Unobserved Heterogeneity, State Dependence, and Health Plan Choices.

Ariel Pakes, Jack Porter, Mark Shepard, and Sophie Calder-Wang *

July 1, 2025

Abstract

We provide a new method to analyze discrete choice models with state dependence and individual-by-product fixed effects, and use it to analyze consumer choices in the policy-relevant environment of a subsidized health insurance exchange. Moment inequalities are used to distinguish state dependence from consumers' switching choices in response to changes in product attributes. We infer much smaller switching costs on the health insurance exchange than is inferred from standard logit and/or random effects methods. A counterfactual policy evaluation illustrates that the policy implications of this difference can be substantive.

Keywords: health insurance market, state dependence, fixed effects, discrete choice, partial identification

*Pakes: Harvard University, apakes@fas.harvard.edu. Porter: University of Wisconsin-Madison, jrporter@ssc.wisc.edu. Shepard: Harvard University, mark_shepard@hks.harvard.edu. Calder-Wang: University of Pennsylvania, sophiecw@wharton.upenn.edu. We thank Hanbin Yang and Xinmei Yang for truly outstanding research assistance, and Liran Einav, Jeremy Fox, Ben Handel, and Amanda Starc for comments. We also acknowledge the Massachusetts Health Connector (and especially Marissa Woltmann) for help in providing and interpreting the data. We gratefully acknowledge data funding from Harvard's Lab for Economic Applications and Policy.

1 Introduction

This paper develops a new method to estimate state dependence in a choice model that allows for flexible unobserved heterogeneity through individual-by-product fixed effects. We apply the method to unravelling the sources of persistence in health insurance plan choices, an issue which has led to considerable policy debate. Our method is nonparametric, but we also consider parametric models. We then compare the empirical results from the models that allow for individual-by-product fixed effects to a familiar set of parametric models that do not. The models with fixed effects find an upper bound to switching costs that is considerably lower than the estimates from models that do not allow for fixed effects and a counterfactual indicates that the difference is likely to have substantial implications for the analysis of price effects.

Distinguishing the impacts of unobserved heterogeneity from those of state dependence in discrete choice models has implications for the interpretation and policy implications of many observed phenomena.[1] Both the increase in the availability of panel data discrete choice data sets, and the recognition that inattention and/or switching costs have dynamic implications, have made the distinction even more salient in empirical work. Inattention is likely particularly relevant for choices that can be reversed if needed, and/or where repeated evaluations of alternative choices have cognitive costs that rival implied utility gains. So it is not surprising that so much attention has been paid to the impact of state dependence in health insurance choices.[2]

Addressing inattention requires a model which conditions both on past choices and on preferences. Our empirical analysis asks whether past choices (their "state") or unobserved preferences underlies the low price responsiveness that has been found in health insurance choice. This question is both economically important and policy-relevant. Governments set rules for

[1]The analysis of unemployment durations seeks to separate out the causal effects of being unemployed on future employment from unobserved heterogeneity in worker employability (see Kroft, Lange, and Notowidigdo (2013) and the articles cited therein). Both the marketing and I.O. literatures face the problem of distinguishing switching costs from unobserved preferences in explaining the constancy of individual purchasing patterns over time (see the review by Keane (1997)). Network models often need to distinguish between common preferences and the causal effects of the network (see for example, Conley and Udry (2010)), A similar problem arises in distinguishing the effects of moral hazard from adverse selection in evaluating policies designed to monitor behavior in insurance markets (Abbring, Heckman, Chiappori, and Pinquet, 2003)

[2]Directly relevant here is the literature on *mechanisms* behind state dependence in health insurance choices, distinguishing factors like search costs, rational inattention, and true switching hassles (Heiss, McFadden, Winter, Wuppermann, and Zhou, 2021; Abaluck and Adams-Prassl, 2021; Brown and Jeon, 2024; Brot-Goldberg, Layton, Vabson, and Wang, 2023).

market-based health insurance programs in the Affordable Care Act exchanges, Medicare Part D, and Medicaid managed care that cover more than 100 million people and cost over $750 billion in public spending per annum in the U.S. alone. Recent applied work suggests that choice persistence driven by state dependence (or "switching costs") may lead to larger insurance markups (Ho, Hogan, and Scott Morton, 2017), may interact with problems created by adverse selection (Handel, 2013; Polyakova, 2016), and may lead to invest-then-harvest pricing dynamics (Ericson, 2014). It is unsurprising, then, that regulators often seek to encourage switching through reminders and outreach, with the idea that active shopping will improve market outcomes. However, as noted by Dafny, Ho, and Varela (2013), if choice persistence is primarily due to preference heterogeneity, those policies may be misguided; it may be better to simply encourage product variety.[3]

Prior econometric models for distinguishing state dependence from unobserved heterogeneity have considered two general approaches. One approach estimates a fully parametric utility model that includes a cost of switching from an individual's lagged choice (their "state"). As emphasized by Heckman (1981), this requires treatment of the "initial conditions" problem.[4] Alternatively one can use the results in Honoré and Kyriazidou (2000) that allow for (very flexible) unobserved preferences, captured by *individual-by-product fixed effects*, but requires finding cases where all product characteristics (including prices) are *constant* (or nearly so) over time. In many settings—including the health insurance context we study—such conditions are rare enough that the conditioning set becomes exceedingly small.

We offer a new way of distinguishing between state dependence and heterogeneity. It uses moment inequalities derived from revealed preference in a model which allows for fully flexible product-by-individual specific fixed effects and non-parametric disturbance and price coefficient distributions to identify bounds on the impact of state dependence. The moment inequalities lead to an analysis which is similar to the "within" analysis in continuous choice models. The

[3]Similar questions about the role and implications of heterogeneity vs. state dependence have been studied in a variety of applied settings. Examples include consumer products markets (Keane, 1997; Dubé, Hitsch, and Rossi, 2009, 2010; Bronnenberg, Dubé, and Gentzkow, 2012), residential electricity markets (Hortaçsu, Madanizadeh, and Puller, 2017), auto insurance (Honka, 2014), and paid television services (Shcherbakov, 2016).

[4]That is it requires data on individuals who make choices without any state dependence (e.g., a first-time product choice), or with the unlikely proposition that their state is unrelated to their preferences. Valid initial conditions are not always available, and even when they are, identification of switching costs comes partly from the parametric specification of their distribution conditional on observable determinants of the choice.

bounds they generate are obtained from patterns of switching choices in response to *changes in prices* (or other product characteristics). The central assumption needed for the analysis is that once we allow for individual-by-product fixed effects and state dependence, the remaining unobservables in the agent's utilities for the various choices is identically distributed over time. This enables a revealed preference argument that makes the economic intuition underlying the bounds transparent.

To see this consider the simple case where price is the only observed product characteristic of interest—i.e. relegate the impact of other observables to the fixed effects. To obtain an upper bound start with a group of agents who were at choice c in $t-2$. Say the relative price of c rises in $t-1$ and a portion switch out of c. Now consider price changes at time t. Regardless of what those changes are there is a value for the state dependence parameter that is large enough to insure that the fraction of the original group who chose c in t should be less then the fraction that chose c in $t-1$. If the fraction of the group who are in c at t is smaller than in $t-1$ the state dependence parameter must be bounded from above.

For the lower bound consider a group of agents who are at c in $t-3$ and stayed in c in $t-2$. Now say the relative price of c goes up in $t-1$ and some of that group switch to other choices. Then the price of c falls to a lower level then it was at $t-2$ yet less chose c in t than in $t-2$. The price of c in t is lower than in $t-2$ yet agents do not switch back because to chose c in t they would incur switching costs whereas when they chose c in $t-2$ they did not have to incur those costs. So there are two changes when we compare the choice of c in the two periods; i) the relative price of c is lower in period t, and ii) the choice of c in t did generate switching costs while that choice in $t-2$ did not. If fewer people chose c in t than in $t-2$ the switching costs must have generated more disutility than the price decrease caused increase in utility, giving us our lower bound.

The next section formalizes these arguments and considers both extensions to allow for more than one variable to change over time, and the simplifications that are available if the data contains individuals who make the choice for the first time. Though our analysis can be used to bound the impact of state dependence on the coefficients of interest, a more detailed model may be needed to uncover the mechanism that generate it. Relatedly, the moment inequalities we derive are unlikely to provide a sharp characterization of the identifying information on κ_0, but we exploit variation in choices in a straightforward way that should appeal to practitioners.

In the nonparametric case we consider, a sharp characterization is, in principle, possible via a computationally intensive algorithm recently developed in Mbakop (2023).[5]

Related Econometric Literature. We build on two strands of the literature: papers that analyze discrete choice models with fixed effects and papers that add state dependence to that problem. Chamberlain (1980) shows how an assumption of "logit" disturbances generates a consistent conditional likelihood estimator for that problem. Manski's (1987) maximum score estimator provides consistent estimates for the binary choice problem with fixed effects and a nonparametric disturbance distribution. Papers by Shi, Shum, and Song (2018) and Pakes and Porter (2024), which we return to below, use an assumption of stationarity of the marginal distribution of disturbances over time to obtain their estimators for multinomial problems. Also related is work by Tebaldi, Torgovitsky, and Yang (2023) that develops a method to estimate static demand for health insurance in a model with flexible, nonparametric preference heterogeneity.

As noted, Honoré and Kyriazidou (2000) allow for state dependence and fixed effects and generate point identification by conditioning on observations that are matched across periods. A recent paper by Honoré and Weidner (2024) considers a binary logit model with state dependence that does not require matching (or situations with constant product characteristics over time).[6] Honoré and Tamer (2006) examine identified sets from a related model, and Khan, Ouyang, and Tamer (2021) investigate different assumptions on disturbances using both conditioning and matching. Torgovitsky (2019) considers state dependence through a nonparametric dynamic binary potential outcome framework and provides an approach to computing sharp bounds on state dependent treatment effects under various assumptions.

Empirical Results. Our empirical work analyzes health insurance choices in the Commonwealth Care ("CommCare") program in Massachusetts, enacted as part of the state's "RomneyCare" reform. The program provided subsidized health insurance for citizens with incomes

[5]Relatedly, our focus is on utility function parameters, rather than on the quantiles or averages of utilities, which is the focus of Chernozhukov, Fernández-Val, Hahn, and Newey (2013), or the treatment effect parameters defined in Torgovitsky (2019). This is largely due to our interest in evaluating counterfactuals, including equilibrium responses to changes in the environment.

[6]Honoré and Weidner (2024) use a "functional differencing" method (Bonhomme, 2012) that allows them to construct moment functions across possible outcomes that exactly difference out choice probabilities from the logit model, generating a mean-zero GMM moment and point identification of the state dependence parameter.

below 300% of the federal poverty level via an insurance exchange that let consumers choose among competing private plans. The program started in 2007 and grew steadily during 2007 and 2008. We begin our analysis in 2009 at the time of the first large price change (conditioning on choices prior to this) and use plan switching behavior from 2009 to 2013 (just before the transition to the Affordable Care Act) for our empirical estimates. Importantly, the program features several large price changes that provide identifying variation for our method.

We use individual-level panel data on insurance choices to estimate switching costs in our model using both our non-parametric and parametric moment inequality approaches. In our non-parametric analysis, we find bounds on κ_0 of \$102 to \$186 per year (with a 99% confidence range of (\$78, \$450)). In our main parametric model, we find a somewhat larger (point) estimate of \$591 per year (with a 99% confidence interval of (\$432, \$754)). These switching costs are meaningful relative to average (subsidized) consumer premiums in the market, which vary from \$575-\$740 per year during this period.

The upper bound on switching costs is a focus of our analysis. This is because we find that the upper bounds from our method are much *smaller* than the point estimates from methods used in the prior applied literature, and the difference is large enough to have a substantial impact on our counterfactual analysis.

To show this we use our data to estimate logit choice models that allow for state dependence but do not allow for individual-by-product fixed effects, instead relying on alternative approaches to capture unobserved heterogeneity. These choice models include plan fixed effects interacted with: (i) increasingly detailed consumer attributes (up to 252 interactions between consumer and product attributes), (ii) individual random effects assumed to be orthogonal to an initial lagged choice, and (iii) individual random effects starting from a plausible initial condition (a consumer's first choice in the market), with the likelihood function simulated over their full sequence of subsequent choices.

Across these comparison models, we estimate much higher switching costs of \$989 to \$1527 per year — values that are two to three times larger than the *upper bound* of the confidence interval from our non-parametrics method (\$450). These higher estimates are consistent with prior work on the CommCare data (see Shepard (2022), who finds $\kappa_0 \approx \$1,000$) and with similarly high estimates in other health insurance settings (e.g., Handel, 2013; Polyakova, 2016; Heiss, McFadden, Winter, Wuppermann, and Zhou, 2021). The much lower switching costs from our method (with flexible fixed effects) suggests a large role for unobserved preferences that is not easily captured by observed consumer attributes or random effects in the health insurance context,[7] implying less inertia and considerably larger price-responsiveness than obtained from prior procedures.

We conclude with an examination of the implications of the difference between the estimates that do and do not allow for fixed effects. The largest plan in our data experimented from 2011-12 with increasing its average annual premium from \$701 to \$1,171. After experiencing sharp losses in market share, it reduced its premium to \$554 in 2013. Using the comparison models to compare the implications of the estimates of κ that do and do not allow for fixed effects, we consider a counterfactual where instead the plan priced at the average of the 2012 and 2013 prices in both years. The difference in the predictions from using the different κ estimates is dramatic. The predicted share decline in 2012 is four to five times larger when we use our estimates, and the predictions for the two-year change actually differed in sign.

Outline of Paper. We begin with a revealed preference inequality that provides the relationship between price (or attribute) changes and switching behavior that underlies all of our results. Next we consider the implications of a method that makes only weak assumptions on the disturbance terms. These implications are then used to investigate the role of state dependence in the choice of plans made by the participants in CommCare. Next we consider the implications of revealed preference when one is willing to make parametric assumptions on the disturbances, first without and then with the additional structure of extreme value disturbances. The implications of the differences between the various estimators are explored in the counterfactual analysis. We conclude with a brief summary. All proofs are provided in Appendix 6.4.

[7] One plausible reason is the key role of varying hospital and physician networks across plans, combined with individual-specific preferences for accessing certain doctors/hospitals with whom patients have an existing relationship (see Shepard, 2022; Tilipman, 2022). Another plausible explanation is varying perceptions of insurer brand quality (Starc, 2014), perhaps based on local advertising or recommendations of family and friends.

2 Overview: Price Changes that Induce Switching.

We consider a panel data setting where consumers (i) choose from a choice set $\mathcal{D}_t$ over a sequence of periods $t \in \{0, 1, ...\}$. We denote the choice of consumer i at time t as y_{it}. We allow for *state dependence* in which an individual's lagged choice, $y_{i,t-1}$ (their "state"), influences the desirability of choosing this same option in period t. For expositional ease we begin with the simple case in which the price coefficient is constant over time but can differ across individuals, while the monetary equivalent of "switching costs", that is the ratio of switching costs to price, does not vary across either time or individuals. So the utility individual i derives from choice d at time t is

$$U_{d,i,t} = \underbrace{(-p_{d,i,t} - \kappa_0 \cdot 1\{y_{i,t-1} \neq d\}) \cdot \gamma_i \; + \; \lambda_{d,i}}_{\text{Structural Utility } (=SU_{d,i,t})} \; + \; \underbrace{\varepsilon_{d,i,t}}_{\text{Error}} \tag{2.1}$$

where we assume that $\gamma_i > 0$. We call the first set of terms the "structural utility" of option d at time t for person i, or $SU_{d,i,t}$. It captures both the value of observed product characteristics, which for expositional simplicity we limit to prices ($p_{d,i,t}$), the role of switching costs (κ_0), and unobserved preferences which vary across products for each individual ($\lambda_{d,i}$). The model also includes a disturbance ($\varepsilon_{d,i,t}$) on which we entertain various assumptions below. Consumers choose the option that maximizes their utility[8] so $y_{i,t} = \arg\max_{d \in \mathcal{D}_t} U_{d,i,t}$.

Below we generalize and allow for additional observables, say $x_{i,t}$, that vary over time and individuals and can impact both the price coefficient, so γ_i becomes $\gamma(x_{i,t})$, and the switching costs, so κ_0 becomes $\kappa_0(x_{i,t})$; generalizations whose importance are accentuated by our empirical findings. The distinction between the $x_{i,t}$ and the $\lambda_{d,i}$ is that the observed characteristics can change over time, whereas the $\lambda_{d,i}$ do not. Our analysis will only compare choices that are no more than two periods apart, so we expect the $\lambda_{d,i}$ to capture much of the individual specific differences in utility across products.

The key identification challenge in applied work is separately distinguishing the switching cost κ_0 from unobserved preferences, $\lambda_{d,i}$. A naive approach that ignores unobserved preferences

[8]By modeling the choice in this way we are not allowing perceptions of the future to have a direct impact on the current choice. This limits the applicability of the model, as there are many instances where these perceptions are integral to analyzing the questions of interest. For both a review of the literature that does incorporate decision models that are explicitly forward looking, and a set of restrictions which enables one to use sufficient statistics to circumvent related technical problems that arise in estimation, see Aguirregabiria, Gu, and Luo (2021).

and simply estimates a specification like (2.1) with the λ's omitted is likely to yield an upward-biased estimate of switching costs (κ_0) and bias the price coefficient (γ_i) towards zero. When prices change the naive estimate of κ_0 will pick up both real switching costs and the fact that people who chose d last period likely had stronger unobserved preferences for it (high $\lambda_{d,i}$). This will increase the response to last period's choice (or the "state"), and decrease the estimate of the response to price. Researchers can partly address this concern by adding more *observed* heterogeneity to the model,[9] but it is difficult to know whether the included variables capture all the relevant sources of heterogeneity in preferences.

Our approach uses the simple logic of revealed preferences to "difference out" or isolate unobserved preferences, yielding moment inequalities that can separate the impact of price movements from state dependence in the presence of arbitrary $\lambda_{d,i}$. To see how this works, consider two choices c and d that are both feasible in period s. If the agent chose c instead of d, revealed preference implies $U_{i,c,s} \geq U_{i,d,s}$, or

$$U_{i,c,s}-U_{i,d,s} = \Big(-[p_{c,i,s}-p_{d,i,s}]-[\mathbf{1}\{y_{i,s-1} \neq c\}-\mathbf{1}\{y_{i,s-1} \neq d\}]\kappa_0\Big)\gamma_i+[\lambda_{c,i}-\lambda_{d,i}]+[\epsilon_{c,i,s}-\epsilon_{d,i,s}] \geq 0.$$

Now consider a second period, say $t > s$, where the same two choices were available but a relative price change induced the agent to choose d instead of c. Then

$$U_{d,i,t}-U_{c,i,t} = \Big(-[p_{d,i,t}-p_{c,i,t}]-[\mathbf{1}\{y_{i,t-1} \neq d\}-\mathbf{1}\{y_{i,t-1} \neq c\}]\kappa_0\Big)\gamma_i+[\lambda_{d,i}-\lambda_{c,i}]+[\epsilon_{d,i,t}-\epsilon_{c,i,t}] \geq 0.$$

Denote the difference in these two utility differences between periods t and s by

$$\Delta\Delta U^{d,c}_{i,t,s} \equiv (U_{d,i,t} - U_{c,i,t}) - (U_{d,i,s} - U_{c,i,s})\,.$$

Analogously, denote the difference in the associated price differences between the two periods by $\Delta\Delta p^{d,c}_{i,t,s}$, and the difference in disturbances by $\Delta\Delta\epsilon^{d,c}_{i,t,s}$. Adding the two revealed preference inequalities yields $\Delta\Delta U^{d,c}_{i,t,s}$ and, crucially, cancels the fixed effect terms:

$$\Delta\Delta U^{d,c}_{i,t,s} = \Big(-\Delta\Delta p^{d,c}_{i,t,s} - \Delta\Delta sw^{d,c}_{i,t,s} \cdot \kappa_0\Big)\gamma_i + \Delta\Delta\epsilon^{d,c}_{i,t,s} \geq 0, \tag{2.2}$$

[9]For a health insurance example, see Figure 5 in Heiss, McFadden, Winter, Wuppermann, and Zhou (2021).

where $\Delta\Delta sw_{i,t,s}^{d,c}$ is the difference between the current and prior period of the difference in the indicator for whether the individual was at state c or d when the two choices were made, or formally

$$\begin{aligned}\Delta\Delta sw_{i,t,s}^{d,c} &= \Delta\Delta sw_{i,t,s}^{d,c}(y_{i,t-1}, y_{i,s-1}) \\ &\equiv (\mathbf{1}\{y_{i,t-1} \neq d\} - \mathbf{1}\{y_{i,t-1} \neq c\}) - (\mathbf{1}\{y_{i,s-1} \neq c\} - \mathbf{1}\{y_{i,s-1} \neq d\}).\end{aligned}$$

This observable term will play an important role in what can be learned about switching costs, κ_0, from this sequence of choices. We show below that when $\Delta\Delta sw_{i,t,s}^{d,c}$ is nonzero, its sign determines whether upper or lower bound information about κ_0 is available.

For a given distribution of the errors, equation (2.2) can be used to bound the probability of the sequence of choices described (i.e. $y_{i,s} = c$, $y_{i,t} = d$). In section 4, we implement this inequality under the logistic assumption. However, our main results do not rely on parametric distributional assumptions on the error term. Next, we show how within differencing can translate to bounds on κ_0 under weak assumptions on the distribution of errors.

2.1 Nonparametric Bounds

We are allowing for fixed effects, so we are analyzing the determinants of changes in individuals' choices over time (the analogue of analyzing the "within" dimension in models with continuous left hand side variables). Equation (2.2) implies that the response of the choices to the price differences in the two periods depends only on κ_0 and $\Delta\Delta\epsilon_{i,t,s}^{d,c}$. Typically estimation and inference of dynamic discrete choice models is based on logistic or other parametric distributional assumptions for the disturbance term. We will focus on the non-parametric case where the main assumption on the $\epsilon_{i,t} \equiv [\epsilon_{1,i,t}, \dots \epsilon_{\mathcal{D},i,t}]$ is a stationarity assumption. Let $\lambda_i \equiv [\lambda_{1,i}, \dots, \lambda_{\mathcal{D},i}]$, $p_{i,t} \equiv [p_{1,i,t}, \dots p_{\mathcal{D},i,t}]$, and $p_i \equiv [p_{i,1}, \dots, p_{i,T}]$. Additional covariates $x_{i,t}$ will be included below, so we will also include covariates in the statement of the assumption with $x_i \equiv [x_{i,1}, \dots, x_{i,T}]$. The distribution of the error terms must satisfy the following.

Assumption 2.1.

$$\epsilon_{i,t} \,|\, p_i, x_i, y_{i,t-1}, y_{i,t-2}, \dots, y_{i,1}, \gamma_i, \lambda_i \;\sim \epsilon_{i,s} | \, \gamma_i, \lambda_i$$

for all time periods t and s. $\square$

Assumption 2.1 covers many panel settings.[10] It does not impose any restrictions on the correlation between the $\{\lambda_i\}_i$ and the observable determinants of the choice (there can be "correlated effects"), and it allows $\epsilon_{d,i,t}$ to be freely correlated with $\epsilon_{c,i,t}$, $\forall(c,d) \in \mathcal{D}^2$. We also assume the ϵ_i are identically distributed across individuals, though the identification results could be re-written to allow for non-identical distributions.

Note, however, that this assumption does imply that the current disturbance is conditionally independent of past disturbances, so we are attributing all dependence of choices over time that are not a function of changes in observables to state dependence. This contrasts with prior non-parametric work on models with state dependence and fixed effects which allocates all additional time dependence in choices to serial correlation in disturbances.[11] By attributing all the additional time dependence in choices to κ_0 here, we hope to get a robust upper bound to its value.

Since Assumption 2.1 places no restrictions on the marginal distribution of errors, the scale of random utilities is not identified. Without loss of generality, then, we can normalize $\gamma_i \equiv 1$ when taking this nonparametric approach. In this formulation, only the ratio of the switching cost coefficient to the price coefficient is potentially identifiable. Below we will generalize the specification to allow the ratio of state dependence to price to depend on observable individual characteristics that may change over time.

From the point of view of revealed preference, Assumption 2.1 implies that the difference in an individual's probability of choosing d instead of c between any two periods cannot arise from differences in the distributions of the disturbance in those periods. Instead, it must result solely from the difference in the *structural utility* between the two periods. So if we eliminate

[10] In our discussion paper (Pakes et. al. 2021) we show that equation (2.2) can identify bounds when all that is assumed is that the conditional median of $\Delta\Delta\epsilon^{d,c}_{i,t,s} = 0$, if and only if there are cells in the data where the probability of switching twice is greater than a half. If this condition is satisfied then one can allow the distribution of $\epsilon_{i,t}$ to differ over time. Unfortunately this rules out most studies of insurance choices, though it often does not rule out retail market choices, particularly those with periodic sales.

[11] Pakes and Porter (2024) provide sharp bounds for "static" panel data choice models under the alternative assumption that the disturbances are freely correlated over time but there is neither state dependence nor correlation between the disturbances, which represent unobserved preferences, and the included regressors. Here we allow for both arbitrary unobserved preferences that are fixed over comparison periods and state dependence, but no serial correlation in the disturbances. By eliminating serial correlation we maximize the role for state dependence in accounting for the persistence of individual choices over time.

the impact of the $\lambda_{d,i}$, as we do in equation (2.2), the differences in choices following a relative price change will be determined by the tradeoff between the importance of price and that of state dependence in determining utility.

We now show how this logic can generate moments that identify upper and lower bounds on κ_0. In doing so we describe the moments that we actually use in our empirical analysis. For the upper bound, the intuition is that a relatively high probability of switching to a particular choice will be evidence that switching costs cannot be too large. We use the discrete choice model, under Assumption 2.1, to determine which choices allow us to make use of this intuition.

Given a switching cost of κ, suppose that choice d^* satisfies

$$SU_{d^*,i,t}(d^*;\kappa) - SU_{d^*,i,t-1}(y_{i,t-2};\kappa) = -(p_{d*,i,t} - p_{d*,i,t-1}) + \kappa \geq SU_{c,i,t}(d^*;\kappa) - SU_{c,i,t-1}(y_{i,t-2};\kappa) \tag{2.3}$$

for all choices $c \neq d^*$, where the lagged dependent variable $y_{i,t-2}$ is a choice other than d^*, and we are writing structural utility as a function of the lagged dependent variable and the switching costs, κ. Note that since we are comparing the same choice in different periods the fixed effects drop out of both sides of this equation.

Equation (2.3) states that choice d^* experiences the largest increase in structural utility from $t-1$ to t. If the given value of κ is the true value of switching costs, then this increase in structural utility for choice d^* will necessarily lead to a corresponding increase in the choice probability, so

$$\Pr(y_{i,t} = d^* \mid y_{i,t-1} = d^*, \lambda_i) \geq \Pr(y_{i,t-1} = d^* \mid y_{i,t-2} = c, \lambda_i). \tag{2.4}$$

Inequality (2.4) is stated for individual specific probabilities. Since these probabilities depend on λ_i, they are not directly measurable from the data. Note that the inequality in equation (2.3), which implies (2.4), does not depend on the fixed effects λ_i. Still the differing conditioning sets of the two choice probabilities in (2.4), and the fact that we do not know the distribution of the λ_i, means that one cannot simply average or integrate out the λ_i from both sides of the inequality. Below, in (2.12), we find that a more subtle argument which uses Jensen's Inequality shows that equation (2.3) implies that for a sufficiently large κ the proportion of a set of individuals who chose c in $t-2$, that chose d^* in $t-1$ and in $t-2$ must satisfy

$$\Pr(y_{i,t} = d^* \,|\, y_{i,t-1} = d^*, y_{i,t-2} = c) \geq \Pr(y_{i,t-1} = d^* \,|\, y_{i,t-2} = c), \tag{2.5}$$

where now $Pr(\cdot)$ refers to the probability (or proportions), and these can be consistently estimated.

This result is the "positive" implication of a κ that satisfies equation (2.3). That is, if (2.3) is true for those at c in $t-2$, then the inequality in equation (2.5) must be true. Empirically, we make use of the "contrapositive,"[12] that is, if the inequality in (2.5) is violated, then κ_0 can not satisfy (2.3). Since equation (2.3) can be rewritten

$$\kappa \geq \max\left\{ \frac{\Delta\Delta p^{d^*,c}_{i,t,t-1}}{2}, \max_{c' \neq d^*, c} \Delta\Delta p^{d^*,c'}_{i,t,t-1} \right\} \equiv MaxP(d^*), \tag{2.6}$$

if we find that (2.5) is not satisfied we will learn that

$$\kappa_0 < MaxP(d^*).$$

When are we likely to get this upper bound to κ_0? Consider individuals who start at some $c \neq d^*$ in $t-2$. If the price of c relative to that of d^* rises in period $t-1$, we expect some of those who chose c in $t-2$ to switch to d^* in $t-1$. Now say the price of all $c \neq d^*$ fall in period t. If enough people who switched out in $t-1$ do not switch back in t, it must be the case that there is an upper bound to switching cost.

We next provide moments which generate a *lower bound* for κ_0. Lower bounds are generated by turning around the argument for upper bounds. To get a lower bound, we compare the probability of staying with a given choice in period s to the probability of switching to that choice in period t. To consider a switch in the later period t, we need to allow s and t to be separated by at least two periods, e.g. $s = t-2$. This pattern of choices is the opposite of the pattern for the upper bound, and so it will lead to a different sign on the switching variable, that is on the coefficient of κ, in the difference of utilities. This sign flip on κ coincides with obtaining lower bound information rather than upper bound information. More intuitively,

[12] That is, if A implies B, then B is not true implies A is not true.

when the switch at time t has low probability relative to the probability of staying at time $t-2$, then this will potentially provide evidence that switching costs cannot be too low.

We now show how particular choices generate a lower bound. Suppose choice $d*$ has the largest increase in structural utility between periods t and $t-2$, that is

$$SU_{d_*,i,t}(y_{i,t-1};\kappa)-SU_{d_*,i,t-2}(d*;\kappa)=-(p_{d_*,i,t}-p_{d_*,i,t-2})-\kappa\geq SU_{c,i,t}(y_{i,t-1};\kappa)-SU_{c,i,t-2}(d*;\kappa) \tag{2.7}$$

for all $c\neq d*$ and $y_{i,t-1}$ is a choice other than $d*$. Again, since we are considering differences in structural utilities of the same choice at different times, the fixed effects drop out of all of these differenced expressions.

If (2.7) holds for κ equal to the true value κ_0, then choice $d*$ experiencing the largest increase in structural utility implies that this choice will have a corresponding increase in probability,

$$\Pr(y_{i,t}=d*\,|\,y_{i,t-1},y_{i,t-2}=d*,\lambda_i)\geq\Pr(y_{i,t-2}=d*\,|\,y_{i,t-3}=d*,\lambda_i), \tag{2.8}$$

where $y_{i,t-1}\neq d_*$.

As with the upper bound, the inequality in (2.7) does not depend on λ_i, but averaging out over the individual probabilites in (2.8) presents a challenge due to the presence of the different lagged dependent variables in the conditioning sets. Below, in (2.12), we invoke Jensen's Inequality to deal with this issue, and find that equation (2.7) with $y_{i,t-1}=c$ ($\neq d*$) will also imply that if we consider the people who started at d_* and chose d_* at $t-2$, and then look at the proportions who chose d_* in t after choosing c in y_{t-1},

$$\Pr(y_{i,t}=d*\,|\,y_{i,t-1}=c,y_{i,t-2}=d_*)\geq\Pr(y_{i,t-1}=c,y_{i,t-2}=d*\,|\,y_{i,t-2}=d*), \tag{2.9}$$

where now $Pr(\cdot)$ refers to the probability (or proportions) which can be consistently estimated.

Finally, using the contrapositive of this result, if the inequality in (2.9) is violated then (2.7) will also fail. Failure of (2.7) can be re-written as a lower bound for κ_0,

$$\kappa_0>\min\left\{-\frac{\Delta\Delta p^{d_*,c}_{i,t,t-2}}{2},\min_{c'\neq d_*,c}-\Delta\Delta p^{d_*,c'}_{i,t,t-2}\right\}.$$

The rest of this subsection provides an overview of the three extensions to these results which we use in our empirical work. A more formal presentation of the results leading to them is given in the Appendix.

Initial Condition Estimators. Heckman (1981) introduces an initial condition estimator for the problem of analyzing discrete choice models with state dependence and versions of it have been used extensively since. It requires data on individuals who are making the specified choice for the first time, and we do have individuals who enter CommCare for the first time. The initial condition estimator is typically used in conjunction with a parametric assumption on the distribution of the disturbances conditional on the observable determinants of the choice.

The estimator makes the assumption that the agent knows the value of the individual specific effects (the vector $\{\lambda_i\}$) before making their initial choice. When we add this assumption to our stationarity assumption we can obtain a moment inequality estimator that is less demanding of the data. This is because there is no state dependence when the initial choice is made. So adding the initial revealed preference inequality to that from a later period differences out the fixed effects and generates inequalities that bound κ_0. The form of the upper and lower bounds for κ_0 is in Appendix 6.3, and our empirical work considers this generalization in the context of the health insurance choices in the CommCare data.

Other observable determinants of the importance of switching costs. The next section explains and the Appendix formally derives non-parametric bounds that allow the ratio of switching costs to price to differ across individuals and/or over time. Our empirical work does not use individuals who experience differences in choice set characteristics other than price, and allows the preferences for those characteristics to vary freely across cells constructed from individuals with similar value of observable covariates. Given these constraints we investigate whether the ratio of switching costs to price within cells differs with income and/or health status.

Parametric Methods. Our focus is on non-parametric analysis, but there are two reasons to also consider parametric versions of the revealed preference analysis that allows for fixed effects and state dependence. First, we want to understand the sources of the difference between the bounds obtained from our nonparametric procedure and the estimates from prior parametric models that *do not* allow for fixed effects: are they a result of the parametric assumptions, or due to the absence of the fixed effects? To this end we also use our data to provide estimates from prior models that allow for state dependence but not our λ_i and compare them to our non-parametric results. Second, once we make parametric assumptions we can derive inequalities that do not require the choice set to be constant over the periods comparred, an extension which is likely to be particularly useful in panel data studies of purchases at retail outlets.

There are parametric inequalities that are "generic" in the sense that they can be derived for any distribution of the disturbance terms. Unfortunately these inequalities do not provide upper bounds for the magnitude of the ratio of state dependence to price which is our focus. As a result we consider adding the additional information available from models which assume a logit distribution for the disturbances, the dominant parametric assumption used in prior research on health insurance choice.

2.2 Generalizations and the Use of Jensen's Inequality.

We start with a generalization of the discussion of the last section which allows switching costs to depend on covariates and considers probabilities of choosing a subset of the available choices. The reader who is not interested in these derivations should be able to go directly to section 3, and follow the remainder of the text.

We maintain the stationarity Assumption 2.1 and the normalization that $\gamma_i \equiv 1$, and consider variation in choice probabilities across two periods t and s, with $t > s$. The structural utility for the more general specification is

$$SU_{d,i,t}(y_{i,t-1}) \equiv -p_{d,i,t}\gamma_0(x_{i,t}) - \mathbf{1}\{y_{i,t-1} \neq d\}\delta_0(x_{i,t}) + \lambda_{d,i}, \tag{2.10}$$

where we have made explicit the dependence of coefficients on the observable, $x_{i,t}$. Switching costs will be captured by the ratio of the coefficients on state dependence and price, $\kappa_0(x_{i,t}) = \delta_0(x_{i,t})/\gamma_0(x_{i,t})$.

We first establish an inequality on the choice probabilities that extends an analogous result from the panel data discrete choice literature that does not allow for state dependence. The inclusion of state dependence in the choice probabilities distinguishes our case from that earlier case and makes the inequality derived in that literature not directly applicable to a model with state dependence. We then show how to use Jensen's Inequality to transform that inequality in a way that generates a new probability inequality that can be implemented when state dependence is present.

For any choice $d \in \mathcal{D}$, we can consider the change or difference in the structural utility, $SU_{d,i,t}(y_{i,t-1}) - SU_{d,i,s}(y_{i,s-1})$. Since these are differences in the utility of the same choice in two periods, the fixed effects cancel. We can then order these changes by choice from largest to smallest for a given $x_i \equiv (x_{i,1}, \ldots, x_{i,T})$. The result from the previous literature that does not allow for state dependence (Pakes and Porter 2024) is that if we order the choices based on the differences in structural utility (with the true values of the parameters), then the probability of choosing one of the highest ranked choices will be greater in period t than in period s. That is, the choices with the largest structural utility difference will have a larger probability in t than in s. Given that the distribution of the unobservable part of utility (including the fixed effects) is stationary over time, it is intuitive that increases in the structural utility lead to higher probabilities.

We can summarize this result as follows. Let D_0 be a set of choices containing the top ranked changes in structural utility. That is, if D_0 is a singleton, then it will contain only the choice with the largest increase in structural utility. If D_0 contains two choices, then these will be the choices with the largest and second largest increases in the structural utility, and so on. More explicitly, suppose D_0 satisfies

$$\min_{d \in D_0} SU_{d,i,t}(y_{i,t-1}) - SU_{d,i,s}(y_{i,s-1}) \;\geq\; \max_{c \notin D_0} SU_{c,i,t}(y_{i,t-1}) - SU_{c,i,s}(y_{i,s-1}).$$

Then,

$$\Pr(y_{i,t} \in D_0 | y_{i,t-1}, p_i, x_i; \lambda_i) \;\geq\; \Pr(y_{i,s} \in D_0 | y_{i,s-1}, p_i, x_i; \lambda_i). \tag{2.11}$$

In the model without state dependence the conditional probabilities in this inequality would not include a lagged dependent variable, and hence the conditioning sets on either side of the inequality would be identical. In that case, the fixed effects can be integrated out from both sides of the inequality to obtain conditional probabilities that could be used in estimation. In our current setting, the conditioning sets on the two sides of the inequality contain different lagged endogenous variables that depend on the fixed effects, so that simple method of integrating out the fixed effects is not applicable.

To see how we deal with this challenge, take the case where D_0 is a singleton, $\{d^*\}$, and $s = t-1$. Then, the probability inequality above is $\Pr(y_{i,t} = d^* | y_{i,t-1} = d^*, y_{i,t-2}, p_i, x_i; \lambda_i) \geq \Pr(y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i; \lambda_i)$. Multiplying both sides of this inequality by $\Pr(y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i, ; \lambda_i)$ and applying Jensen's Inequality, we have

$$\begin{aligned}
\Pr(y_{i,t} = d^*, y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i) &\equiv E_{\lambda_i}[\Pr(y_{i,t} = d^*, y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i; \lambda_i) | y_{i,t-2}, p_i, x_i] \\
&\geq E_{\lambda_i}[(\Pr(y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i; \lambda_i))^2 \, | y_{i,t-2}, p_i, x_i] \\
&\geq [E_{\lambda_i}(\Pr(y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i, \lambda_i))]^2 \\
&\equiv (\Pr(y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i))^2 , \qquad (2.12)
\end{aligned}$$

Finally divide both sides of (2.12) by $\Pr(y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i)$ to obtain

$$\Pr(y_{i,t} = d^* \mid y_{i,t-1} = d^*, y_{i,t-2}, p_i, x_i) \geq \Pr(y_{i,t-1} = d^* \mid y_{i,t-2}, p_i, x_i). \qquad (2.13)$$

Given $\gamma(x_i)$ and $\delta(x_i)$, we compute a value d^*. The inequality in (2.13) can then be checked empirically. Violations of (2.13) can then be used to rule out or provide evidence against the given values of $\gamma(x_i)$ and $\delta(x_i)$.

Two additional points are worth noting. First the slackness in (2.13) is due to the fact that it ignores the variance in $\Pr(y_{i,t-1} = d^* | y_{i,t-2}, p_i, x_i; \lambda_i)$ conditional only on $(y_{i,t-2}, p_i, x_i)$. This conditional variance, in turn, depends on the variance of the λ_i conditional on those variables. The λ_i are explained by both the observable and unobservable determinants of utility, and the richer the set of observable characteristics that we condition on, the lower the conditional variance of the λ_i in the data, and the more powerful this inequality. Second, though the result holds for all $s < t$, it requires the assumption that the $\{\lambda_i\}$ is fixed over the two comparison periods so s should not be too distant from t. Below we form cells with common observable characteristics, and only use choices made at an $s \geq t-2$.

The result in the theorem to follow extends the argument above in two ways. First we broaden the argument to apply to choice probabilities of non-singleton sets. When $s = t-1$ the extension just requires replacing d^* with D_0 as defined above. Second, if $s < t-1$, constructing the inequality for D_0 requires specifying $y_{i,t-1}$, and as in our discussion of the lower bound above, multiple values of $y_{i,t-1}$ can imply that the inequality is satisfied. The next condition defines the set of values $y_{i,t-1}$ can take, and labels that set as D_1.

Condition 2.2. *Given $t > s$ and choice sets $D_0, D_1 \subset \mathcal{D}$, for all $d' \in D_1$,*

$$\min_{d \in D_0}[SU_{d,i,t}(d') - SU_{d,i,s}(y_{i,s-1})] \;\geq\; \max_{c \notin D_0}[SU_{c,i,t}(d') - SU_{c,i,s}(y_{i,s-1})].$$

The next theorem generalizes the result in (2.13) to take account of these changes.

Theorem 2.3. *Suppose Assumption 2.1 holds.*

(a) For $s = t-1$, for any choice set $D_0 = D_1$ satisfying Condition 2.2,

$$\Pr(y_{i,t} \in D_0 \,|\, y_{i,t-1} \in D_0, y_{i,t-2}, p_i, x_i) \;\geq\; \Pr(y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i)$$

(b) For $s < t-1$, for any choice sets D_0 and D_1 satisfying Condition 2.2,

$$\Pr(y_{i,t} \in D_0 \,|\, y_{i,t-1} \in D_1, y_{i,s} \in D_0, y_{i,s-1}, p_i, x_i) \;\geq\; \Pr(y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i).$$

3 Data, Estimation, and Non-parametric Results.

We analyze health insurance plan choices made by enrollees in the Commonwealth Care ("Comm-Care") program in Massachusetts between 2009-2013. The program provided heavily subsidized insurance to low-income adults (earning less than 300% of the Federal Poverty Level) via a market featuring competing private health insurers. Five insurers participate in the market during our data period, with each insurer (by rule) offering a single plan. Program rules required each enrollee to make a separate choice; there was no family coverage, and kids were covered in the separate Medicaid program. Individuals make plan choices at two times: (1) when they join the market as a new enrollee, and (2) during an annual open enrollment month when they are allowed to switch plans. Because our focus is on switching costs, we study open enrollment choices, setting the prior choice (the state, $y_{i,t-1}$) equal to the individual's plan in the month prior to open enrollment.[13] For more detail on the data and the CommCare program see Shepard (2022); Finkelstein, Hendren, and Shepard (2019); McIntyre, Shepard, and Wagner (2021); Shepard and Wagner (2025).

Treatment of the data. We want to capture switching costs that are not induced by changes in the individual's choice environment, other than by prices, $y_{i,t-1}$, and $x_{i,t}$ (when we allow for heterogeneity in switching costs), and we need the choice set to be the same in the two periods we compare. We therefore remove comparisons for individuals who changed regions (there are five in the data), or who faced different plan offerings in the comparison periods and, as noted, only consider $s = t - 1$ and $t - 2$.

We begin by assuming that the relative cost of switching, relative to individual specific aversions to price, does not vary across individuals, and then generalize to allow the cost of switching relative to price to differ between health and income groups. The health groups are defined by whether one is below or above the median of the medical risk score distribution in year t.[14] There are five income groups in the data, defined by income relative to the federal poverty line (FPL): 0-100% FPL, 100-150% FPL, 150-200% FPL, 200-250% FPL, and 250-300% FPL. Subsidies—and therefore post-subsidy premiums—vary across these groups, with lower-income groups both paying lower premiums overall and having narrower premium differences across plans.[15] Besides variation across income groups due to subsidies, price variation was limited by regulations. Prices could vary by region in 2009-2010 but not from 2011-on. No variation was allowed on other factors including age, gender, health status, or any other characteristics.

Our model assumes that individual-level unobserved plan preferences ($\lambda_{d,i}$) are stable over the time periods we compare. This is consistent with the characteristics of the CommCare market. Coverage is heavily regulated, with all cost sharing and covered medical services completely standardized across insurers. The only flexible plan attributes are provider networks.

[13]In rare circumstances, individuals are allowed to switch plans mid-year (e.g., if they move across regions). Though we do not include these mid-year switching opportunities in our estimation, we do condition on any switches that occur mid-year and update the lagged plan accordingly for the next switching opportunity at open enrollment.

[14]We use risk scores from the Department of HHS's Hierarchical Condition Categories (HHS-HCC) method that is used for risk adjustment in the ACA Marketplaces. We apply the method as posted on the HHS website to demographics and diagnoses observed on our claims data.

[15]When we include switching cost heterogeneity across incomes (see below), we only allow heterogeneity for two income groups: 100-200% FPL versus 200-300% FPL. This reduces the number of parameters we need to estimate for the more time-intensive method.

These were largely stable during our sample period with one major exception. Network Health (one of our plans) dropped Partners Healthcare (the state's largest medical system) from its hospital network at the start of 2012. This decreases the value of Network Health. By transitivity of preferences those who either stayed with Network or switched to Network would have preferred Network with the prior values of their individual specific fixed effects, so their 2012 choices can be used below. However those who left Network could have left because the value they assigned to Network decreased when Network dropped Partners, so their 2012 choices are dropped from the data. There were no other major changes in the networks of the plans during our study period. However, one plan enters mid-sample (Celticare during 2010-11), and one plan (Fallon) exits several areas in 2011.

Forming Inequalities. To form the sample analogues of the inequalities in Theorem 2.3 we form cells with the same observed characteristics and $y_{i,s-1}$. The observed characteristics of a cell are denoted by z_i and are defined by the Cartesian product of: a) couple of years, b) the five income groups (separately in year t and s), c) region, and d) plan availability (separately in year s and t). Note that this implies that the λ_i represent differences in tastes among consumers with the same z_i and $y_{i,s-1}$.

Table 1 provides summary statistics on both the full underlying dataset, and the estimation sample that is limited to cells with $N \geq 50$ members (which we use for reasons described presently). If two sets of inequalities are positive at the true parameter value we can sum them into a single inequality which is also positive at that value. Since we are focused on price responsiveness, we summed our cell inequalities into groups that exhibited the same price changes and initial choice. Our groups are then defined by (a) couple of years, (b) income group in t and s (which set subsidies) and (c) prior choice. In total, our estimation sample includes 158,994 member-choices, which fall into 722 cells and 446 groups. Because of many plan availability changes in 2011, there is insufficient data with stable plan choice sets to use year-pairs where s and t cross 2011. Therefore, our final sample includes four sets of year-pairs: (2009, 2010), (2011, 2012), and (2012, 2013) for $s = t - 1$ (from which we derive upper bounds on κ), and (2011, 2013) for $s = t - 2$ (from which we derive lower bounds).[16]

[16]Given our assumptions, the variables defining our groups are conditionally independent of the disturbances in the comparison periods. This implies that we can use subsets of the groups in estimation without incurring a selection bias. This assumes that the plan-specific effects are constant over time; but regulation insured that all

Table 1: Sample Sizes for the Nonparametric Estimator

Year Pair	Full dataset		Estimation sample ($N \geq 50$)		
	Number of Members	Number of Groups	Number of Members	Number of Cells	Number of Groups
Year pairs: $s = t - 1$					
$(2009, 2010)$	51,548	378	35,941	194	73
$(2011, 2012)$	55,657	462	47,168	184	127
$(2012, 2013)$	60,544	480	50,479	205	138
Year pairs: $s = t - 2$					
$(2011, 2013)$	33,747	458	25,406	139	108
Total	**201,496**	**1,778**	**158,994**	**722**	**446**

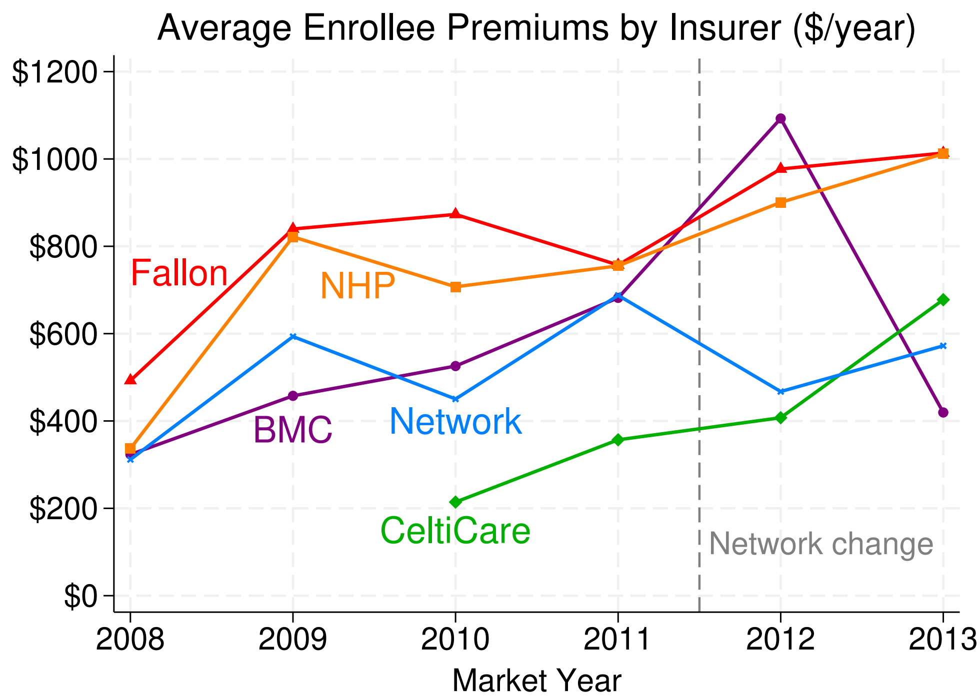
Notes: The table shows summary statistics for the nonparametric estimator sample, by pair of years (s, t). See the text for definitions of cells and groups used in the estimation. The table lists the number of members and groups, both before and after applying the minimum cell-size cutoff of 50 members.

Price movements. Figure 1 provides the average prices paid by consumers (i.e. after subsidy) by year and plan. It indicates that the prices of each of the plans do go both up and down over time. Large price changes occurred between 2011 and 2012. This reflects a new set of rules introduced by the market regulators in 2012 that changed the nature of competition in the market. In all years, enrollees with incomes 0-100% FPL were fully subsidized, meaning they paid $0 for all available plans. Prior to 2012 these enrollees were fully subsidized regardless of the plan they chose; however from 2012 on these new enrollees were only allowed to choose from the two lowest-price plans in their area. This created an auction-like dynamic in which the two lowest-bidding plans "won" access to this large group, representing about half of new enrollees.

The different plans reacted differently to the rule change. Boston Medical Center's plan (BMC), the largest plan as of 2011, increased its price sharply in 2012, essentially ceding the market for full-subsidy new enrollees to the other insurers, but then lowered its price by an even greater amount in 2013. This generated the two largest price changes in Figure 1, though there were also significant changes in the prices of Network and Celticare plans in that period, and smaller changes in the other plans. As noted there were no major changes in BMC's network or

non-price non-network plan characteristics, including cost sharing and covered medical services, were constant over time, and we have controlled for the only notable change in networks.

Figure 1: Massachusetts Health Insurance Market: Plan Choice and Premiums



Notes: The figure shows average annual enrollee premiums (after subsidies) for each of the Massachusetts CommCare plans over our sample period of 2008-2013. Averages shown are for above-poverty enrollees only; below-poverty enrollees pay $0 for all plans in all years.

other quality attributes over this period. Instead the change in 2012 appears to reflect BMC's strategic response to the new competitive rule. BMC chose to raise its price in 2012 and earn a larger margin on those members who did not leave. In contrast, Network Health and CeltiCare bid low in 2012 and "won" the auction. As a result of these choices, BMC lost almost half of its market share during 2012, and then decided to reverse course in 2013 and undercut both its competitors. This allowed it to rebuild its market share in 2013, leading into the important transition of CommCare into an Affordable Care Act exchange in 2014.

Table 2 summarizes the path of enrollment between 2011 and 2013 for people enrolled in CommCare in 2011. The top panel indicates that about two thirds of those enrolled in CommCare in 2011 had moved out of CommCare by 2013. This is a market with a lot of "churn" (largely induced by movements in and out of low-income eligibility due to employment changes). The bottom panel reports on switching behavior among subscribers who stayed in CommCare between 2011 and 2013. Among those who were in CommCare for the three years,

Table 2: Statistics on Enrollment and Switching for 2011 Enrollees over 2011-2013

	All 2011 Enrollees	**By 2011 Plan**	
		BMC	All Other Plans
Number of Enrollees			
Total Enrollees in 2011	114,752	38,131	76,621
Leave Market before 2013	78,654	26,113	52,541
Stay in Market 2011-13	36,098	12,018	24,080
Switching Rates (among stayers in market)			
Switch Plans from 2011-2012	13.4%	16.9%	11.6%
Switch in 2012, Switch Back in 2013	2.2%	5.3%	0.6%
Switch in 2012, Do Not Switch Back 2013	11.2%	11.6%	11.1%

Note: The table shows statistics on enrollment and switching rates over the 2011-13 period. The sample is people enrolled in CommCare in 2011 who are not in the below-poverty income group (who do not pay premiums so do not experience the premium changes shown in Figure 1), and the columns separate this group by their plan in 2011. The top panel shows enrollment numbers, and the bottom panel shows switching rates among people who stay in the market from 2011-13.

the fraction who switch plans in 2012 is 13.4%; BMC, the plan with the largest price increase, loses 17% of its 2011 subscribers. Though all prices changed in 2013, BMC is the only plan whose average price decreased. Of the subscribers who left BMC in 2012, a little less that a third switch back in 2013, while only five percent of those who switched out of other plans in 2012 switched back in 2013.

Estimation and Inference. We estimate the identified set using the sample moment inequalities computed at the group level. For confidence sets, we adopt a Bayesian approach, as proposed by Kline and Tamer (2020). In particular, we construct confidence sets using the "Bayesian bootstrap," which is relatively easy to apply and understand. We draw samples that mimic the observed sample and check which parameters satisfy the inequalities from the simulated samples. A parameter value that is accepted in (1-α) percentage of the simulated samples is in the $\alpha-$level confidence set.[17]

Formally, the Bayesian bootstrap combines an uninformative prior with the data to generate

[17] A prior version of this paper initially used a common inequality approach that obtained confidence sets based on an approximate distribution for the objective function test statistic (that minimizes the squared negative part of the inequalities derived directly from the data). This procedure is based on Gaussian approximations to the distribution of the empirical choice probabilities. When cell sizes are small and some empirical choice probabilities are near zero, the simulated moments can imply probabilities that are negative and produce degenerate confidence sets. These issues led to the use of the Bayesian bootstrap, which is especially well suited to the case of near zero choice probabilities.

a Dirichlet posterior distribution for the choice probabilities for each cell in the data. Given the uninformative prior, the Dirichlet posterior parameters are set to mimic the observed frequencies and sample size of the cell, as in Chamberlain and Imbens (2003). The posterior simulation draws values of probabilities for possible choice sequences in each cell. For the upper bound we focus on choices from consecutive time periods, so if the choice set has four plans, we draw probabilities for the $4^2 = 16$ choice sequences in each cell. The lower bound uses choices from three consecutive time periods, so we draw probabilities for $4^3 = 64$ choice sequences.

For each set of random draws we construct the inequalities based on the simulated probabilities implied by Theorem 2.3 with singleton sets D_0 ($D_0 = d^*$ or d_* in the notation of section 2.1), $s \geq t-2$, and all t for each cell. We then sum these inequalities over our groups, divide each aggregated inequality by its standard error (derived from the multinomial formula for each cell), and look for a set of parameters that satisfy these inequalities. If none do (the usual case for the Bayesian bootstrap), we minimize a quadratic norm of the negative parts of the inequalities. These steps are repeated a large number of times and then a credible set (or confidence interval) is formed that covers the upper and lower bounds of our parameters a given percentage of the simulated samples.

3.1 Nonparametric Empirical Results.

Table 3 provides estimates of the identified set and both 95% and 99% confidence sets for the model with a single κ and different cell size cutoffs. Since the 99% confidence set was identical for a minimum cell size of 50 and 100, we use a minimum cell size of 50 and 99% confidence sets throughout the paper.

Figure 2 plots the objective function (the sum of squared negative parts of moments) for this case, where lower values imply a better fit to the data. The units on the x-axis are dollars per annum. The jumps in the objective function occur when an additional constraint(s) is violated. The flat segment enclosed by the gray lines defines our estimate of the identified set, which is $\kappa \in (\$102, \$186)$, while the 99% credible set is the interval between the red dashed lines ($\kappa \in (\$78, \$450)$).[18] Though the confidence set is rather wide, its upper bound is less than half of the average estimate of $\kappa = \$1047$ per year from prior research on this data using

[18]We also investigated the robustness of our result with respect to the number of bootstrap samples, say ns. We report results with $ns = 500$, but the results did not vary provided $ns \geq 100$.

Table 3: Nonparametrics ID Set and Confidence Intervals for κ

Sample	**ID Set (κ)**		**95% CI**		**99% CI**	
	LB	UB	LB	UB	LB	UB
Min cell size = 20	\$150	\$168	\$150	\$270	\$150	\$354
Min cell size = 50 (main sample)	**\$102**	**\$186**	**\$78**	**\$294**	**\$78**	**\$450**
Min cell size = 100	\$78	\$186	\$78	\$450	\$78	\$450

Note: The table shows estimates from our nonparametrics estimator of the identified set and 95% and 99% confidence sets (or credible sets) for the switching cost κ in dollars per year.

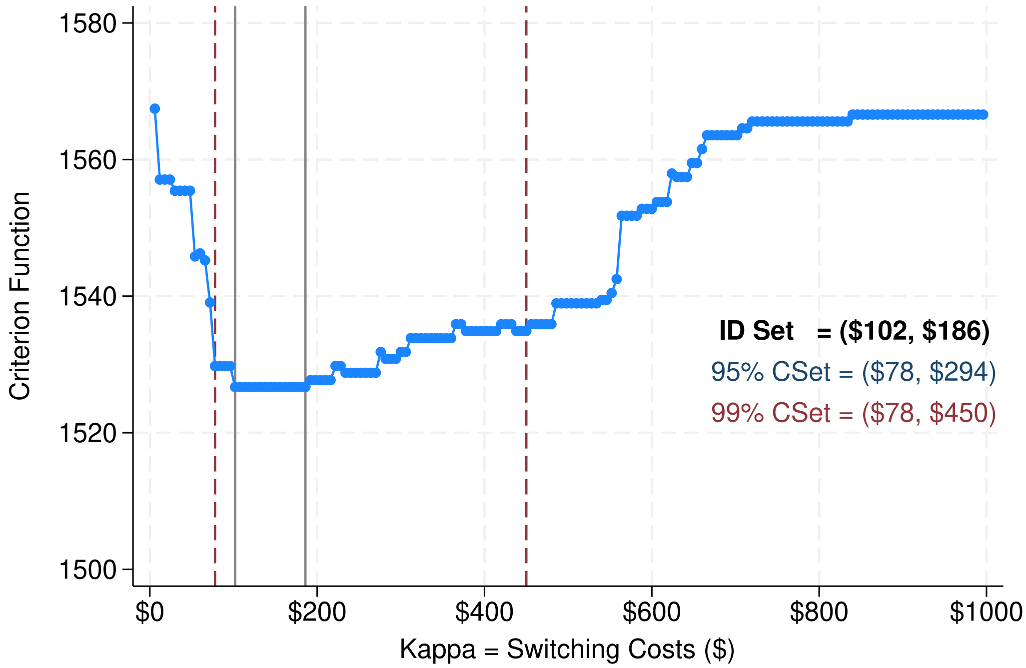
a standard multinomial logit model with observable preference heterogeneity and coarse fixed effects (see Shepard (2022)). We come back to implications of this difference below.

Figure 3 shows the underlying estimates from the lower-bound ($s = t - 2$) and upper bound ($s = t - 1$) moments; the objective function in Figure 2 is the sum of these two. The lower bound moments show that the data strongly reject $\kappa < 0$, and, for positive values of κ, the objective function is relatively flat. The objective function for the upper bound moments rises gradually and then more steeply after about $\kappa = \$300$, corresponding to the upper bound of the 95% confidence set.

Initial Conditions Estimators. We know the first time a consumer enters the Massachusetts exchange. So provided we are willing to add the assumption that the agents know their own choice specific fixed effects before making their first choice on the exchange, we can apply the non-parametric generalization of the initial condition estimator described above and in Appendix 6.3 to estimate κ.

When we combine both the "standard" moments (used in Figure 2) and the initial conditions moments, we get a larger value for the identified set (in this case, a point) of $\kappa = \$480$ but a relatively similar 99% confidence set of $\kappa \in (\$180, \$480)$ (see Appendix Figure 6). However, when we use *only* the initial conditions moments, we get rather different results. The size of the sample that only uses the initial condition moments is about a third of that used for the estimates that condition on the prior choice but, as noted in section 2, the initial condition moments can obtain bounds from fewer switches. Using the initial condition moments alone

Figure 2: Nonparametrics Estimator Objective Function



Notes: The figure shows the objective function (or "criterion function," which is the sum of squared moments) for the main nonparametrics estimator. The x-axis is the switching cost parameter (κ) in dollars per year. The on-graph note indicates the identified set and 95% and 99% credible sets.

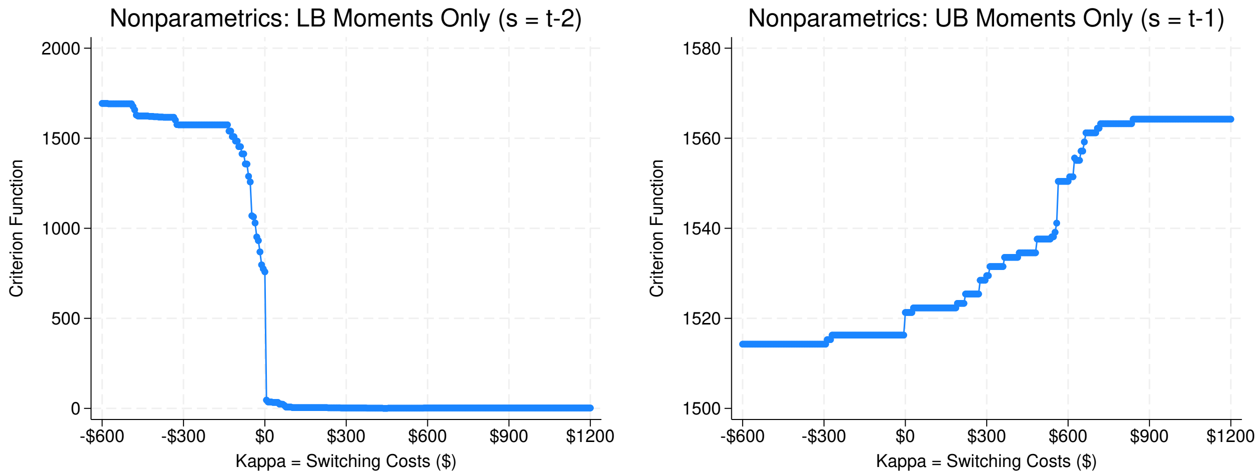
generates a lower bound on κ of $600 and does not generate an upper bound. So the confidence set generated by the initial condition estimator does not intersect that from the data that conditions on the initial choice and is relatively uninformative.

This seems largely to be a result of a lack of power in the initial condition moments. The upper bound is found by comparing the proportion of switches in the period following entry to the shares at entry within our cells. To get an upper bound the switches must be larger than the shares resulting from the initial choices. The initial choices roughly distribute like the shares of the various plans in the data, but the proportion of switches is always less than the share of the smallest plans.[19]

Heterogeneity in the cost of switching relative to price. Next we consider whether the relationship of switching cost to price varies by income. Observations with income (denoted $I_{i,t}$)

[19]This likely would be problematic if we were interested on the determinants of plan preferences, but is much less so when the interest is in the determinants of switching behavior and we condition on the individual specific plan preferences embodied in the fixed effects.

Figure 3: Nonparametrics Estimator: Lower and Upper Bound Moments Separately



Notes: The figure shows the objective function (or "criterion function," which is the sum of squared moments) for the main nonparametrics estimator, separately for lower bound ($s = t - 2$) and upper bound ($s = t - 1$) moments. The x-axis is the switching cost parameter (κ) in dollars per year. The criterion function in Figure 2 is the sum of these two sets of moments.

less that 100% of the federal poverty level (or FPL) were fully subsidized.[20] So for income we distinguish between those with income between 100-200% of the FPL and those with 200-300% of the FPL and use the structural utility model

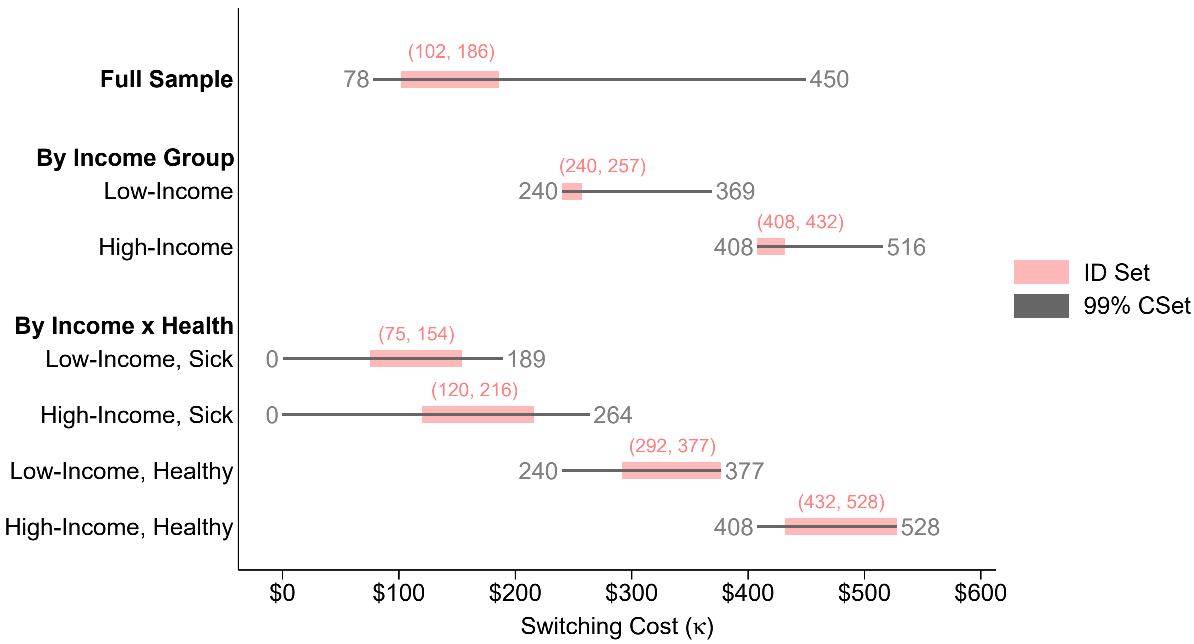
$$SU_{d,i,t} = -p_{d,i,t}\big(1 + \gamma_0 \cdot \{2FPL < I_{i,t} \leq 3FPL\}\big) - \{y_{i,t-1} \neq d\}\delta_0 + \lambda_{d,i}.$$

Recall that we can divide by a person-specific positive constant without changing the implications of the model, so we set the low-income price coefficient equal to one. This makes the ratio of switching costs to price for low-income observations equal to δ_0, but that ratio for high-income people is now $\kappa_0 = \delta_0/(1 + \gamma_0)$.

Estimates of the identified set and confidence sets for the nonparametrics estimator with heterogeneity are shown in Figure 4 and Table 4. The top panel (labeled "Full Sample") shows the results reported above from the model without heterogeneity. The next panel (labeled "By Income Group") shows estimates separately for the low-income and high-income groups. The two confidence sets do not intersect, and as might have been expected, their bounds correspond roughly to kinks in the objective function in Figure 2. We conclude that the relative importance

[20]The FPL was $10,380 in 2010, and was adjusted by the CPI-U thereafter.

Figure 4: Nonparametrics Estimator with Heterogeneity by Income and Sickness



Notes: The figure shows estimates from the nonparametrics estimator with heterogeneity by income and sickness. The top panel (labeled "Full Sample") shows results reported in Table 3 from the model without heterogeneity. The next panel (labeled "By Income Group") shows estimates separately for the low-income (100-200% FPL) and high-income (200-300% FPL) groups. The bottom panel (labeled "By Income x Health") shows estimates separately for income x health groups, where sickness is defined as above or below the median of the medical risk score distribution in year t. The red bar indicates the identified set, and the gray line indicate the 99% confidence sets.

of switching costs, relative to price, is lower for higher income people among this low income population.[21]

Notice also that once we distinguish different income groups the confidence intervals narrow considerably (they are about a third the width of the confidence set for the model with a single κ_0). This convinced us to push the analysis one step further and distinguish between more and less healthy observations. To that end we use the medical risk score (described above) and let the sickness variable, our $S_{i,t}$, equal one for those above its median value and zero for those below.

The impact of health on the cost of switching insurers is not clear a priori. It could be that

[21] The alternative non-parametric panel data discrete choice model in Pakes and Porter (2024) which allows for fixed effects and arbitrarily serially correlated disturbances, but did not allow for state dependence and used different moments from the same underlying data, also obtained the result that higher income people in this low income population were less sensitive to price.

Table 4: Switching Cost (κ) Estimates in Nonparametrics Heterogeneity Model

	Low-Income Group	**High-Income Group**
By Income-Only Model		
ID Set	**[\$240, \$257]**	**[\$408, \$432]**
95% CSet	*[\$240, \$343]*	*[\$408, \$480]*
99% CSet	*[\$240, \$369]*	*[\$408, \$516]*
By Income x Health Model		
Healthy ($s_{it} = 0$): ID Set	**[\$292, \$377]**	**[\$432, \$528]**
95% CSet	*[\$240, \$377]*	*[\$408, \$528]*
99% CSet	*[\$240, \$377]*	*[\$408, \$528]*
Sick ($s_{it} = 1$): ID Set	**[\$75, \$154]**	**[\$120, \$216]**
95% CSet	*[\$0, \$171]*	*[\$0, \$240]*
99% CSet	*[\$0, \$189]*	*[\$0, \$264]*

Note: This table shows estimates of identified sets and 95% and 99% confidence sets for the nonparametrics estimator with heterogeneity in switching costs, as also shown in Figure 4. Low-income indicates income between 100-200% of the Federal Poverty Line (FPL) while high-income indicates income between 200-300% FPL. People with income less than poverty did not pay any premiums, so their price coefficients are not identified. Sick ($S_{it} = 1$) indicates that an individual's medical risk score in year t is above the median in the population distribution.

being sick makes one more sensitive to price as it impacts perception of likely future wealth. Alternatively if patients have developed relationships with plan specific health care providers, being sick could make patients more hesitant to switch. The model we use to investigate this is

$$SU_{d,i,t} = -p_{d,i,t}\big(1 + \gamma_0 \cdot \{2FPL < I_{i,t} \leq 3FPL\}\big) - \{y_{i,t-1} \neq d\}\big(\delta_0 + \delta_s \cdot \{s_{i,t} = 1\}\big) + \lambda_{d,i},$$

so now the relative cost of switching differs with both income and health.

The final panel of Figure 4 and Table 4 (labeled "By Income x Health") presents the results. The confidence intervals for the healthy group are shifted up from those that did not condition on health status, the shift being most noticeable for the upper bound. The sicker part of both income groups have much lower switching costs, with an upper bound less than half of that for the non-sick group. The confidence sets of the sicker and less sick groups for a given income group do not intersect. Moreover, though the identified set for the less sick groups are bounded away from zero and fairly narrow, the data does not have enough power to reject a lower bound greater than zero for the sick groups.

4 Parametric Models With and Without F.E.s.

The prior literature on panel data discrete choice models that allowed for state dependence used parametric models and did not allow for fully flexible individual and choice specific fixed effects. In the case of health insurance choice, this generated estimates of "switching costs" from the CommCare data that were more than twice as high as the maximal upper bound reported in Table 3. We first show that we obtain similar results when we estimate the models without fixed effects using the cuts of the data we use. We then consider parametric functional forms with fixed effects. Throughout whenever we require a specific parametric distribution for the disturbances we assume that distribution to be logistic.[22]

Parametric models without fixed effects.

Table 5 presents the results from an assortment of models that allow for state dependence but do not allow for choice specific fixed effects that differ by person. The first four columns present the results for models with increasing numbers of observed explanatory variables. These are standard multinomial logit choice models with utility function:

$$U_{i,d,t}^{MNL} \;=\; \gamma \cdot P_{i,d,t} \;+\; X_{i,d,t}'\beta \;+\; \xi_{d,t}(Z_i) \;+\; \delta \cdot 1\{y_{i,t-1} \neq d\} \;+\; \epsilon_{i,d,t} \tag{4.1}$$

where $P_{i,d,t}$ is consumer price; $X_{i,d,t}'$ is a vector of plan attributes (e.g. hospital networks) whose value can across individuals; $\xi_{d,t}(Z_i)$ are a set of plan dummies that can vary across regions, demographics, and health status; δ is the switching cost; and $\epsilon_{i,d,t}$ is the "logit" error.

Like prior research on alternative data sets (e.g., Figure 5 in Heiss et. al. 2021), as we add explanatory variables the estimates of the importance of state dependence relative to price falls. Our base case, which includes only plan fixed effects and price, generates an estimate of \$1382.5 (column (1)). When we allow for plan dummies interacted with each of age-sex, region, network and chronic illness dummies, a model with 249 parameters, we get an estimate of the ratio of switching costs relative to price of \$1018 (column (3)). Then we add three variables constructed

[22] We also add back in the data on individuals who changed their choice set across periods as a constant choice set is not needed once we specify the distribution of disturbances. This does increase sample size (see Appendix Table 7). We also estimated all models used in this section with the data used in the non-parametric section of the paper, but there was very little difference in any of the results.

Table 5: Multinomial Logit Estimation

	Simple (1)	Plan Dummies (2)	Detailed Plan Dum. (3)	Detailed Plan Dum. + Network (4)	Plan Dum. + Rand. Eff. (5)	Include New Enr (6)	New Enr + Rand. Eff. (7)
Normalize $\epsilon_{i,d,t}$ to EV1							
Switching Cost (δ)	-4.340	-4.481	-4.466	-4.427	-4.809	-4.221	-5.003
	(0.008)	(0.008)	(0.009)	(0.009)	(0.058)	(0.007)	(0.065)
Price ($ per month) (γ)	-0.038	-0.049	-0.053	-0.054	-0.050	-0.035	-0.039
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)
Hospital Network Utility	—	—	—	0.201	—	—	—
				(0.012)			
Prev. Used Hospitals Covered	—	—	—	0.656	—	—	—
				(0.025)			
Prev. Used×Partners Hosp.	—	—	—	1.122	—	—	—
				(0.034)			
Normalize $\gamma \equiv 1$							
Switching Cost ($/year)	1,382.5	1,101.3	1,018.4	986.8	1,149.1	1,429.8	1,527.0
($\kappa = 12 * \delta/\gamma$)	(7.1)	(4.8)	(4.5)	(4.3)	(16.5)	(5.8)	(20.4)
Plan Dummies	—	Yes	Yes	Yes	Yes	Yes	Yes
Plan ×(Area, Age–Sex, Illness)	—	—	Yes	Yes	—	—	—
Plan Random Effects	—	—	—	—	Yes	—	Yes
N Parameters	2	7	249	252	11	7	11
N Individuals × Years	2,134,763	2,134,763	2,134,763	2,134,763	213,575	3,476,550	348,242

by Shepard (2022) to capture an individual-specific value assigned to each plan and individual-specific prior experience with each plan's network (column (4)).[23] As in Shepard (2022) all three are highly significant, but they only reduce the estimate of switching costs relative to price to $987.

We also considered models with random normal plan effects. These are estimated using simulated maximum likelihood. Column (5) provides the results of a model which conditions on the initial choice and assumes the random normal effects are independent of it, as well as the other right hand side variables. Column (7) is an initial conditions estimator. It adds the initial choices to the choices to be explained and allows for the random plan effects. For comparison column (6) adds the initial choice but does not use the random normal plan effects. All of these generate estimates of the switching cost relative to price that are considerably more than $1,000.

[23] These variables are a "network utility" variable calculated from the indirect utility from a hospital choice model, which captures the option demand value of access to the plan's network, and two variables which measure the share of a patient's previously used hospitals covered by each plan, capturing prior experience with the network's hospitals (one allows for interaction with Partners Healthcare hospitals, for which loyalty seemed to be especially strong).

We conclude that models of state dependence on our data that do not allow for fixed effects generate estimates that are twice as large as the upper bound to the estimates that we obtained once we allowed for fixed effects.

Parametric models with fixed effects.

Once we specify a particular distribution for the disturbances the shape restrictions of that distribution will generally allow for identification of separate price and state dependence coefficients. For ease of exposition we develop the relevant parametric inequalities for a model in which the switching cost and price coefficients do not vary over time or across individuals. So equation (2.1) becomes

$$U_{d,i,t} = -p_{d,it}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq d\}\delta_0 + \lambda_{d,i} + \epsilon_{d,i,t}, \tag{4.2}$$

and we assume $(\gamma_0, \delta_0) > (0, 0)$. Our focus is still on $\kappa_0 \equiv \delta_0/\gamma_0$. We begin with "generic" parametric inequalities; i.e. inequalities that are available regardless of the parametric error distribution assumed.

Generic Parametric Inequalities. Provided d and c are feasible choices in both t and s, and d was chosen in period t while c was chosen in period s, the revealed preference double difference that enables us to difference out the individual and choice specific fixed effects (the analogue of equation 2.2 above) becomes

$$-\Delta\Delta p_{i,t,s}^{d,c}\gamma_0 - \Delta\Delta sw_{i,t,s}^{d,c}(y_{i,t-1}, y_{i,s-1})\delta_0 + \Delta\Delta\epsilon_{i,t,s}^{d,c} \geq 0,$$

where we now assume $\Delta\Delta\epsilon_{i,t,s}^{d,c}$ has the known distribution $\mathcal{F}(\cdot)$. Since the model implies that $\Delta\Delta\epsilon_{i,t,s}^{d,c} \geq \Delta\Delta p_{i,t,s}^{d,c}\gamma_0 + \Delta\Delta sw_{i,t,s}^{d,c}\delta_0$, we have

$$\underbrace{Pr(y_{i,t} = d, y_{i,s} = c | y_{i,s-1}, p_i, \lambda_i)}_{\equiv Pr_{t,s}^{d,c}} \leq 1 - \mathcal{F}\left(\Delta\Delta p_{i,t,s}^{d,c}\gamma_0 + \Delta\Delta sw_{i,t,s}^{d,c}(y_{i,t-1}, y_{i,s-1})\delta_0\right).$$

Notice that this inequality did not require the same choice set in both periods, just that we are willing to specify the joint distribution of the disturbances in t and s. This is true for all the parametric inequalities. It enables us to increase the size of the data set used in our comparisons by more than 40%, and is likely to generate even larger increments in studies of retail markets where choice set changes are typically more frequent.

Since $\mathcal{F}(.)$ is a strictly increasing distribution function, we can rearrange and use its inverse to yield

$$\Delta\Delta p_{i,t,s}^{d,c}\gamma_0 + \Delta\Delta sw_{i,t,s}^{d,c}(y_{i,t-1}, y_{i,s-1})\delta_0 \leq \mathcal{F}^{-1}(1 - Pr_{t,s}^{d,c}). \tag{4.3}$$

$\Delta\Delta p_{i,t,s}^{d,c}$ and $\Delta\Delta sw_{i,t,s}^{d,c}(\cdot)$ can be calculated for each cell in our data. So consistent estimates of $Pr_{t,s}^{d,c}(\cdot)$, say $\hat{Pr}_{t,s}^{d,c}(\cdot)$, and this inequality can be used to bound estimates of (γ_0, δ_0).

As noted in our data $\hat{Pr}_{d,c}^{t,s} < .5$ in all cells, so $\mathcal{F}^{-1}(1 - Pr_{t,s}^{d,c})$ is always positive. By plotting the boundary of inequality (4.3) for the cases where $\Delta\Delta p_{i,t,s}^{d,c}$ is either positive or negative, and $\Delta\Delta sw_{i,t,s}^{d,c}(\cdot)$ is either positive, negative, or zero, Appendix 6.2 shows that if $\mathcal{F}^{-1}(1 - Pr_{t,s}^{d,c})$ is always positive inequality (4.3) can provide upper bounds for both parameters, but lower bounds for neither. Since $\kappa_0 = \delta_0/\gamma_0$, without a lower bound on γ_0 we do not get an upper bound on κ_0.

Distribution Specific Inequalities. Having made parametric distributional assumptions, additional bounds will typically still be available, but they will differ with the form of the assumed ϵ distribution. We work with the logit case, as that has been used extensively in the related literature. This also generates a distribution for the double difference of disturbances (for $\Delta\Delta\epsilon_{i,t,s}^{d,c}$) that is analytic, simplifying computation.[24]

Assumption 4.1. *The utility function is equation (2.1), and $\varepsilon_{1,i,t}, \ldots, \varepsilon_{\mathcal{D},i,t}$ are independent and identically distributed across choices, where $\varepsilon_{1,i,t}$ has a standard Gumbel distribution.*

Recall how we obtained the non-parametric lower bound. We considered a group who started at c in $t-3$ and stayed at c in $t-2$, then substituted away from c in $t-1$, and never switched back in t despite the fact that the relative price of c in t was lower then it was in $t-2$ (see the discussion leading to equation 2.4). That logic also generates a lower bound of the same

[24]That distribution and its density are

$$\mathcal{F}(y) = \frac{exp(y)(y-1)+1}{(exp(y)-1)^2}, \text{ and } f(y) = \frac{exp(y)(exp(y)(y-2)+y+2)}{(exp(y)-1)^3}.$$

form for the parametric logit case. That is, if for the group of people who started at c in $t-3$, $Pr(y_{i,t}=c|y_{i,t-3}=c,p_i,x_i) < Pr(y_{i,t-2}=c|y_{i,t-3}=c,p_i,x_i)$, then $\delta_0/\gamma_0 \geq .5\min_{d\in\mathcal{D}}\Delta\Delta p^{d,c}_{i,t,t-2}$ which will lead to a positive lower bound for δ_0/γ_0 if the relative price of c in t, relative to all other goods, is less than it was in $t-2$. Of course if Assumption 4.1 holds, then use of the parametric form to obtain the bounds generates different lower bounds than those obtained in the non-parametric section.

The method of proof for the parametric and non-parametric lower bounds are different. The non-parametric proofs use Jensen's inequality to bound the squared average probability by its variance. With parametric functional forms that variance depends on the unknown distribution of the λ_i, so to obtain the lower bound from the logit assumption we work directly with the individual specific inequalities generated by assumption (4.1). We were unable to find an upper bound in this way. However there is a special feature of the logit distribution which does generate both upper and lower bounds. If we condition on being at c in period $s-1$, then the log odds ratio of continuing to c and then switching to d, relative to choosing d and then switching to c, also bounds γ_0/δ_0. The next theorem formalizes the bounds for the logit case and is proven in the Appendix.

Theorem 4.2. *Suppose Assumption 4.1 holds. Then, in addition to the generic inequalities in equation (4.3), we have*

A. *if* $\mathcal{D}_t=\mathcal{D}_{t-2}\equiv\mathcal{D}$, $\min_{d\neq c,d\in\mathcal{D}}\Delta\Delta p^{d,c}_{i,t,t-2}>0$ *and* $Pr(y_{i,t}=c|y_{i,t-3}=c,p_i,x_i) < Pr(y_{i,t-2}=c|y_{i,t-3}=c,p_i,x_i)$ *then*

$$\kappa_0=\delta_0/\gamma_0>\frac{1}{2}\Delta\Delta p^{d,c}_{i,t,t-2}$$

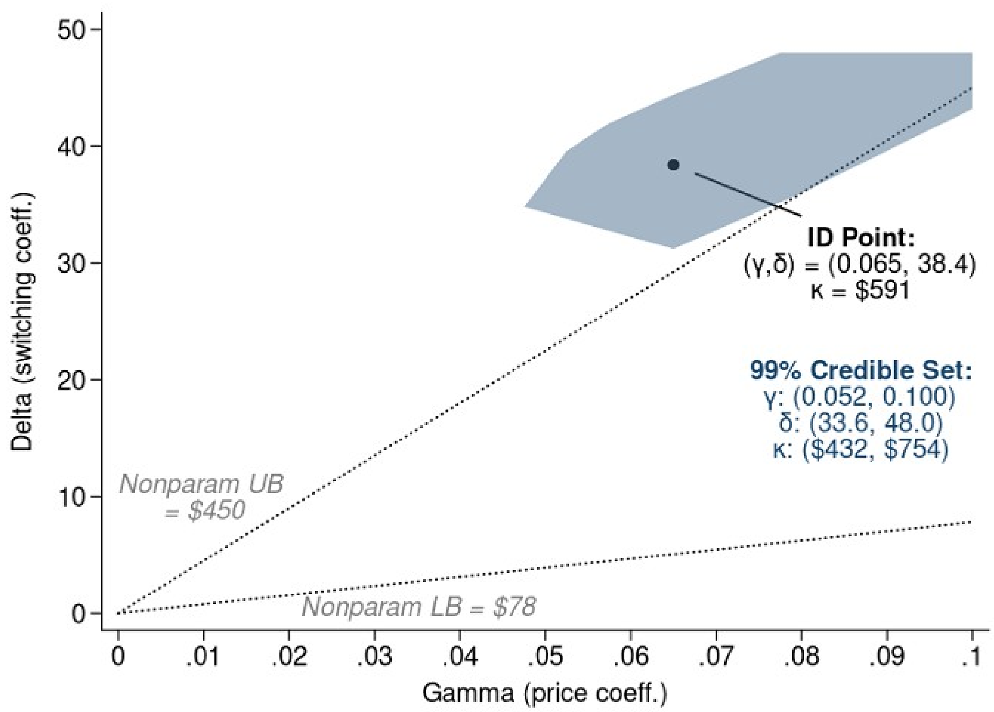
B. *and if* $s\in\{t-1,t-2\}$ *and* $(d,c)\in\mathcal{D}_t\cap\mathcal{D}_s$, *then*

$$exp\left[\Delta\Delta p^{c,d}_{i,t,s}\gamma_0\right] \;\leq\; \frac{\Pr(y_{i,t}=d,y_{i,s}=c\,|\,y_{i,s-1}=c,p_i,x_i)}{\Pr(y_{i,t}=c,y_{i,s}=d\,|\,y_{i,s-1}=c,p_i,x_i)} \leq\; exp\left[2\delta_0+\Delta\Delta p^{c,d}_{i,t,s}\gamma_0\right].$$

Parametric Empirical Results with Fixed Effects. As noted once we allow for a parametric distribution of the disturbances we can compare sequences of choices that faced different choice sets. Appendix Table 7 provides the analogue of the sample size Table 1 for the parametrics estimator data set that allows for these changes, and there you can see that it expands the number of sequences we can use. However, the results from the parametric model with the expanded data and with the non-parametric data set were not noticeably different from one another.

Figure 5 provides the results from the expanded data set when we stack the generic parametric inequalities in equation (4.3) with the inequalities from Theorem 4.2. The dotted line provides the upper and lower bound for $\kappa_0 = \delta_0/\gamma_0$ from the non-parametric model with a single κ_0. The shaded area is the 99% confidence set for (δ_0, γ_0) from the parametric inequalities.

Figure 5: Parametrics Estimator with All Parametric Inequalities



Notes: The figure shows results from the parametric model with fixed effects when we stack all moments, including the generic parametric inequalities in equation (4.3) and the inequalities from Theorem 4.2. For references, the 99% confidence set from the non-parametric model is shown as a dotted line. The black dot is the identified point, and the blue shaded area is the 99% confidence set for (δ, γ).

There are no values of (δ_0, γ_0) that satisfy all the inequalities so the "identified set" generated by these inequalities is the point that minimizes the objective function. The estimated value of κ_0 and its confidence interval are \$591 and (\$432, \$754) respectively. The point estimate is clearly outside the non-parametric confidence set, and the two confidence sets barely overlap. Still we note that the confidence sets obtained from the parametric model that allows for fixed effects only about half of those obtained from the models that did not allow for fixed effects.

The difference between the parametric and non-parametric results seems to be a result of part B of theorem (4.2); i.e. that the ratio of the log odds ratios do not depend on the fixed effects. When we only used the generic parametric inequalities, the logit inequality for the lower bound that is the analogue of the non-parametric inequality for the lower bound (part A of that theorem), and the non-parametric inequalities, we obtained an estimate of the identified set for κ_0 of $(\$141, \$187)$, very similar to the nonparametric estimate of ($102, $186). The upper and lower bound of the confidence set for κ_0 this generated were $(\$76, \$452)$; almost identical to the non-parametric confidence set of $(\$78, \$450)$.

4.1 Counterfactual Comparisons.

We now explore whether the difference between the κ_0 bounds obtained from the inequality estimator, and the κ_0 estimates obtained from the comparison models that allow for state dependence but not individual-by-product specific fixed effects, is likely to have economically important implications for a counterfactual of interest. Figure 1 showed that BMC, the largest plan with over a third of the market in 2011 (see Table 2), increased its relative price dramatically in 2012 and then decreased it by an even greater amount in 2013. We consider predictions for what would have happened had they instead kept their price constant at the average of the 2012 and 2013 prices in those two years.

The calculation conditions on the 2011 choices of enrolled individuals. We then predict BMC's market share in 2012 twice; once using the actual and once the counterfactual prices. Finally, we use these predictions and the actual and counterfactual prices in 2013 to obtain the predicted shares from the counterfactual policy for the two year period from 2011-2013. The predictions for these sequences are done in pairs, one of which uses the (γ_0, δ_0) estimates from a comparison model in Table 5, the other uses the γ_0 estimate from the relevant comparison model but restricts δ_0 to equal $\gamma_0 \cdot \hat{\kappa}$ where $\hat{\kappa} = \$186$, the upper bound of the identified set from the non-parametric estimates. The latter need not equal what our model would predict, as that would require either a model or bounds for the $\{\lambda_{i,d}\}_{i,d}$. Still, the difference between the two predictions should provide an indication of whether the implications of a model that allowed for fixed effects are likely to be different than a model which does not.

Table 6 provides the results. The bottom row shows that the average actual BMC premiums, averaged over all incumbent enrollees who were not in the below-poverty group (and hence paid

Table 6: Counterfactual Comparisons ($\kappa = \$186$)

Specification	2011 market shares	2012			2013		
		status-quo	cfactual	diff	status-quo	cfactual	diff
Market shares without imposing κ							
Plan FE	0.357	0.290	0.327	0.037	0.338	0.333	-0.005
Plan × Region FE	0.357	0.291	0.327	0.036	0.341	0.334	-0.007
Plan FE + RE	0.357	0.284	0.324	0.040	0.327	0.328	-0.001
Market shares imposing κ							
Plan FE	0.357	0.029	0.151	0.123	0.545	0.156	-0.389
Plan × Region FE	0.357	0.050	0.197	0.147	0.531	0.215	-0.316
Plan FE + RE	0.357	0.048	0.205	0.157	0.607	0.264	-0.343
Annual Premium	\$701	\$1,171	\$847	—	\$554	\$878	—

premiums), was \$701 per year in 2011. In 2012 that average increased to \$1171, and in 2013 it fell to \$554; the changes that generated the sharp spike in the price plot in Figure 1. We consider counterfactual prices that equal the average of the prices in 2012 and 2013 in each income group, and then hold that price fixed in both years. That results in an average price of \$847 in 2012 and \$878 in 2013 (with the slight difference coming from changes in the relative size of different income groups in the two years).

The actual predictions differ somewhat between the pairs defined by the comparison models but their qualitative nature does not. The fall in price in 2012 from the \$1,171 to \$847 leads to a prediction of a 3.7 to 4.0 point increase in share when we use the parameters estimated by the comparison models, but a prediction of a share increase that was more than three times that (12.3 to 15.7 points) when we constrain $\hat{\kappa} = 186$. In 2013 when the counterfactual average price was \$878 compared to the actual average price of \$554, the estimates predict a change of share of less than 1.0 point, whereas when we use $\hat{\kappa} = 186$ the higher counterfactual price in 2013 generates a two period prediction of a 30 to 40 point share fall. The impact of the higher κ_0 estimates on the comparison models' prediction in any one year spills over to the following years, making longer term predictions particularly problematic.

5 Conclusion.

We have provided both empirical results on switching costs in health insurance choices and methodological results on estimating models which allow for both individual by choice specific fixed effects and state dependence.

Our empirical results indicate lower estimates of state dependence than do models that do not allow for very flexible unobserved heterogeneity in our low income sample. When we hold the ratio of the switching cost to price constant but allow the price coefficient to vary arbitrarily, we estimate an upper bound to that ratio of $516 per annum and a lower bound of $78. The upper bound is about half of what prior research and our own analysis generates from models that do not allow for individual by choice specific fixed effects on the same data. Our estimates are based on comparing sequences of choices of individuals who have similar observed sequences of characteristics, the same choice set, and the same initial choice leading into the comparison periods.

When, in addition, we allow the ratio of switching cost to price to differ between the lower and the higher income parts of our low income sample, the results sharpen considerably. The higher income subset of the data generate a confidence set with a lower bound of $408 and an upper about $528, while the lower income subsets generate a confidence set of [$240, $377]. Differentiating the ratio of switching cost to price further by allowing it to depend on health status seems also to be important. The sicker half of our subset in each income class have confidence sets entirely below those we obtain from the sicker parts of each income class; their upper bounds are $189 and $264 respectively, about half of those for the less sick group in the two income classes.

The differences between these estimates and estimates that do not allow for individual by plan fixed effects seem large enough to make significant differences in the likely impacts of price movements on switching behavior, and the longer the prediction horizon the larger the range of possible biases. So it appears important to allow for flexible individual-level preferences, likely because of the very heterogeneous way that similar consumers value plan provider networks (the key plan attribute in our context). For instance, people may care very strongly about whether *their* current doctor is covered in a given plan (Shepard, 2022; Tilipman, 2022), an individual-by-plan specific match factor that is not likely to be captured with coarse plan interactions.

Our methodological results apply to the analysis of the increasingly available panel data discrete choice data sets. They are likely to be most relevant for repeated decisions on a set of choices that can be reversed if needed and/or have cognitive costs of re-evaluating the choices that rival the expected benefits of re-evaluation. The estimation algorithm stems from the implications of revealed preference when the utility function allows for flexible individual by product fixed effects, state dependence, and a non-parametric vector of product specific disturbances that distribute independently over time. This contrasts with related prior work on panel data discrete choice models with individual by product fixed effects that allows for arbitrary serial correlation in disturbances but no state dependence (Pakes and Porter 2024). The latter attributes all the dependence in choices over time that cannot be explained by observables to serial correlation in the disturbances, while the dynamic environment considered here attributes all this dependence to state dependence in an attempt to make our upper bounds to state dependence robust.

To obtain our bounds we use a double difference which differences out the fixed effects and then employ the inequalities implied when evaluated at the true value of the parameters of the utility function in a moment inequality estimation algorithm. Use of a Bayesian bootstrap (Kline and Tamer, 2020) makes the algorithm both easy to understand and to program. We show that the algorithm can accommodate switching costs that vary with observed characteristics, as well as arbitrary heterogeneity in the coefficient of price. Though the algorithm is likely to generate especially sharp parameter estimates from panel data problems in which there typically is a lot of switching (for e.g. sales in retail outlets), it also can be effective in cases where switching is typically less frequent, as in the repeated choices of insurance policies studied here.

From a methodological point of view much remains to be done. The non-parametric framework requires a constant choice set over comparison periods. We also provide a model that assumes a parametric form for the distribution of the disturbance vector and it can accommodate different choice sets over comparison periods. There are "generic" moment inequalities that apply to any parametric form of the disturbance vector, but they do not generate upper bounds to the switching cost coefficient. Upper bounds can typically be found but they will depend on the parametric form of the disturbance vector. In addition we have not investigated new computational methods for sharp identification (Mbakop (2023)) in our framework, nor have we asked when the framework would lead to point identification; both questions that are central to the methodological literature on moment inequalities.

References

Abaluck, J., and A. Adams-Prassl (2021): "What do consumers consider before they choose? Identification from asymmetric demand responses," *The Quarterly Journal of Economics*, 136(3), 1611–1663.

Abbring, J. H., J. J. Heckman, P.-A. Chiappori, and J. Pinquet (2003): "Adverse Selection and Moral Hazard in Insurance: Can Dynamic Data Help to Distinguish?," *Journal of the European Economic Association*, 1(2-3), 512–521.

Aguirregabiria, V., J. Gu, and Y. Luo (2021): "Sufficient statistics for unobserved heterogeneity in structural dynamic logit models," *Journal of Econometrics*, 223(2), 280–311.

Bonhomme, S. (2012): "Functional differencing," *Econometrica*, 80(4), 1337–1385.

Bronnenberg, B. J., J.-P. H. Dubé, and M. Gentzkow (2012): "The evolution of brand preferences: Evidence from consumer migration," *American Economic Review*, 102(6), 2472–2508.

Brot-Goldberg, Z., T. Layton, B. Vabson, and A. Y. Wang (2023): "The behavioral foundations of default effects: theory and evidence from Medicare Part D," *American Economic Review*, 113(10), 2718–2758.

Brown, Z. Y., and J. Jeon (2024): "Endogenous information and simplifying insurance choice," *Econometrica*, 92(3), 881–911.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 47(1), 225–238.

Chamberlain, G., and G. W. Imbens (2003): "Nonparametric Applications of Bayesian Inference," *Journal of Business & Economic Statistics*, 21(1), 12–18.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81(2), 535–580.

Conley, T. G., and C. R. Udry (2010): "Learning About a New Technology: Pineapple in Ghana," *American Economic Review*, 100(1), 35–69.

Dafny, L., K. Ho, and M. Varela (2013): "Let Them Have Choice: Gains from Shifting Away from Employer-Sponsored Health Insurance and Toward an Individual Exchange," *American Economic Journal: Economic Policy*, 5(1), 32–58.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2009): "Do switching costs make markets less competitive?," *Journal of Marketing research*, 46(4), 435–445.

——— (2010): "State dependence and alternative explanations for consumer inertia," *The RAND Journal of Economics*, 41(3), 417–445.

Ericson, K. M. M. (2014): "Consumer Inertia and Firm Pricing in the Medicare Part D Prescription Drug Insurance Exchange," *American Economic Journal: Economic Policy*, 6(1), 38–64.

Finkelstein, A., N. Hendren, and M. Shepard (2019): "Subsidizing Health Insurance for Low-income Adults: Evidence from Massachusetts," *American Economic Review*, 109(4), 1530–67.

Handel, B. R. (2013): "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103(7), 2643–82.

Heckman, J. J. (1981): "Statistical models for discrete panel data," *Structural analysis of discrete data with econometric applications*, 114, 178.

Heiss, F., D. McFadden, J. Winter, A. Wuppermann, and B. Zhou (2021): "Inattention and switching costs as sources of inertia in medicare part d," *American Economic Review*, 111(9), 2737–81.

Ho, K., J. Hogan, and F. Scott Morton (2017): "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program," *The RAND Journal of Economics*, 48(4), 877–905.

Honka, E. (2014): "Quantifying search and switching costs in the US auto insurance industry," *The RAND Journal of Economics*, 45(4), 847–884.

Honoré, B. E., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68(4), 839–874.

Honoré, B. E., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74(3), 611–629.

Honoré, B. E., and M. Weidner (2024): "Moment conditions for dynamic panel logit models with fixed effects," *Review of Economic Studies*, p. rdae097.

Hortaçsu, A., S. A. Madanizadeh, and S. L. Puller (2017): "Power to choose? An analysis of consumer inertia in the residential electricity market," *American Economic Journal: Economic Policy*, 9(4), 192–226.

Keane, M. P. (1997): "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business & Economic Statistics*, 15(3), 310–327.

Khan, S., F. Ouyang, and E. Tamer (2021): "Inference on Semiparametric Multinomial Response Models," *Quantitative Economics*, 12(3), 743–777.

Kline, B., and E. Tamer (2020): "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 7, 329–366.

Kroft, K., F. Lange, and M. J. Notowidigdo (2013): "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 128(3), 1123–1167.

Manski, C. F. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, pp. 357–362.

Mbakop, E. (2023): "Identification in Some Discrete Choice Models: A Computational Approach," ArXiv Working Paper.

McIntyre, A., M. Shepard, and M. Wagner (2021): "Can automatic retention improve health insurance market outcomes?," in *AEA Papers and Proceedings*, vol. 111, pp. 560–66.

Pakes, A., and J. Porter (2024): "Moment inequalities for multinomial choice with fixed effects," *Quantitative Economics*, 15(1), 1–25.

Polyakova, M. (2016): "Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D," *American Economic Journal: Applied Economics*, 8(3), 165–95.

Shcherbakov, O. (2016): "Measuring consumer switching costs in the television industry," *The RAND Journal of Economics*, 47(2), 366–393.

Shepard, M. (2022): "Hospital network competition and adverse selection: evidence from the Massachusetts health insurance exchange," *American Economic Review*, 112(2), 578–615.

Shepard, M., and M. Wagner (2025): "Do ordeals work for selection markets? Evidence from health insurance auto-enrollment," *American Economic Review*, 115(3), 772–822.

Shi, X., M. Shum, and W. Song (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *Econometrica*, 86(2), 737–761.

Starc, A. (2014): "Insurer pricing and consumer welfare: Evidence from medigap," *The RAND Journal of Economics*, 45(1), 198–220.

Tebaldi, P., A. Torgovitsky, and H. Yang (2023): "Nonparametric estimates of demand in the california health insurance exchange," *Econometrica*, 91(1), 107–146.

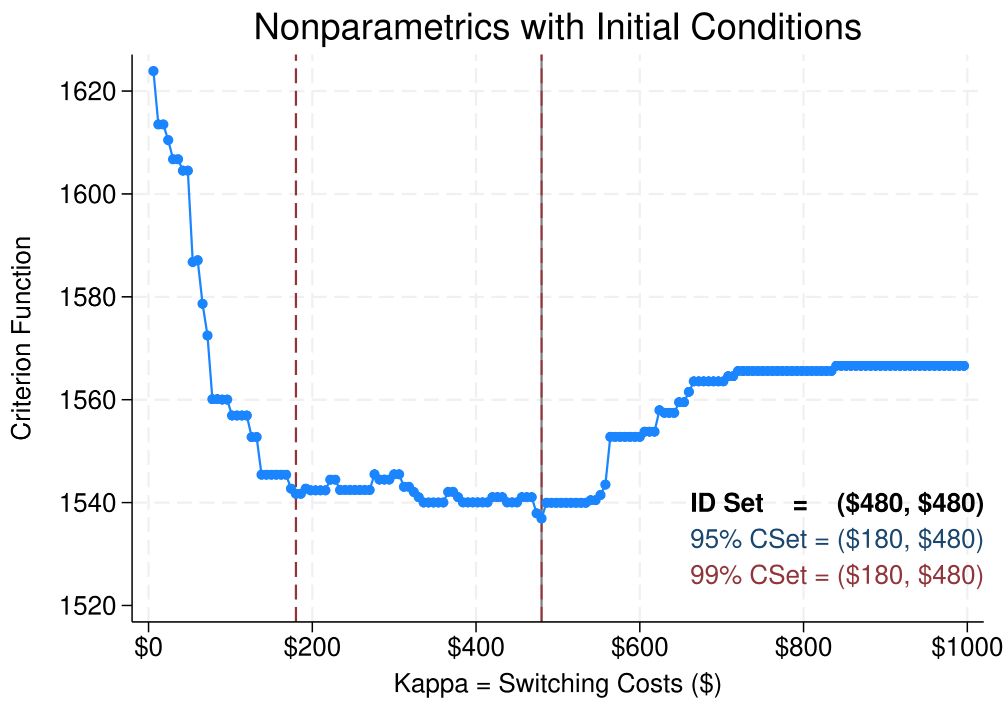
Tilipman, N. (2022): "Employer Incentives and Distortions in Health Insurance Design: Implications for Welfare and Costs," *American Economic Review*, 112(3), 998–1037.

Torgovitsky, A. (2019): "Nonparametric inference on state dependence in unemployment," *Econometrica*, 87(5), 1475–1505.

6 Appendix: For Online Publication Only

6.1 Additional Figures and Tables

Figure 6: Nonparametrics Estimator: Including Continuing Enrollees and Initial Conditions



Notes: The figure shows the objective function (or "criterion function," which is the sum of squared moments) for the nonparametrics estimator that includes both continuing enrollees and initial conditions moments. The graph is analogous to Figure 2, which is for the main nonparametrics estimator with only continuing enrollee moments. The x-axis is the switching cost parameter (κ) in dollars per year. The on-graph note indicates the identified set and 95% and 99% credible sets.

Table 7: Sample Sizes for the Parametrics Estimator

	Full dataset		Estimation sample ($N \geq 50$)		
Year Pair	Number of Members	Number of Groups	Number of Members	Number of Cells	Number of Groups
Year pairs: $s = t - 1$					
$(2009, 2010)$	53,167	100	6,332	14	10
$(2010, 2011)$	56,971	100	2,095	4	4
$(2011, 2012)$	58,359	125	31,160	30	23
$(2012, 2013)$	60,725	125	31,869	27	22
Year pairs: $s = t - 2$					
$(2009, 2011)$	27,753	100	5,840	25	18
$(2010, 2012)$	33,516	100	4,760	23	20
$(2011, 2013)$	35,302	125	16,589	54	42
Total	**325,793**	**775**	**98,645**	**177**	**139**

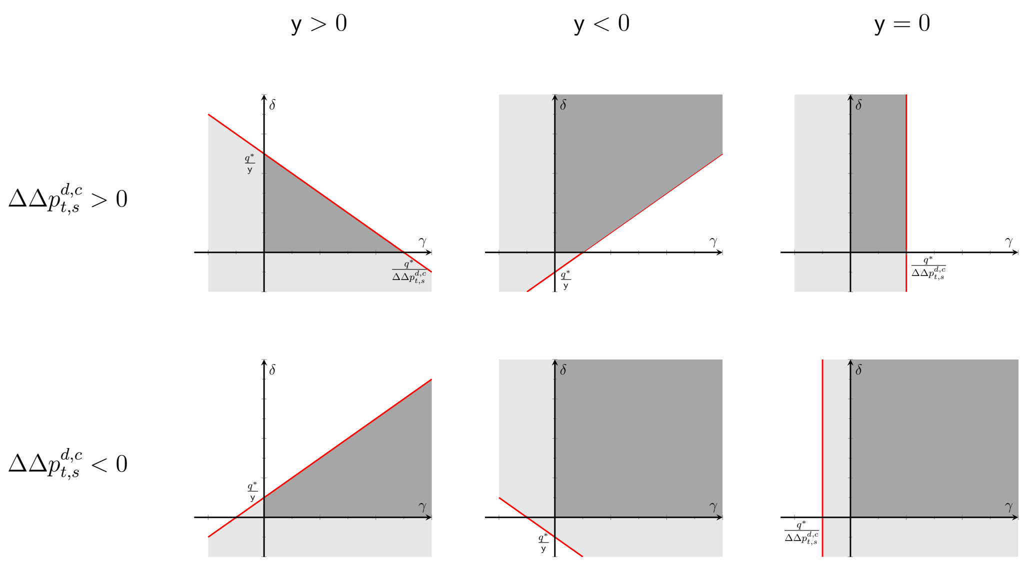
Notes: The table shows summary statistics for the parametrics estimator sample, by pair of years (s, t). See the text for definitions of cells and groups used in the estimation. The table lists the number of members and groups, both before and after applying the minimum cell-size cutoff of 50 members.

6.2 Generic Inequality Cases

Below we graphically display the information on the parameters (γ_0, δ_0) contained in the equation (4.3). Define $\mathsf{y} = \Delta\Delta sw^{d,c}_{i,t,s}(y_{i,t-1}, y_{i,s-1})$ and $q^* = \mathcal{F}^{-1}(1 - Pr^{d,c}_{t,s})$.

Figure 7 [8] considers the case where $q^* > 0$ [< 0]. If $\mathcal{F}(0) = 0.5$, then the case $q^* > 0$ [< 0] corresponds to $Pr^{d,c}_{t,s} < 0.5$ [> 0.5].

Figure 7: Inequalities for $q^* > 0$



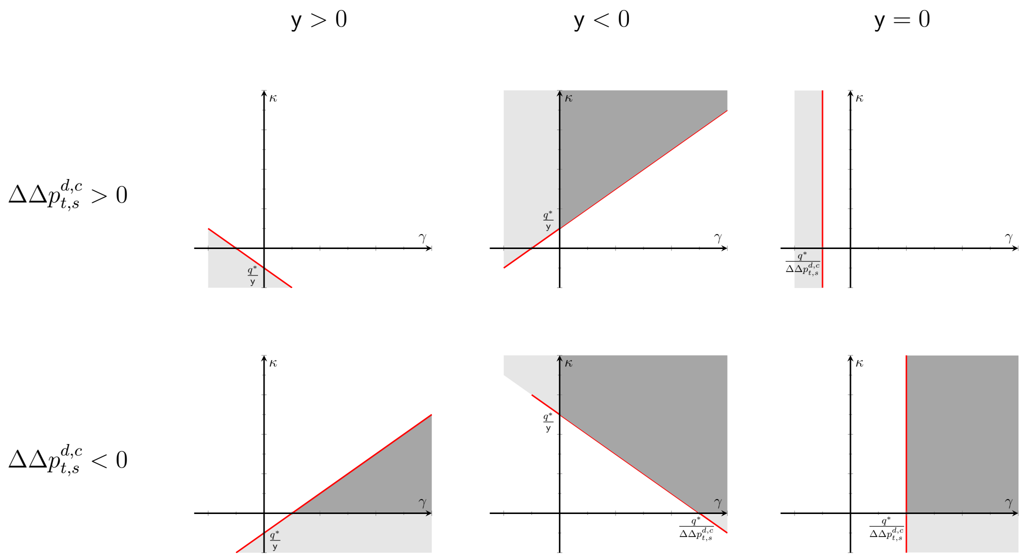
Two cases in Figure 7 corresponding to $\Delta\Delta p^{d,c}_{i,t,s} < 0$ and $\mathsf{y} \leq 0$ are uninformative. For these cases, the whole first quadrant satisfies the inequality. In our empirical work, we use the remaining four cases to inform bounds on κ_0. As noted previously, the case $q^* < 0$ does not occur in the empirical work, so we do not make use of the cases in Figure 8.

6.3 Initial Condition Bounds

In section 2.1, we illustrate moment inequalities that generate upper and lower bounds for switching costs. Here we repeat the analogous exercise for the case where the first period under consideration is the initial period for individual i with no lagged dependent variable. Suppose $s = 1$ is the initial period.

An upper bound for κ_0 can be obtained based on the first two periods. Suppose, for choice d^*, $SU_{d^*,i,2}(y_{i,1} = d^*) - SU_{d^*,i,1} \geq \max_{c \neq d^*} [SU_{c,i,2}(y_{i,1} = d^*) - SU_{c,i,1}]$. Then, $Pr(y_{i,2} =$

Figure 8: Inequalities for $q^* < 0$



$d^*|y_{i,1} = d^*, p_i, x_i) \geq Pr(y_{i,1} = d^*|p_i, x_i)$. So, violation of this probability inequality implies that $\kappa_0 < \max_{c \neq d^*} \Delta\Delta p^{d^*,c}_{i,2,1}$.

Next, consider the first three periods for a lower bound. Suppose, for choice d_* and $c\ ned$, $SU_{d_*,i,3}(y_{i,2} = c) - SU_{d_*,i,1} \geq \max_{c' \neq d_*} [SU_{c',i,3}(y_{i,2} = c) - SU_{c',i,1}]$. Then, $Pr(y_{i,3} = d^*|y_{i,2} = c, y_{i,1} = d^*, p_i, x_i) \geq Pr(y_{i,2} = c, y_{i,1} = d^*|p_i, x_i)$. So, if $\Delta\Delta p^{c',d}_{i,3,1} \geq 0$ for all $c' \neq c$, then violation of the probability inequality implies that $\kappa_0 > \Delta\Delta p^{c,d}_{i,3,1}$.

6.4 Proofs

First, we formalize (2.11) in a lemma. Recall the form of structural utility in (2.10).

Lemma 6.1. *Suppose Assumption 2.1 holds. Assume $t > s$, and $D_0 \subset \mathcal{D}$. If*

$$\min_{d \in D_0} [SU_{d,i,t}(y_{i,t-1}) - SU_{d,i,s}(y_{i,s-1})] \;\geq\; \max_{c \notin D_0} [SU_{c,i,t}(y_{i,t-1}) - SU_{c,i,s}(y_{i,s-1})],$$

then

$$\Pr(y_{i,t} \in D_0|y_{i,t-1}, p_i, x_i; \lambda_i) \;\geq\; \Pr(y_{i,s} \in D_0|y_{i,s-1}, p_i, x_i; \lambda_i). \;\square$$

Proof of Lemma 6.1:

For all $c \notin D_0$, $d \in D_0$,

$$SU_{c,i,s}(y_{i,s-1}) - SU_{d,i,s}(y_{i,s-1}) \geq SU_{c,i,t}(y_{i,t-1}) - SU_{d,i,t}(y_{i,t-1}).$$

Hence, for all $d \in D_0$,

$$\begin{aligned}
&\left\{\varepsilon_{i,s} \,\middle|\, \varepsilon_{d,i,s} \geq \max_{c \notin D_0} [SU_{c,i,s}(y_{i,s-1}) - SU_{d,i,s}(y_{i,s-1}) + \varepsilon_{c,i,s}]\right\} \\
\subseteq \;& \left\{\varepsilon_{i,t} \,\middle|\, \varepsilon_{d,i,t} \geq \max_{c \notin D_0} [SU_{c,i,t}(y_{i,t-1}) - SU_{d,i,t}(y_{i,t-1}) + \varepsilon_{c,i,t}]\right\} \\
= \;& \left\{\varepsilon_{i,t} \,\middle|\, SU_{d,i,t}(y_{i,t-1}) + \varepsilon_{d,i,t} \geq \max_{c \notin D_0} [SU_{c,i,t}(y_{i,t-1}) + \varepsilon_{c,i,t}]\right\}
\end{aligned}$$

So,

$$\begin{aligned}
&\Pr(y_{i,t} \in D_0 \,|\, y_{i,t-1}, p_i, x_i; \lambda_i) \\
&= \Pr\left(\bigcup_{d \in D_0} \left\{\varepsilon_{i,t} \,\middle|\, SU_{d,i,t}(y_{i,t-1}) + \varepsilon_{d,i,t} \geq \max_{c \notin D_0} [SU_{c,i,t}(y_{i,t-1}) + \varepsilon_{c,i,t}]\right\} \,\middle|\, \lambda_i\right) \\
&\geq \Pr\left(\bigcup_{d \in D_0} \left\{\varepsilon_{i,s} \,\middle|\, SU_{d,i,s}(y_{i,s-1}) + \varepsilon_{d,i,s} \geq \max_{c \notin D_0} [SU_{c,i,s}(y_{i,s-1}) + \varepsilon_{c,i,s}]\right\} \,\middle|\, \lambda_i\right) \\
&= \Pr(y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i; \lambda_i)
\end{aligned}$$

In the second and third probabilities, the terms p_i, x_i, $y_{i,t-1}$, and $y_{i,s-1}$ denote the realized values of these variables from the conditioning statement.

⊠

Proof of Theorem 2.3:

(a) The supposition of Lemma 6.1 is satisfied for D_0 and $y_{i,t-1} = d'$ for any $d' \in D_0$. Hence,

$$\begin{aligned}
&\Pr(y_{i,t} \in D_0, y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \\
&= \sum_{d' \in D_0} \Pr(y_{i,t} \in D_0, y_{i,t-1} = d' \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \\
&= \sum_{d' \in D_0} \Pr(y_{i,t} \in D_0 \,|\, y_{i,t-1} = d', p_i, x_i, \lambda_i) \cdot \Pr(y_{i,t-1} = d' \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \\
&\geq \sum_{d' \in D_0} \Pr(y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \cdot \Pr(y_{i,t-1} = d' \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \\
&= [\Pr(y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i, \lambda_i)]^2
\end{aligned}$$

Next, apply Jensen's Inequality to integrate out λ_i.

$$\begin{aligned}
&\Pr(y_{i,t} \in D_0, y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i) \\
&= E\left[\Pr(y_{i,t} \in D_0, y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \,|\, y_{i,t-2}, p_i, x_i\right] \\
&\geq E\left[[\Pr(y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i, \lambda_i)]^2 \,|\, y_{i,t-2}, p_i, x_i\right] \\
&\geq [E\left[\Pr(y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \,|\, y_{i,t-2}, p_i, x_i\right]]^2 \\
&= [\Pr(y_{i,t-1} \in D_0 \,|\, y_{i,t-2}, p_i, x_i)]^2
\end{aligned}$$

(b) $s < t - 1$.

The supposition of Lemma 6.1 is satisfied for D_0 and $y_{i,t-1} = d'$ for any $d' \in D_1$. Hence,

$$\begin{aligned}
&\Pr(y_{i,t} \in D_0, y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \\
&= \sum_{d' \in D_1} \Pr(y_{i,t} \in D_0, y_{i,t-1} = d', y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \\
&= \sum_{d' \in D_1} \Pr(y_{i,t} \in D_0 \,|\, y_{i,t-1} = d', p_i, x_i, \lambda_i) \cdot \Pr(y_{i,t-1} = d', y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \\
&\geq \sum_{d' \in D_1} \Pr(y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \cdot \Pr(y_{i,t-1} = d', y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \\
&= \Pr(y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \cdot \Pr(y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \\
&\geq [\Pr(y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i)]^2
\end{aligned}$$

Next, apply Jensen's Inequality to integrate out λ_i.

$$\begin{aligned}
&\Pr(y_{i,t} \in D_0, y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i) \\
&= E\left[\Pr(y_{i,t} \in D_0, y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \,|\, y_{i,s-1}, p_i, x_i\right] \\
&\geq E\left[\Pr(y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i)^2 \,|\, y_{i,s-1}, p_i, x_i\right] \\
&\geq \left(E\left[\Pr(y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i, \lambda_i) \,|\, y_{i,s-1}, p_i, x_i\right]\right)^2 \\
&= \left[\Pr(y_{i,t-1} \in D_1, y_{i,s} \in D_0 \,|\, y_{i,s-1}, p_i, x_i)\right]^2
\end{aligned}$$

⊠

Derivation of Equation (4.3):

$$\begin{aligned}
A \equiv \Big\{ &(\varepsilon_{i,t}, \varepsilon_{i,s}) \,|\, -p_{d,i,t}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq d\}\delta_0 + \lambda_{d,i} + \varepsilon_{d,i,t} \geq \\
&\qquad \max_{d' \neq d}\left[-p_{d',i,t}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq d'\}\delta_0 + \lambda_{d',i} + \varepsilon_{d',i,t}\right], \\
&-p_{c,i,s}\gamma_0 - \mathbf{1}\{y_{i,s-1} \neq c\}\delta_0 + \lambda_{c,i} + \varepsilon_{c,i,s} \geq \\
&\qquad \max_{c' \neq c}\left[-p_{c',i,s}\gamma_0 - \mathbf{1}\{y_{i,s-1} \neq c'\}\delta_0 + \lambda_{c',i} + \varepsilon_{c',i,s}\right] \Big\} \\
\subset \Big\{ &(\varepsilon_{i,t}, \varepsilon_{i,s}) \,|\, -p_{d,i,t}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq d\}\delta_0 + \lambda_{d,i} + \varepsilon_{d,i,t} \geq \\
&\qquad -p_{c,i,t}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq c\}\delta_0 + \lambda_{c,i} + \varepsilon_{c,i,t}, \\
&-p_{c,i,s}\gamma_0 - \mathbf{1}\{y_{i,s-1} \neq c\}\delta_0 + \lambda_{c,i} + \varepsilon_{c,i,s} \\
&\qquad \geq -p_{d,i,s}\gamma_0 - \mathbf{1}\{y_{i,s-1} \neq d\}\delta_0 + \lambda_{d,i} + \varepsilon_{d,i,s} \Big\} \\
\subset \Big\{ &(\varepsilon_{i,t}, \varepsilon_{i,s}) \,|\, \Delta\Delta\varepsilon^{d,c}_{i,t,s} \geq \Delta\Delta p^{d,c}_{i,t,s}\gamma_0 + \Delta\Delta sw^{d,c}_{i,t,s}(y_{i,t-1}, y_{i,s-1})\delta_0 \Big\},
\end{aligned}$$

which implies

$$\begin{aligned}
&Pr(y_{i,t} = d, y_{i,s} = c | y_{i,s-1}, p_i, \lambda_i) \leq \Pr((\varepsilon_{i,t}, \varepsilon_{i,s}) \in A \,|\, y_{i,s-1}, p_i, \lambda_i) \\
&\leq \Pr(\Delta\Delta\varepsilon^{d,c}_{i,t,s} \geq \Delta\Delta p^{d,c}_{i,t,s}\gamma_0 + \Delta\Delta sw^{d,c}_{i,t,s}(y_{i,t-1}, y_{i,s-1})\delta_0 \,|\, y_{i,s-1}, p_i, \lambda_i)
\end{aligned}$$

Integrate both sides with respect to the conditional distribution of λ_i, and the result follows.

Proof of Theorem 4.2(A): Recall $SU_{d,i,t}(y_{i,t-1}) = -p_{d,i,t}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq d\}\delta_0 + \lambda_{d,i}$.

We consider choice c that satisfies $\min_{d\neq c, d\in\mathcal{D}} \Delta\Delta p^{d,c}_{i,t,t-2} > 0$. We will prove the result via the contrapositive. Suppose for all $d' \neq c$, $\Delta\Delta p^{d',c}_{i,t,t-2}\gamma_0 \geq 2\delta_0$. It follows straightforwardly that for all $d', f \in \mathcal{D}$ with $d' \neq c$,

$$SU_{d',i,t-2}(c) - SU_{c,i,t-2}(c) \geq SU_{d',i,t}(f) - SU_{c,i,t}(f),$$

where the fixed effects cancel from this inequality and so play no role. This inequality, in turn, implies that (for all λ_i),

$$1 \leq \sum_{f\in\mathcal{D}} \Pr(y_{i,t-1} = f \,|\, y_{i,t-3} = c, p_i, x_i, \lambda_i) \frac{\sum_{d'\in\mathcal{D}} \exp(SU_{d',i,t-2}(c) - SU_{c,i,t-2}(c))}{\sum_{d'\in\mathcal{D}} \exp(SU_{d',i,t}(f) - SU_{c,i,t}(f))}.$$

This inequality implies that (for all λ_i),

$$\frac{\exp(SU_{c,i,t-2}(c))}{\sum_{d'\in\mathcal{D}} \exp(SU_{d',i,t-2}(c))} \leq \sum_{f\in\mathcal{D}} \Pr(y_{i,t-1} = f \,|\, y_{i,t-3} = c, p_i, x_i, \lambda_i) \frac{\exp(SU_{c,i,t}(f))}{\sum_{d'\in\mathcal{D}} \exp(SU_{d',i,t}(f))}$$

which is equivalent to for all λ_i,

$$\sum_{f\in\mathcal{D}} \Pr(y_{i,t} = c \,|\, y_{i,t-1} = f, y_{i,t-3} = c, p_i, x_i, \lambda_i) \Pr(y_{i,t-1} = f \,|\, y_{i,t-3} = c, p_i, x_i, \lambda_i)$$
$$\geq \Pr(y_{i,t-2} = c \,|\, y_{i,t-3} = c, p_i, x_i, \lambda_i),$$

or, for all λ_i,

$$\Pr(y_{i,t} = c \,|\, y_{i,t-3} = c, p_i, x_i, \lambda_i) \geq \Pr(y_{i,t-2} = c \,|\, y_{i,t-3} = c, p_i, x_i, \lambda_i),$$

which finally yields

$$\Pr(y_{i,t} = c \,|\, y_{i,t-3} = c, p_i, x_i) \geq \Pr(y_{i,t-2} = c \,|\, y_{i,t-3} = c, p_i, x_i).$$

The result follows as the contrapositive. $\boxtimes$

Lemma 6.2. *Let* $\mathcal{M}_t(c,\lambda_i) = \sum_{r\in\mathcal{D}_t} \exp[-p_{r,t}\gamma_0 - \mathbf{1}\{c \neq r\}\delta_0 + \lambda_{r,i}]$. *For any* c, d,

$$e^{-\delta_0} \;\le\; \frac{\mathcal{M}_t(d,\lambda_i)}{\mathcal{M}_t(c,\lambda_i)} \;\le\; e^{\delta_0}$$

Proof of Lemma 6.2:

For any c, d, r, $-1 - \mathbf{1}\{c \neq r\} \;\le\; -\mathbf{1}\{d \neq r\} \;\le\; 1 - \mathbf{1}\{c \neq r\}$, and so

$$\begin{aligned} e^{-\delta_0}\mathcal{M}_t(c,\lambda_i) &= \sum_{r\in\mathcal{D}_t} e^{(-1-\mathbf{1}\{c\neq r\})\delta_0}\;\exp[-p_{r,t}\gamma_0 + \lambda_{r,i}] \;\le\; \sum_{r\in\mathcal{D}_t} e^{-\mathbf{1}\{d\neq r\}\delta_0}\;\exp[-p_{r,t}\gamma_0 + \lambda_{r,i}] \\ &= \mathcal{M}_t(d,\lambda_i) \;\le\; \sum_{r\in\mathcal{D}_t} e^{(1-\mathbf{1}\{c\neq r\})\delta_0}\;\exp[-p_{r,t}\gamma_0 + \lambda_{r,i}] \\ &= e^{\delta_0}\mathcal{M}_t(c,\lambda_i) \end{aligned}$$

The result follows. ⊠

The following theorem extends the result in Theorem 4.2(B).

Theorem 6.3. *Suppose Assumption 4.1 holds, and* $(d,c) \in \mathcal{D}_t \cap \mathcal{D}_s$. *Then, for* $s \in \{t-1, t-2\}$ *and* $y_{i,s-1} \neq d$,

$$\Pr(y_{i,t} = c, y_{i,s} = d \,|\, y_{i,s-1}, p_i, x_i)\; e^{(\mathbf{1}\{y_{i,s-1}=c\}-1)\delta_0} e^{\Delta\Delta p_{t,s}^{c,d}\gamma_0}$$

$$\le \Pr(y_{i,t} = d, y_{i,s} = c \,|\, y_{i,s-1}, p_i, , x_i)$$

$$\le \Pr(y_{i,t} = c, y_{i,s} = d \,|\, y_{i,s-1} p_i, x_i)\; e^{(\mathbf{1}\{y_{i,s-1}=c\}+1)\delta_0} e^{\Delta\Delta p_{t,s}^{c,d}\gamma_0}$$

Remark 6.4. The case $y_{i,s-1} = d$ is implied by the case $y_{i,s-1} = c$.

Proof of Theorem 6.3:

(i) $s = t - 1$.

$$\Pr(y_{i,t} = d, y_{i,t-1} = c \,|\, y_{i,t-2}, p_i, x_i, \lambda_i)$$

$$\begin{aligned}
&= \Pr(y_{i,t} = d \,|\, y_{i,t-1} = c, p_i, x_i, \lambda_i) \Pr(y_{i,t-1} = c \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \\
&= \frac{e^{-p_{d,i,t}\gamma_0 - \mathbf{1}\{c \neq d\}\delta_0 + \lambda_{d,i}}}{\mathcal{M}_t(c, \lambda_i)} \frac{e^{-p_{c,i,t-1}\gamma_0 - \mathbf{1}\{y_{i,t-2} \neq c\}\delta_0 + \lambda_{c,i}}}{\mathcal{M}_{t-1}(y_{i,t-2}, \lambda_i)} \\
&= \frac{e^{-p_{d,i,t}\gamma_0} e^{-p_{c,i,t-1}\gamma_0 - \mathbf{1}\{y_{i,t-2} \neq c\}\delta_0} e^{-\mathbf{1}\{c \neq d\}\delta_0} e^{\lambda_{d,i} + \lambda_{c,i}}}{\mathcal{M}_t(c, \lambda_i)\mathcal{M}_{t-1}(y_{i,t-2}, \lambda_i)}
\end{aligned}$$

Similarly, for $\Pr(y_{i,t} = c, y_{i,t-1} = d \,|\, y_{i,t-2}, p_i, x_i, \lambda_i)$.

So,

$$\begin{aligned}
\frac{\Pr(y_{i,t} = d, y_{i,t-1} = c \,|\, y_{i,t-2}, p_i, x_i, \lambda_i)}{\Pr(y_{i,t} = c, y_{i,t-1} = d \,|\, y_{i,t-2}, p_i, x_i, \lambda_i)} &= \frac{e^{-p_{d,i,t}\gamma_0} e^{-p_{c,i,t-1}\gamma_0 - \mathbf{1}\{y_{i,t-2} \neq c\}\delta_0} \mathcal{M}_t(d, \lambda_i)}{e^{-p_{c,i,t}\gamma_0} e^{-p_{d,i,t-1}\gamma_0 - \mathbf{1}\{y_{i,t-2} \neq d\}\delta_0} \mathcal{M}_t(c, \lambda_i)} \\
= \exp\left[-\Delta\Delta p^{d,c}_{i,t,t-1}\gamma_0\right] \exp\left[-\left(\mathbf{1}\{y_{i,t-2} \neq c\} - \mathbf{1}\{y_{i,t-2} \neq d\}\right)\delta_0\right] \cdot \frac{\mathcal{M}_t(d, \lambda_i)}{\mathcal{M}_t(c, \lambda_i)}
\end{aligned}$$

By Lemma 6.2,

$$\begin{aligned}
&\Pr(y_{i,t} = c, y_{i,t-1} = d \,|\, y_{i,t-2}, p_i, x_i, \lambda_i)\; e^{-\delta_0} e^{-\Delta\Delta p^{d,c}_{i,t,t-1}\gamma_0} e^{-(\mathbf{1}\{y_{i,t-2} \neq c\} - \mathbf{1}\{y_{i,t-2} \neq d\})\delta_0} \\
&\quad \leq \Pr(y_{i,t} = d, y_{i,t-1} = c \,|\, y_{i,t-2}, p_i, x_i, \lambda_i) \\
&\quad \leq \Pr(y_{i,t} = c, y_{i,t-1} = d \,|\, y_{i,t-2}, p_i, x_i, \lambda_i)\; e^{\delta_0} e^{-\Delta\Delta p^{d,c}_{i,t,t-1}\gamma_0} e^{-(\mathbf{1}\{y_{i,t-2} \neq c\} - \mathbf{1}\{y_{i,t-2} \neq d\})\delta_0}
\end{aligned}$$

The result for the case $s = t - 1$ follows by integrating out λ_i.

(ii) $s = t-2$.

$$\begin{aligned}
&\Pr(y_{i,t}=d, y_{i,t-2}=c \,|\, y_{i,t-3}, p_i, x_i, \lambda_i) \\
&= \sum_{r\in\mathcal{D}_{t-1}} \Pr(y_{i,t}=d \,|\, y_{i,t-1}=r, p_i, x_i, \lambda_i) \Pr(y_{i,t-1}=r \,|\, y_{i,t-2}=c, p_i, x_i, \lambda_i) \\
&\qquad \cdot \Pr(y_{i,t-2}=c \,|\, y_{i,t-3}, p_i, x_i, \lambda_i) \\
&= \left[\sum_{r\in\mathcal{D}_{t-1}} \frac{e^{-(\mathbf{1}\{r\neq d\}+\mathbf{1}\{c\neq r\})\delta_0} e^{-p_{r,i,t-1}\gamma_0+\lambda_{r,i}}}{\mathcal{M}_t(r,\lambda_i)} \right] \cdot \frac{e^{-p_{d,i,t}\gamma_0} e^{-p_{c,i,t-2}\gamma_0 - \mathbf{1}\{y_{i,t-3}\neq c\}\delta_0} e^{\lambda_{d,i}+\lambda_{c,i}}}{\mathcal{M}_{t-1}(c,\lambda_i)\mathcal{M}_{t-2}(y_{i,t-3},\lambda_i)}
\end{aligned}$$

Similarly, for $\Pr(y_{i,t}=c, y_{i,t-2}=d \,|\, p_i, y_{i,t-3}, x_i = x, \lambda_i)$.

Hence,

$$\begin{aligned}
\frac{\Pr(y_{i,t}=d, y_{i,t-2}=c \,|\, y_{i,t-3}, p_i, x_i, \lambda_i)}{\Pr(y_{i,t}=c, y_{i,t-2}=d \,|\, y_{i,t-3}, p_i, x_i, \lambda_i)} &= \frac{e^{-(p_{d,i,t}+p_{c,i,t-2})\gamma_0} e^{-\mathbf{1}\{c\neq y_{i,t-3}\}\delta_0} \mathcal{M}_{t-1}(d,\lambda_i)}{e^{-(p_{c,i,t}+p_{d,i,t-2})\delta_0} e^{-\mathbf{1}\{d\neq y_{i,t-3}\}\delta_0} \mathcal{M}_{t-1}(c,\lambda_i)} \\
= \exp\left[-\Delta\Delta p^{d,c}_{i,t,t-2}\gamma_0\right] \exp\left[-\left(\mathbf{1}\{y_{i,t-3}\neq c\} - \mathbf{1}\{y_{i,t-3}\neq d\}\right)\delta_0\right] &\cdot \frac{\mathcal{M}_{t-1}(d,\lambda_i)}{\mathcal{M}_{t-1}(c,\lambda_i)}
\end{aligned}$$

As in the $s = t-1$ case, the result for $s = t-2$ now follows by application of Lemma 6.2(a) and integrating out λ_i.

⊠