdots, J\}$, then obviously $b < 0$. Therefore, we have for all $k \neq m$,

$$\begin{aligned}
\sigma_m(\delta_{kt}, \delta_{qt} - c_{kt}\ \forall q \neq k | u_{i0t} = -c_{kt}) &= \sigma_m(\delta_{kt} - b + c_{kt}, \delta_{qt} - b\ \forall q \neq k | u_{i0t} = -b) \\
&< \sigma_m(\widetilde{\delta}_{kt} + \widetilde{c}_{kt}, \widetilde{\delta}_{qt}\ \forall q \neq k | u_{i0t} = 0) \\
&= \sigma_m(\widetilde{\delta}_{kt}, \widetilde{\delta}_{qt} - \widetilde{c}_{kt}\ \forall q \neq k | u_{i0t} = -\widetilde{c}_{kt})
\end{aligned}$$

Note that the inequality holds since $\delta_{qt} - \widetilde{\delta}_{qt} \geq b$, or equivalently $\delta_{qt} - b \geq \widetilde{\delta}_{qt}$ for all q, $c_{kt} \geq \widetilde{c}_{kt}$, $-b > 0$, but $\delta_{mt} - b = \widetilde{\delta}_{mt}$, and σ_m is strictly decreasing in the arguments other than the m-th one. Also

$$\sigma_m(\delta_t) = \sigma_m(\delta_t - b | u_{i0t} = -b) < \sigma_m(\widetilde{\delta}_t)$$

for the same reason, and finally

$$\begin{aligned}
\sigma_m(\delta_{mt}, \delta_{qt} - c_{mt}\ \forall q \neq m | u_{i0t} = -c_{mt}) &= 1 - h_{mt} \\
&= \sigma_m(\widetilde{\delta}_{mt}, \widetilde{\delta}_{qt} - \widetilde{c}_{mt}\ \forall q \neq m | u_{i0t} = -\widetilde{c}_{mt})
\end{aligned}$$

Since $s_{k,t-1} \geq 0$ for all $k = 0, \cdots, J$, we have

$$\sum_{k=1}^{J} \sigma_m(\delta_{kt}, \delta_{qt} - c_{kt} \ \forall q \neq k | u_{i0t} = -c_{kt}) s_{k,t-1} + \sigma_m(\delta_t) s_{0,t-1}$$
$$< \sum_{k=1}^{J} \sigma_m(\widetilde{\delta}_{kt}, \widetilde{\delta}_{qt} - \widetilde{c}_{kt} \ \forall q \neq k | u_{i0t} = -\widetilde{c}_{kt}) s_{k,t-1} + \sigma_m(\widetilde{\delta}_t) s_{0,t-1}$$

which violates (19) for $j = m$.

The second subcase is when $c_{jt} - \widetilde{c}_{jt}$ is positive for some j but non-positive for other j, and $\delta_{qt} - \widetilde{\delta}_{qt}$ is positive for some q. Define a set $J_+ := \{j : c_{jt} - \widetilde{c}_{jt} > 0\}$, and let $\delta_{lt} - \widetilde{\delta}_{lt} =: a > 0$ be the largest among $\{\delta_{qt} - \widetilde{\delta}_{qt} : q = 1, \cdots, J\}$. I claim that $l \notin J_+$. To see this, recall that

$$\begin{aligned} \sigma_l(\delta_{lt} + c_{lt}, \delta_{qt} \ \forall q \neq l | u_{i0t} = 0) &= 1 - h_{lt} \\ &= \sigma_l(\widetilde{\delta}_{lt} + \widetilde{c}_{lt}, \widetilde{\delta}_{qt} \ \forall q \neq l | u_{i0t} = 0) \\ &= \sigma_l(\delta_{lt} + \widetilde{c}_{lt}, \widetilde{\delta}_{qt} + a \ \forall q \neq l | u_{i0t} = a) \end{aligned}$$

Since $\delta_{qt} - \widetilde{\delta}_{qt} \leq a$, or equivalently $\delta_{qt} \leq \widetilde{\delta}_{qt} + a$ for all q, and $a > 0$, the strict monotonicity of σ_l implies that $c_{lt} < \widetilde{c}_{lt}$. I next claim that $\delta_{kt} - a + c_{kt} \leq \widetilde{\delta}_{kt} + \widetilde{c}_{kt}$ for all $k \neq l$. To see this, note that for any $k \in J_+$,

$$\begin{aligned} \sigma_k(\delta_{kt} + c_{kt}, \delta_{qt} \ \forall q \neq k | u_{i0t} = 0) &= 1 - h_{kt} \\ &= \sigma_k(\widetilde{\delta}_{kt} + \widetilde{c}_{kt}, \widetilde{\delta}_{qt} \ \forall q \neq k | u_{i0t} = 0) \\ &= \sigma_k(\widetilde{\delta}_{kt} + a + \widetilde{c}_{kt}, \widetilde{\delta}_{qt} + a \ \forall q \neq k | u_{i0t} = a) \end{aligned}$$

and that $\delta_{jt} \leq \widetilde{\delta}_{jt} + a$ for all j and $a > 0$. So the monotonicity of σ_k implies that $\delta_{kt} + c_{kt} < \widetilde{\delta}_{kt} + a + \widetilde{c}_{kt}$, or equivalently $\delta_{kt} - a + c_{kt} < \widetilde{\delta}_{kt} + \widetilde{c}_{kt}$. For any $k \notin J_+ \cup \{l\}$,

$$(\delta_{kt} - a + c_{kt}) - (\widetilde{\delta}_{kt} + \widetilde{c}_{kt}) = (\delta_{kt} - \widetilde{\delta}_{kt} - a) + (c_{kt} - \widetilde{c}_{kt}) \leq 0$$

since $\delta_{kt} - \widetilde{\delta}_{kt} \leq a$ and $c_{kt} \leq \widetilde{c}_{kt}$ for such k. Therefore, we have for all $k \neq l$,

$$\begin{aligned} \sigma_l(\delta_{kt}, \delta_{qt} - c_{kt} \ \forall q \neq k | u_{i0t} = -c_{kt}) &= \sigma_l(\delta_{kt} - a + c_{kt}, \delta_{qt} - a \ \forall q \neq k | u_{i0t} = -a) \\ &> \sigma_l(\widetilde{\delta}_{kt} + \widetilde{c}_{kt}, \widetilde{\delta}_{qt} \ \forall q \neq k | u_{i0t} = 0) \\ &= \sigma_l(\widetilde{\delta}_{kt}, \widetilde{\delta}_{qt} - \widetilde{c}_{kt} \ \forall q \neq k | u_{i0t} = -\widetilde{c}_{kt}) \end{aligned}$$

Note that the inequality holds since $\delta_{kt} - a + c_{kt} \leq \widetilde{\delta}_{kt} + \widetilde{c}_{kt}$, $\delta_{qt} - \widetilde{\delta}_{qt} \leq a$, or equivalently $\delta_{qt} - a \leq \widetilde{\delta}_{qt}$ for all q, $-a < 0$, but $\delta_{lt} - a = \widetilde{\delta}_{lt}$, and σ_l is strictly decreasing in all arguments other than the l-th one. Also

$$\sigma_l(\delta_t) = \sigma_l(\delta_t - a | u_{i0t} = -a) > \sigma_l(\widetilde{\delta}_t)$$

for the same reason, and finally

$$\sigma_l(\delta_{lt}, \delta_{qt} - c_{lt} \ \forall q \neq l | u_{i0t} = -c_{lt}) = 1 - h_{lt} = \sigma_l(\widetilde{\delta}_{lt}, \widetilde{\delta}_{qt} - \widetilde{c}_{lt} \ \forall q \neq l | u_{i0t} = -\widetilde{c}_{lt})$$

Since $s_{k,t-1} \geq 0$ for all $k = 0, \cdots, J$, we have

$$\begin{aligned} &\sum_{k=1}^{J} \sigma_l(\delta_{kt}, \delta_{qt} - c_{kt} \ \forall q \neq k | u_{i0t} = -c_{kt}) s_{k,t-1} + \sigma_l(\delta_t) s_{0,t-1} \\ &\qquad > \sum_{k=1}^{J} \sigma_l(\widetilde{\delta}_{kt}, \widetilde{\delta}_{qt} - \widetilde{c}_{kt} \ \forall q \neq k | u_{i0t} = -\widetilde{c}_{kt}) s_{k,t-1} + \sigma_l(\widetilde{\delta}_t) s_{0,t-1} \end{aligned}$$

which violates (19) for $j = l$.

The third subcase is when $c_{jt} - \widetilde{c}_{jt}$ is positive for some j but non-positive for other j, and $\delta_{qt} - \widetilde{\delta}_{qt}$ is non-positive for all q. Define $J_+ := \{j : c_{jt} - \widetilde{c}_{jt} > 0\}$ in the same way. It cannot be the case that $\delta_{qt} - \widetilde{\delta}_{qt} = 0$ for all q, since if it is true, for some $k \in J_+$, (20) implies that

$$\begin{aligned} \sigma_k(\delta_{kt} + c_{kt}, \delta_{qt} \ \forall q \neq k | u_{i0t} = 0) &= 1 - h_{kt} \\ &= \sigma_k(\widetilde{\delta}_{kt} + \widetilde{c}_{kt}, \widetilde{\delta}_{qt} \ \forall q \neq k | u_{i0t} = 0) \\ &= \sigma_k(\delta_{kt} + \widetilde{c}_{kt}, \delta_{qt} \ \forall q \neq k | u_{i0t} = 0) \end{aligned}$$

which contradicts the strict monotonicity of σ_k in its k-th argument. Hence $\delta_{qt} - \widetilde{\delta}_{qt} < 0$ for some q. Take the minimum $\delta_{mt} - \widetilde{\delta}_{mt} =: b < 0$ among those values. Note that

$$\begin{aligned} \sigma_m(\delta_{mt} + c_{mt}, \delta_{qt} \ \forall q \neq m | u_{i0t} = 0) &= 1 - h_{mt} \\ &= \sigma_m(\widetilde{\delta}_{mt} + \widetilde{c}_{mt}, \widetilde{\delta}_{qt} \ \forall q \neq m | u_{i0t} = 0) \\ &= \sigma_m(\delta_{mt} + \widetilde{c}_{mt}, \widetilde{\delta}_{qt} + b \ \forall q \neq m | u_{i0t} = b) \end{aligned}$$

Since $\delta_{qt} - \widetilde{\delta}_{qt} \geq b$, or equivalently $\delta_{qt} \geq \widetilde{\delta}_{qt} + b$ for all q, and $b < 0$, the strict monotonicity of σ_m implies that $c_{mt} > \widetilde{c}_{mt}$, and thus $m \in J_+$. I next claim that $\delta_{kt} - b + c_{kt} \geq \widetilde{\delta}_{kt} + \widetilde{c}_{kt}$

for all $k \neq m$. To see this, note that for any $k \notin J_+$,

$$\begin{aligned}\sigma_k(\delta_{kt} + c_{kt}, \delta_{qt}\ \forall q \neq k | u_{i0t} = 0) &= 1 - h_{kt} \\ &= \sigma_k(\widetilde{\delta}_{kt} + \widetilde{c}_{kt}, \widetilde{\delta}_{qt}\ \forall q \neq k | u_{i0t} = 0) \\ &= \sigma_k(\widetilde{\delta}_{kt} + b + \widetilde{c}_{kt}, \widetilde{\delta}_{qt} + b\ \forall q \neq k | u_{i0t} = b)\end{aligned}$$

and that $\delta_{qt} \geq \widetilde{\delta}_{qt} + b$ for all q and $b < 0$. So the monotonicity of σ_k implies that $\delta_{kt} + c_{kt} > \widetilde{\delta}_{kt} + b + \widetilde{c}_{kt}$, or equivalently $\delta_{kt} - b + c_{kt} > \widetilde{\delta}_{kt} + \widetilde{c}_{kt}$. It is obvious that for $k \in J_+ \setminus \{m\}$,

$$(\delta_{kt} - b + c_{kt}) - (\widetilde{\delta}_{kt} + \widetilde{c}_{kt}) = (\delta_{kt} - \widetilde{\delta}_{kt} - b) + (c_{kt} - \widetilde{c}_{kt}) \geq 0$$

since $\delta_{kt} - \widetilde{\delta}_{kt} \geq b$ and $c_{kt} \geq \widetilde{c}_{kt}$ for such k. Therefore we have for all $k \neq m$,

$$\begin{aligned}\sigma_m(\delta_{kt}, \delta_{qt} - c_{kt}\ \forall q \neq k | u_{i0t} = -c_{kt}) &= \sigma_m(\delta_{kt} - b + c_{kt}, \delta_{qt} - b\ \forall q \neq k | u_{i0t} = -b) \\ &< \sigma_m(\widetilde{\delta}_{kt} + \widetilde{c}_{kt}, \widetilde{\delta}_{qt}\ \forall q \neq k | u_{i0t} = 0) \\ &= \sigma_m(\widetilde{\delta}_{kt}, \widetilde{\delta}_{qt} - \widetilde{c}_{kt}\ \forall q \neq k | u_{i0t} = -\widetilde{c}_{kt})\end{aligned}$$

Note that the inequality holds since $\delta_{kt} - b + c_{kt} \geq \widetilde{\delta}_{kt} + \widetilde{c}_{kt}$, $\delta_{qt} - \widetilde{\delta}_{qt} \geq b$, or equivalently $\delta_{qt} - b \geq \widetilde{\delta}_{qt}$ for all q, $-b > 0$, but $\delta_{mt} - b = \widetilde{\delta}_{mt}$, and σ_m is strictly decreasing in all arguments other than the m-th one. Also

$$\sigma_m(\delta_t) = \sigma_m(\delta_t - b | u_{i0t} = -b) < \sigma_m(\widetilde{\delta}_t)$$

for the same reason, and finally

$$\begin{aligned}\sigma_m(\delta_{mt}, \delta_{qt} - c_{mt}\ \forall q \neq m | u_{i0t} = -c_{mt}) &= 1 - h_{mt} \\ &= \sigma_m(\widetilde{\delta}_{mt}, \widetilde{\delta}_{qt} - \widetilde{c}_{mt}\ \forall q \neq m | u_{i0t} = -\widetilde{c}_{mt})\end{aligned}$$

Since $s_{k,t-1} \geq 0$ for all $k = 0, \cdots, J$, we have

$$\begin{aligned}&\sum_{k=1}^{J} \sigma_m(\delta_{kt}, \delta_{qt} - c_{kt}\ \forall q \neq k | u_{i0t} = -c_{kt}) s_{k,t-1} + \sigma_m(\delta_t) s_{0,t-1} \\ &\quad < \sum_{k=1}^{J} \sigma_m(\widetilde{\delta}_{kt}, \widetilde{\delta}_{qt} - \widetilde{c}_{kt}\ \forall q \neq k | u_{i0t} = -\widetilde{c}_{kt}) s_{k,t-1} + \sigma_m(\widetilde{\delta}_t) s_{0,t-1}\end{aligned}$$

which violates (19) for $j = m$.

Therefore we have $c_t = \widetilde{c}_t$ for all t. In the next step, I show it cannot be the case that $\delta_t \neq \widetilde{\delta}_t$ for some t given $c_t = \widetilde{c}_t$ for all t. Suppose by way of contradiction that $\delta_{qt} - \widetilde{\delta}_{qt} \neq 0$ for some q and t. Assume $\delta_{qt} - \widetilde{\delta}_{qt} > 0$ for some q without loss of generality. Let $\delta_{lt} - \widetilde{\delta}_{lt} =: a > 0$ be the largest value among $\{\delta_{qt} - \widetilde{\delta}_{qt} : j = 1, \cdots, J\}$. (20) implies that

$$\begin{aligned}
1 - h_{lt} &= \sigma_l(\widetilde{\delta}_{lt} + \widetilde{c}_{lt}, \widetilde{\delta}_{qt}\ \forall q \neq l | u_{i0t} = 0) \\
&= \sigma_l(\widetilde{\delta}_{lt} + a + c_{lt}, \widetilde{\delta}_{qt} + a\ \forall q \neq l | u_{i0t} = a) \\
&> \sigma_l(\delta_{lt} + c_{lt}, \delta_{qt}\ \forall q \neq l | u_{i0t} = 0) = 1 - h_{lt}
\end{aligned}$$

where the second equality holds by $\widetilde{c}_{lt} = c_{lt}$, and the inequality holds since $a \geq \delta_{qt} - \widetilde{\delta}_{qt}$, or equivalently $\widetilde{\delta}_{qt} + a \geq \delta_{qt}$ for all j, $a > 0$, but $\widetilde{\delta}_{lt} + a = \delta_{lt}$, and σ_l is strictly increasing in its l-th argument, and strictly decreasing in the others. There is a contradiction, and thus we can conclude that $\delta_t = \widetilde{\delta}_t$ for all t. ∎

Lemma 5 *Suppose Assumptions 1-3 and 18 hold. Let $\mathcal{S}$ be the convex and symmetric support of (p_t, z_t, ξ_t) and $\mathcal{S}_1 := \{(p_{1t}, z_{1t}, \xi_{1t}) : (p_t, z_t, \xi_t) \in \mathcal{S}\}$. If $\{\xi_t\}_{t=1}^T$ are given and the function $S(p_t, z_t, \xi_t)$ defined by (7) is known, the utility function $\delta(p_{jt}, z_{jt}, \xi_{jt})$ is identified on the set $\mathcal{S}_1$, and the joint distribution F_ε is identified on the set $-\delta(\mathcal{S}_1)^J$.*

Proof. Pick $(p_{1t}, z_{1t}, \xi_{1t}) \in \mathcal{S}_1$, and let (p_t, z_t, ξ_t) be such that $(p_{jt}, z_{jt}, \xi_{jt}) = (\overline{p}, \overline{z}, \overline{\xi})$ for all $j \neq 1$. Assumption 18 (i) implies that

$$\begin{aligned}
s_t =& S(p_t, z_t, \xi_t) \\
=& \sigma(\delta(p_{1t}, z_{1t}, \xi_{1t}), \delta(p_{2t}, z_{2t}, \xi_{2t}), \cdots, \delta(p_{Jt}, z_{Jt}, \xi_{Jt})) \\
=& \sigma(\delta(p_{1t}, z_{1t}, \xi_{1t}), \overline{\delta}, \cdots, \overline{\delta})
\end{aligned}$$

Since σ_1 is known at such points by Assumption 18 (ii), we have

$$\delta(p_{1t}, z_{1t}, \xi_{1t}) = \sigma_1^{-1}(s_{1t} | \delta_j = \overline{\delta}\ \forall j \geq 2)$$

As the choice of $(p_{1t}, z_{1t}, \xi_{1t})$ is arbitrary, δ is identified on $\mathcal{S}_1$. Once δ is identified, we can follow a method employed in the proof of Lemma 2 to identify σ from δ and S. Pick any $\delta_t \in \delta(\mathcal{S}_1)^J$, and find $(p_t, z_t, \xi_t) \in \mathcal{S}$ such that $\delta(p_{jt}, z_{jt}, \xi_{jt}) = \delta_{jt}$ for all j. Then $\sigma(\delta_t) = S(p_t, z_t, \xi_t)$ identifies σ on $\delta(\mathcal{S}_1)^J$. Now F_ε is identified on $-\delta(\mathcal{S}_1)^J$ from σ by Lemma 1. ∎

C. Tables and Figures

TABLE 1. Results from the logit model

Variables	i OLS	ii 2SLS	iii GMM	iv OLS	v 2SLS	vi GMM
t	0.0751	0.0624	0.0627	0.0536	0.0487	0.0511
White St.Err.	(0.0047)	(0.0047)	(0.0045)	(0.0033)	(0.0032)	(0.0031)
HAC St.Err.	(0.0086)	(0.0083)	(0.0079)	(0.0054)	(0.0050)	(0.0046)
p	−0.0263	−0.0901	−0.0915	−0.0258	−0.0505	−0.0560
White St.Err.	(0.0049)	(0.0132)	(0.0120)	(0.0040)	(0.0077)	(0.0074)
HAC St.Err.	(0.0092)	(0.0239)	(0.0215)	(0.0068)	(0.0126)	(0.0118)
constant	−2.0135	1.8080	1.8788	−3.8638	−2.3879	−2.1436
White St.Err.	(0.3613)	(0.6943)	(0.6353)	(0.2547)	(0.4124)	(0.4077)
HAC St.Err.	(0.6788)	(1.2367)	(1.1237)	(0.4333)	(0.6577)	(0.6366)
switching costs	not used	not used	not used	included	included	included
R^2	0.7535	0.4680	0.4553	0.7640	0.6872	0.6455

TABLE 2. Results from other models

Variables	i Nested Logit	ii Nested Logit (naïve)	iii Nested Logit (2SLS)	iv Nested Logit (GMM)	v Random coefft. on group dummy
t	0.0544	0.0070	0.0487	0.0511	0.0511
White St.Err	(0.0319)	(0.3292)	(0.0032)	(0.0035)	(0.0031)
HAC St.Err.	(0.0494)	(0.5102)	(0.0050)	(0.0057)	(0.0046)
p	0.0826	0.8501	−0.0505	−0.0560	−0.0560
White St.Err	(0.6642)	(6.8716)	(0.0077)	(0.0074)	(0.0074)
HAC St.Err.	(1.0069)	(10.4301)	(0.0126)	(0.0119)	(0.0118)
constant	−4.3477	−34.4913	−2.3880	−2.1436	−2.1436
White St.Err	(23.7596)	(245.8417)	(0.4124)	(0.4079)	(0.4077)
HAC St.Err.	(36.0452)	(373.3842)	(0.6576)	(0.6397)	(0.6366)
σ	1.9180	10.0029	0.0000	0.0000	0.0000
White St.Err	(7.3173)	(75.7090)	(0.6724)	(0.0235)	(8.3035)
HAC St.Err.	(11.0701)	(114.6779)	(0.9591)	(0.0337)	(11.8448)
switching costs	not used	included	included	included	included
R^2	0.5485	−86.2598	0.6872	0.6455	0.6455

FIGURE 1. The actual churn rates vs
the implied ones by zero switching costs

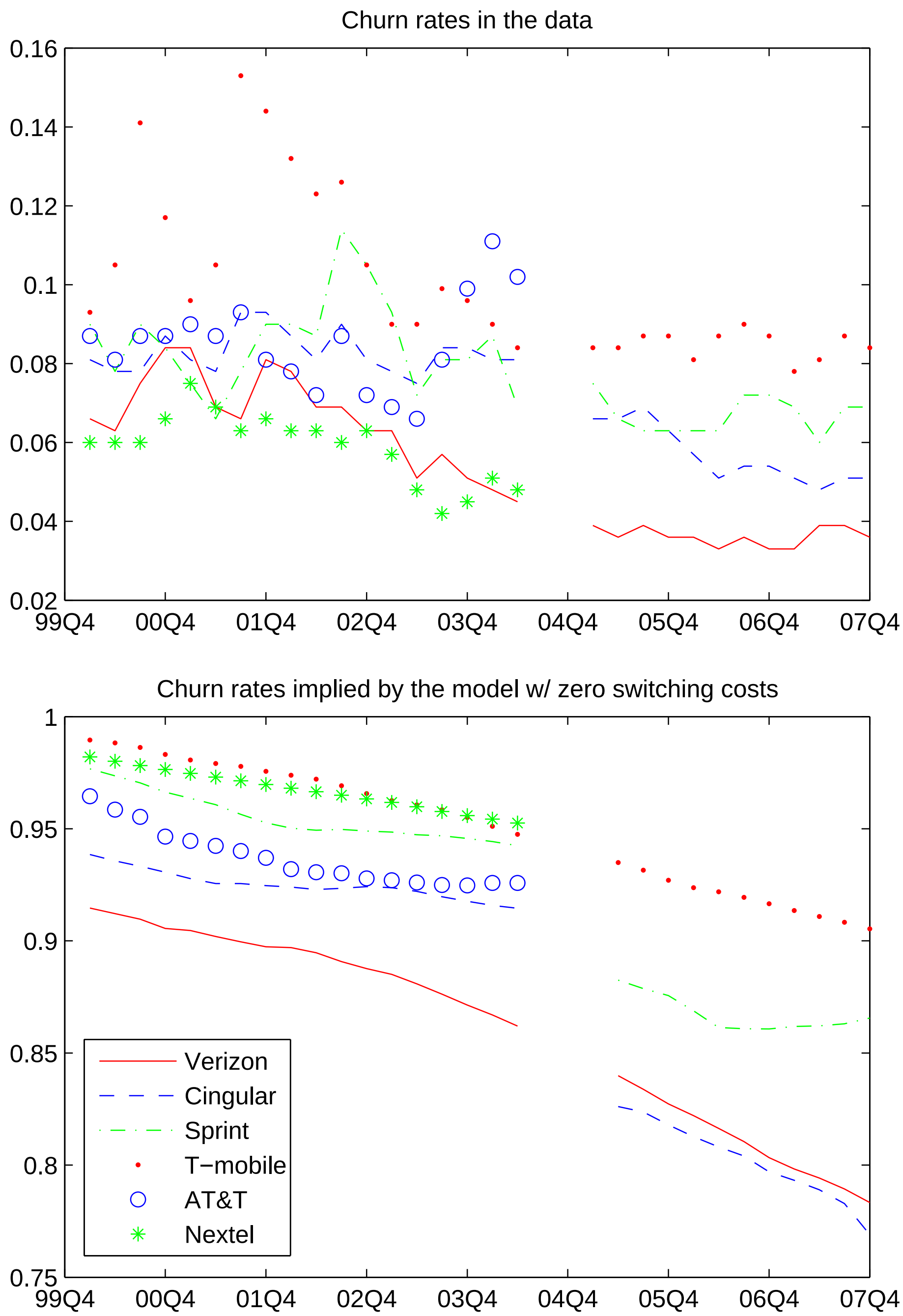

FIGURE 2. The mean utilities δ_{jt}
without and with switching costs

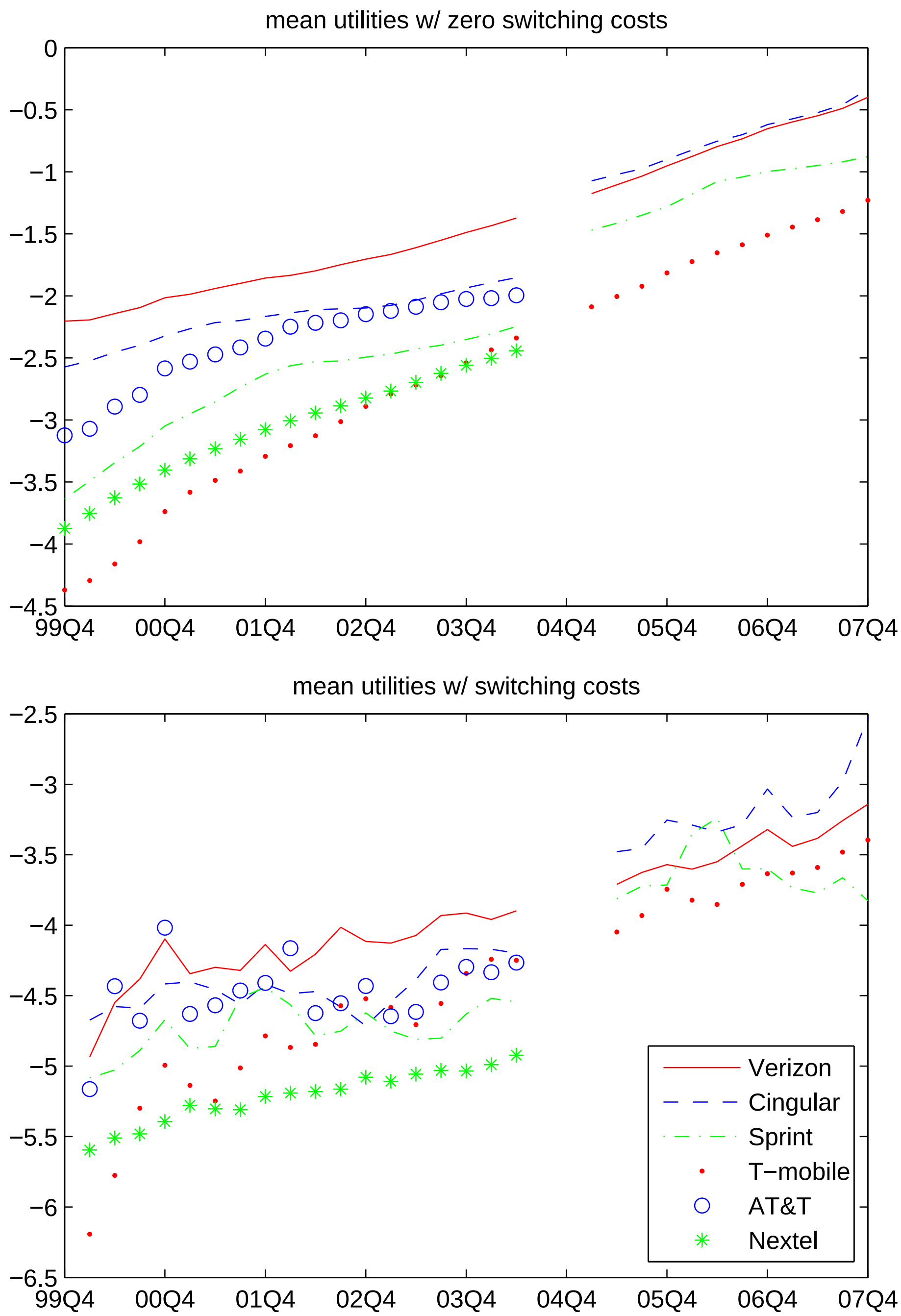

FIGURE 3. The switching costs consistent with the churn rates