Unobserved Heterogeneity, State Dependence, and Health Plan Choices.

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

September 2, 2022

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 a policy-relevant environment (a subsidized health insurance exchange). Moment inequalities are used to infer 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 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

Since Heckman (1978, 1981), distinguishing the impacts of unobserved heterogeneity from those of state dependence has been a central issue in empirical work in economics, as the distinction has implications for the interpretation and policy implications of many observed phenomena. 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 et al. (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 et al., 2003).

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 the issue of understanding persistence in health insurance plan choices, an issue which has led to considerable policy debate. Our method is nonparametric, but we also develop a parametric analogue. We then compare the empirical results from the nonparametric model to both the parametric model that allows for individual-by-product fixed effects, and to a more familiar set of parametric models that do not. In both models that allow for fixed effects we find an upper bound to switching costs that is considerably lower than the estimates found in the literature. A counterfactual indicates that the difference is likely to have substantial implications for the analysis of price effects.

The question of whether state dependence or unobserved heterogeneity underlies low price responsiveness in health insurance choice is economically important and policy-relevant. Governments set rules for market-based health insurance programs in the Affordable Care Act exchanges, Medicare Part D, and Medicaid managed care that cover more than 75 million people and cost over \$700 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).[1] 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.[2]

We offer a new way of distinguishing between state dependence and heterogeneity. Prior work has focused on one of two general approaches. The first, followed by most applied work, estimates a fully parametric utility model that includes a cost of switching from an individual's lagged choice (their "state"). As Heckman (1981) emphasized, this approach relies on finding "initial conditions" in which individuals 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 utility.

The other approach, following the seminal work of Honoré and Kyriazidou (2000), does not require initial conditions but instead focuses on subsets of the data where the role of (very flexible) unobserved preferences, captured by *individual-by-product fixed effects*, can be differenced out. This provides an attractive nonparametric alternative, but in practice, it 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.

Our key contribution is to show how to use partial identification in a model with flexible fixed effects to identify state dependence from patterns of switching choices in response to *changes in prices* (or other product characteristics). In doing so, we develop new economic

[1] A related literature studies the *mechanisms* behind state dependence in health insurance choices, distinguishing factors like search costs, rational inattention, and true switching hassles (Heiss et al., 2021; Abaluck and Adams-Prassl, 2021; Brown and Jeon, 2020; Brot-Goldberg et al., 2021).

[2] 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 et al., 2009, 2010; Bronnenberg et al., 2012), residential electricity markets (Hortaçsu et al., 2017), auto insurance (Honka, 2014), and paid television services (Shcherbakov, 2016).

intuition about how certain patterns of switching choices around price changes can distinguish state dependence from heterogeneity. We show how to use this intuition to guide the selection of moment inequalities based on economic theory, addressing the practical problem of "too many inequalities" that often poses a challenge for moment inequality methods.

Our econometric model is a semiparametric dynamic discrete choice model that allows for flexible unobserved preferences via individual-by-product fixed effects. Here we present the simple case where price is the only observed product characteristic coefficient of interest—relegating the impact of other observables to the fixed effects—but the method can be extended to allow the targeted coefficients to be a vector. Consumer i at time t faces a choice set $\mathcal{D}_t$ where $\#\mathcal{D}_t = D_t$. Given price $p_{d,i,t}$ for choice $d \in \mathcal{D}_t$ and last period's choice $y_{i,t-1} \in \mathcal{D}_{t-1}$, the utility associated with choice d at time t is

$$U_{d,i,t} = \Big(- p_{d,i,t} - \kappa_0 \cdot \mathbf{1}\{y_{i,t-1} \neq d\} \Big) \beta_i + \lambda_{d,i} + \varepsilon_{d,i,t} \tag{1.1}$$

with $\beta_i > 0$. The choice in period t is $y_{i,t} = \max_{d \in \mathcal{D}_t} U_{d,i,t}$.

Here κ_0 represents the price-equivalent cost of switching, β_i allows the importance of price to vary by individual, $\lambda_{d,i}$ denotes individual (additive) product preferences, and $\varepsilon_{d,i,t}$ captures the remaining unobserved variation in random utility. This specification lets us estimate the importance of switching costs relative to price, allowing both for individual-specific price coefficients and a flexible tradeoff between these variables and other additively separable preferences. No restrictions are placed on the joint distribution of the fixed effects and the prices or any other observed individual characteristics, including the initial conditions. Different assumptions on the distribution of $\{\varepsilon_{d,i,t}\}_{d \in \mathcal{D}_t, 1 \leq t \leq T}$ are explained and explored.

The parameter κ_0 is our main empirical focus, as it captures the importance state dependence. In general, κ_0 is only partially identified (Honoré and Tamer, 2006). We construct moment inequalities that are true regardless of unobserved preferences $(\lambda_{d,i}, \beta_i)$ that identify bounds on κ_0. These moment inequalities are a direct result of revealed preference, and the bounds they generate require only relatively weak restrictions on $\varepsilon_{d,i,t}$.

The key intuition, formalized in the paper, can be seen in the following example. Suppose that we observe a product d whose relative price falls from $t-2$ to $t-1$ and then rises again in t.[3] If, for consumers in a given initial state y_{t-2} ($\neq d$), we observe that their probability of switching to d when its price improves in $t-1$ exceeds their probability of staying with d when its price worsens in t, then we can infer that switching costs must not be too large. In other words, we infer an *upper bound* on κ_0 from cases where price changes lead to a sufficient amount of switching towards plans whose price falls. This bound holds regardless of unobserved preferences, our $(\lambda_{d,i}, \beta_i)$, from the fact that these tastes are the same when the two choices are made at times $t-1$ and t. The tightness of the upper bound depends on how little prices need to change to generate a large degree of switching to and from d. Similarly, we can infer a *lower bound* on switching costs from cases where consumers stay with a product when its attributes get worse (e.g., its price rises) more frequently than they switch to it when its attributes improve.

Given the model in equation (1.1), the bounds on κ_0 can be used to either explore the factors associated with the $(\{\lambda_{d,i}\}_{d,i}, \{\beta_i\}_i)$ and/or, as we do below, to assess the impact of counterfactual prices when we control for the fixed effects. Alternatively, and equally important, we provide a simple test for the presence of state dependence and a way to control for its impact on the coefficients of observed covariates. A more detailed model may be needed to uncover the mechanism that generated it.[4]

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

[3] More generally there could be more than one observed characteristic of interest whose value changes over time for a given individual, and/or for which the form of its interactions with either β_i or λ_i can be specified a priori, in which case the target parameter would be a vector. In our empirical work we condition on cells with common observed characteristics, so using a single target parameter seems appropriate (and also simplifies the exposition).

[4] 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. One could analyze the distinction between the identified set defined by our moment inequalities and the sharp set in special cases where the sharp set is known as in the dynamic binary response model with discrete covariates (Khan et al., 2020). However, as we demonstrate below, empirical implementation would have to face the challenge of maintaining power with an exceptionally large number of slack inequalities. Also our focus is on κ_0, rather than on the quantiles or averages of utilities, which is the focus of Chernozhukov et al. (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.

to that problem. The literature on discrete choice with fixed effects provides an analogue of "within" estimation in panel data models with continuous dependent variables where the between/within distinction has been a focus of empirical analysis. 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 (2016), 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 et al. (2019) 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 (2020) considers a binary logit model with state dependence that does not require matching (or situations with constant product characteristics over time).[5] Honoré and Tamer (2006) examine identified sets from a related model, and Khan et al. (forthcoming) 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 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

[5]Honoré and Weidner (2020) 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.

(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 semiparametric and parametric moment inequality approaches. We find lower bounds on switching costs (κ_0) ranging from \$20 (from the semiparametric) to \$32 (from the parametric analysis) per month and upper bounds of \$56-57 per month. The closeness of the parametric upper bound (\$56) with that found from the semiparametric analysis (\$57) adds confidence in our estimates of that parameter. These switching costs are meaningful relative to average (subsidized) consumer premiums in the market, which vary from \$48 to \$62 per month 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 \$78 to \$114 per month — values that are 37-100 percent above the *upper bound* from our fixed effects method. These higher estimates are consistent with prior work on the CommCare data (see Shepard (2022), who finds $\kappa_0 \approx \$100$) and with similarly high estimates in other health insurance settings (e.g., Handel, 2013; Polyakova, 2016). 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,[6] 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 premium from \$58.4 to \$91.5. After experiencing sharp losses in market share, it reduced its premium to \$41.5 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. Before going to the parametric revealed preference empirical results, we present the results from the parametric comparison models that do not allow for fixed effects. These are then compared to the revealed preference bounds that allow for fixed effects and the implications of the differences between them are explored in the counterfactual analysis. We conclude with a brief summary. All proofs are provided in Appendix C.

Notation. Let $\epsilon_{i,t} \equiv [\epsilon_{1,i,t}, \ldots \epsilon_{\mathcal{D},i,t}]$, $\epsilon_i \equiv [\epsilon_{i,1}, \ldots \epsilon_{i,T}]$, $\lambda_i \equiv [\lambda_{1,i}, \ldots, \lambda_{\mathcal{D},i}]$, $p_{i,t} \equiv [p_{1,i,t}, \ldots p_{\mathcal{D},i,t}]$, and $p_i \equiv [p_{i,1}, \ldots, p_{i,T}]$. While p_i denotes price in our application, it could include any time-

[6] 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.

varying observed covariates more generally.

2 Price Changes that Induce Switching.

Our approach uses revealed preference to relate switching behavior directly to the price changes that induced it. Let $c, d \in \mathcal{D}_t \cap \mathcal{D}_s$ denote two distinct choices. If an agent chose d instead of c at time t, then $U_{d,i,t} - U_{c,i,t} \geq 0$. Likewise, if they chose c over d in period s, then $U_{c,i,s} - U_{d,i,s} \geq 0$. Summing these inequalities generates an expression known to be non-negative that does not depend on λ_i, and dividing by $\beta_i > 0$ does not change the sign of the inequality. From this starting point, we consider different methods of generating either an upper or lower bound on κ_0 — depending on the observed pattern of switching — that is a function of (observed) price changes and (unobserved) utility error changes. The different methods we employ are generated by first nonparametric, and then parametric, assumptions on the error distribution.

Formally, assume agent i chooses c at time s and d at time t, where $s < t$ and $\{d, c\} \in \mathcal{D}_t \cap \mathcal{D}_s$. Then the model in equation (1.1) implies $U_{d,i,t} - U_{c,i,t} \geq 0$, or

$$\Big(- [p_{d,i,t} - p_{c,i,t}] - [\mathbf{1}\{y_{i,t-1} \neq d\} - \mathbf{1}\{y_{i,t-1} \neq c\}]\kappa \Big)\beta_i + [\lambda_{d,i} - \lambda_{c,i}] + [\epsilon_{d,i,t} - \epsilon_{c,i,t}] \geq 0$$

and since $U_{c,i,s} - U_{d,i,s} \geq 0$,

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

Since adding these two inequalities together cancels the $\lambda_{d,i} - \lambda_{c,i}$ terms, it generates an inequality that does not depend on these unobserved preferences. If we then divide by $\beta_i > 0$, and define $\Delta\Delta x^{d,c}_{i,t,s} \equiv [(x_{d,i,t} - x_{c,i,t}) - (x_{d,i,s} - x_{c,i,s})]$ for any variable x, we obtain

$$-\Delta\Delta p^{d,c}_{i,t,s} - sw^{d,c}_{i,t,s} \cdot \kappa + \Delta\Delta\epsilon^{d,c}_{i,t,s} \geq 0, \quad \text{where} \tag{2.1}$$

$$sw^{d,c}_{i,t,s} = sw^{d,c}(y_{i,t-1}, y_{i,s-1}) \equiv (\mathbf{1}\{y_{i,t-1} \neq d\} - \mathbf{1}\{y_{i,t-1} \neq c\}) - (\mathbf{1}\{y_{i,s-1} \neq d\} - \mathbf{1}\{y_{i,s-1} \neq c\})$$

and we have abused notation slightly by not distinguishing $\Delta\Delta\epsilon^{d,c}_{i,t,s}$ from $\beta_i^{-1}\Delta\Delta\epsilon^{d,c}_{i,t,s}$. This because differences in β_i do not impact our nonparametric identification results, so for simplicity we take $\beta_i \equiv 1$ in the nonparametric section. However, when we move to parametric models we will reintroduce β_i, as the differences in β_i do impact the parametric results.

Inequality (2.1) follows directly from Samuelson's revealed preference inequalities once we allow for fixed effects, state dependence and disturbances (Samuelson, 1938). It shows that the extent of switching induced by price changes depends on κ_0 and the distribution of the disturbances in the utility function (i.e. $\Delta\Delta\epsilon_{i,t,s}^{d,c}$). Across all our methods, we assume that

Assumption 2.1.

$$\epsilon_{i,t} \,|\, p_i, y_{i,t-1}, y_{i,t-2}, \ldots, y_{i,0}, \beta_i, \lambda_i \;\sim\; \epsilon_{i,t} \,|\, \beta_i, \lambda_i \quad \square$$

which implies that all serial dependence in choices not associated with prices or the fixed effects is modeled through the state dependence parameters.

We then consider three assumptions on the contemporaneous distribution of $\epsilon_{i,t}$. In the first two we allow that distribution to be nonparametric. First we consider the direct (or "positive") implication of equation (2.1) when all that is assumed is that the coniditional median of $\Delta\Delta\epsilon_{i,t,s}^{d,c} = 0$. This allows the distribution of $\epsilon_{i,t,}$ to differ over time. Next we make the stationarity assumption that $\epsilon_{i,t}|\lambda_i, \beta_i \sim \epsilon_{i,1}|\lambda_i, \beta_i$. This does not require the median zero assumption but does assume that the conditional distribution of $\epsilon_{i,t}$ does not change over time. Formally, this approach does not difference out fixed effects, but it achieves the same goal by considering only the "within" variation in choices across time as prices and lagged choices vary while fixed effects and the distribution of disturbances stay constant. We show that it can be informative about κ when the assumption that the conditional median of $\Delta\Delta\epsilon_{i,t,s}^{d,c} = 0$ is not.

Finally we consider the case where $\epsilon_{i,t}$ has a parametric distribution, that is $\epsilon_{i,t} \sim F(\cdot|\theta)$. For this case we begin with inequalities that are available regardless of the precise form of that distribution, and then add inequalities that require $F(\cdot|\theta)$ to be a logistic distribution.

3 Nonparametric Approaches

In this section we present two nonparametric approaches to inferring bounds on switching costs and use them to explore the extent of state dependence in the Commcare data. The assumptions underlying inequality (2.1) are used throughout. This inequality implies that if agent i chooses d at time t and c at time s, where $s < t,$ and $d, c \in \mathcal{D}_t \cap \mathcal{D}_s$, then $\Delta\Delta\epsilon^{d,c}_{i,t,s} \geq \Delta\Delta p^{d,c}_{i,t,s} + \kappa \cdot sw^{d,c}_{i,t,s}$. Each nonparametric approach leads to an identification finding that the probability of choices c and d at times s and t can be informative about κ_0.

3.1 Direct Implications of Revealed Preference.

The revealed preference inequality (2.1), together with serial independence (Assumption 2.1), imply that if agent i chooses d at time t and c at time s, where $s < t,$ and $d, c \in \mathcal{D}_t \cap \mathcal{D}_s$, then $\Delta\Delta\epsilon^{d,c}_{i,t,s} \geq \Delta\Delta p^{d,c}_{i,t,s} + \kappa \cdot sw^{d,c}_{i,t,s}$. Denoting the conditional distribution of $\Delta\Delta\epsilon^{d,c}_{i,t,s}$ (given $p_i, y_{i,s-1}$) by $\mathcal{F}^{d,c}_{t,s}(\cdot)$, then

$$\Pr(y_{i,t} = d, y_{i,s} = c \,|\, p_i, y_{i,s-1}) \leq \; 1 - \mathcal{F}^{d,c}_{t,s}(\Delta\Delta p^{d,c}_{i,t,s} + \kappa_0 \cdot sw^{d,c}_{i,t,s}). \tag{3.1}$$

An estimate of the left hand side of this inequality can be obtained from the data and does not depend on κ. The right hand side does depend on κ_0, but how that can be used to reject values of $\kappa \neq \kappa_0$ depends on what is known about the conditional distribution $\mathcal{F}^{d,c}_{t,s}$. A nonparametric approach adopts the following conditional median assumption[7]

Assumption 3.1.

$$median[\Delta\Delta\epsilon^{d,c}_{i,t,s}|p_i, y_{i,s-1}] = 0$$

First consider the implications of $Pr(y_{i,t} = d, y_{i,s} = c|p_i, y_{i,s-1}) < 0.5$. Notice that $\Delta\Delta p^{d,c}_{i,t,s} + \kappa \cdot sw^{d,c}_{i,t,s} < 0$ implies $1 - \mathcal{F}^{d,c}_{t,s}(\Delta\Delta p^{d,c}_{i,t,s} + \kappa_0 sw^{d,c}_{i,t,s}) > 0.5$ which automatically satisfies equation (3.1). If $\Delta\Delta p^{d,c}_{i,t,s} + \kappa_0 sw^{d,c}_{i,t,s} = x > 0$ equation (3.1) will imply that

[7] The median zero assumption does not require stationarity. It suffices to assume $\varepsilon_{d,i,t} - \varepsilon_{c,i,t}$ and $\varepsilon_{d,i,s} - \varepsilon_{c,i,s}$ are symmetrically distributed about zero. For example, the assumption that $\varepsilon_{d,i,t} - \varepsilon_{c,i,t}$ and $\varepsilon_{d,i,s} - \varepsilon_{c,i,s}$ are symmetrically distributed about zero would follow from an assumption of exchangeability of the disturbances across choices (Manski, 1975; Fox, 2007; Yan, 2013)

$\mathcal{F}^{d,c}_{t,s}(x) > (1 - Pr(y_{i,t} = d, y_{i,s} = c|p_i, y_{i,s-1}))$, but since the only restriction on $\mathcal{F}^{d,c}_{t,s}(x)$ is that it be larger than a half, we can always find an $\mathcal{F}^{d,c}_{t,s}(\cdot)$ that satisfies this condition.

So if $Pr(y_{i,t} = d, y_{i,s} = c|p_i, y_{i,s-1}) < 0.5$, which is always the case for the estimated probabilities in our data, Assumption 3.1 cannot rule out any values of κ. If $Pr(y_{i,t} = d, y_{i,s} = c|p_i, y_{i,s-1}) > 0.5$, Assumption 3.1 will imply that $\Delta\Delta p^{d,c}_{i,t,s} + \kappa_0 sw^{d,c}_{i,t,s} < 0$ producing an upper (lower) bound when $sw^{d,c}_{i,t,s} > 0$ (< 0). Though this case is not relevant for our data, it may be for others, so we provide a more formal treatment of it in the Appendix.

3.2 Contrapositive Implications of Revealed Preference.

This subsection adds Assumption 3.2 to Assumption 2.1.

Assumption 3.2. *For any t, the disturbance $\epsilon_{i,t}$ is conditionally stationary over time, i.e.*

$$\epsilon_{i,t} \,|\, \lambda_i \sim \; \epsilon_{i,1} \,|\, \lambda_i. \quad \square$$

Assumption 3.2 is common in dynamic panel settings. It includes strict exogeneity of the time-varying covariates p_i while placing no restrictions on the correlation between λ_i and p_i. Nor does it impose any restriction on the distribution of $\epsilon_{d,i,t}$ across choices d, so $\epsilon_{d,i,t}$ can 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 i, though the identification results could be re-written to allow for non-identical distributions.

Differences in the structural part of utility. Consider any two periods t and s, with $t > s$, and define the structural part of the utility for choice d in period t as

$$SU_{d,i,t}(y_{i,t-1}, p_i; \lambda_i, \kappa) \equiv -p_{i,d,t} - \kappa \cdot \mathbf{1}\{y_{i,t-1} \neq d\} + \lambda_{d,i} = U_{d,i,t} - \epsilon_{d,i,t}. \tag{3.2}$$

For a fixed κ, order choices by the *difference* $SU_{d,i,t}(y_{i,t-1}, p_i; \lambda_i, \kappa) - SU_{d,i,s}(y_{i,s-1}, p_i; \lambda_i, \kappa)$. Comparing preferences for a given choice in different periods differences out the $\{\lambda_i\}$.[8] So if

[8]Were we to reinstate the β_i the difference would also produce an ordering which does not depend on those parameters, which is why we ignore the β_i in the non-parametric analysis.

$d_1(\cdot)$ is the choice with the largest structural utility difference between the periods

$$d_1(y_{i,t-1}, y_{i,s-1}, p_{i,t}, p_{i,s}; \kappa) \equiv \max_{d \in \mathcal{D}} \left[SU_{d,i,t}(y_{i,t-1}, p_i; \lambda_i, \kappa) - SU_{d,i,s}(y_{i,s-1}, p_i; \lambda_i, \kappa), \right] \quad (3.3)$$

while for $j = 2, \ldots D$, the choice with the j^{th} largest difference is

$$d_j(y_{i,t-1}, y_{i,s-1}, p_{i,t}, p_{i,s}; \kappa) \equiv \max_{d \notin \{d_1, \ldots d_{j-1}\}} \left[SU_{d,i,t}(y_{i,t-1}, p_i; \lambda_i, \kappa) - SU_{d,i,s}(y_{i,s-1}, p_i; \lambda_i, \kappa) \right]. \quad (3.4)$$

In words, $d_1(\cdot)$ is the choice whose structural component of random utility improves most between periods s and t (conditional on lagged choices), $d_2(\cdot)$ improves the next most, and so on. Note that for any given value of κ, the ordering $(d_1(\cdot), d_2(\cdot), \ldots)$ depends only on the differences in prices and switching costs over the two periods.

Since $\epsilon_{i,t}$ and $\epsilon_{i,s}$ have identical distributions the relative magnitude of the conditional probabilities for $y_{i,t} = d_1$ and $y_{i,s} = d_1$ depends only on the difference in the structural part of their utilities. So when $\kappa = \kappa_0$ equation (3.3) ensures that the conditional probability of observing $d_1(\cdot; \kappa_0)$ in period t is greater than in period s. More generally, we have[9]

Lemma 3.3. *Suppose Assumption 3.2 holds. Assume $t > s$, and $D_0 \subset \mathcal{D}$. If*

$$\min_{d \in D_0} \left[SU_{d,i,t}(y_{i,t-1}, p_i; \lambda_i, \kappa_0) - SU_{d,i,s}(y_{i,s-1}, p_i; \lambda_i, \kappa_0) \right]$$
$$\geq \max_{c \notin D_0} \left[SU_{c,i,t}(y_{i,t-1}, p_i; \lambda_i, \kappa_0) - SU_{c,i,s}(y_{i,s-1}, p_i; \lambda_i, \kappa_0) \right],$$

then

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

Let d_j^0 denote $d_j(y_{i,t-1}, y_{i,s-1}, p_{i,t}, p_{i,s}; \kappa_0)$. If the d_j^0 are distinct the choice sets (or the D_0) that satisfy the supposition of this lemma are $D_0 = \{d_1^0\}$, $\{d_1^0, d_2^0\}$, …, $\{d_1^0, \ldots, d_{D-1}^0\}$.

The lemma states that if the structural utility for individual i (including any switching costs) for all $d \in D_0$ improves by at least as much as all options $c \notin D_0$ between periods s and t, then individual i will be more likely to choose an option $d \in D_0$ at time t than at time s *regardless* of its value for λ_i.

[9] An analogous finding in the static discrete panel model is found in Pakes and Porter (2016).

Allowing for state dependence. The conditional probabilities in the conclusion of Lemma 3.3 cannot be estimated directly due to the presence of the unobserved values of λ_i in the conditioning set. Without state dependence, both sides of the inequality could be integrated with respect to the conditional distribution of λ_i given prices to yield a conditional probability inequality based only on observables. However, with state dependence, this simple way of integrating out λ_i from the inequality is not applicable, since the conditional distribution of λ_i given prices and $y_{i,t-1}$ cannot be assumed the same as conditional distribution of λ_i given prices and $y_{i,s-1}$. So to identify κ_0 we need more than the Lemma.

To see through how we proceed consider the special case where $s = t-1$, $D_0 = d_1^0$ (a single choice), and $y_{i,t-1} = d_1^0$. Applying Assumption 2.1 and Lemma 3.3 we obtain

$$\begin{aligned}\Pr(y_{i,t} = d_1^0 | p_i, y_{i,t-1} = d_1^0, y_{i,t-2}, \lambda_i) &= \Pr(y_{i,t} = d_1^0 | p_i, y_{i,t-1} = d_1^0, \lambda_i) \\ &\geq \Pr(y_{i,t-1} = d_1^0 | p_i, y_{i,t-2}, \lambda_i).\end{aligned}$$

As noted above, we cannot use this inequality directly because we do not observe λ_i. So we multiply both sides by $\Pr(y_{i,t-1} = d_1^0 | p_i, y_{i,t-2}, \lambda_i)$ and take expectations over λ_i (given prices and $y_{i,t-2}$). Then,

$$\begin{aligned}\Pr(y_{i,t} = d_1^0, y_{i,t-1} = d_1^0 | p_i, y_{i,t-2}) &\equiv E_{\lambda_i}[\Pr(y_{i,t} = d_1^0, y_{i,t-1} = d_1^0 | p_i, y_{i,t-2}, \lambda_i) | p_i, y_{i,t-2}] \\ &\geq E_{\lambda_i}[\left(\Pr(y_{i,t-1} = d_1^0 | p_i, y_{i,t-2}, \lambda_i)\right)^2 | p_i, y_{i,t-2}] \\ &\geq \left[E_{\lambda_i}\left(\Pr(y_{i,t-1} = d_1^0 | p_i, y_{i,t-2}, \lambda_i)\right)\right]^2 \\ &\equiv \left(\Pr(y_{i,t-1} = d_1^0 | p_i, y_{i,t-2})\right)^2, \end{aligned} \tag{3.5}$$

where the first inequality follows from the lemma, and the second from Jensen's inequality. Finally divide both sides of (3.5) by $Pr(y_{i,t-1} = d_1^0 | p_i, y_{i,t-2})$ to obtain

$$\Pr(y_{i,t} = d_1^0 \,|\, y_{i,t-1} = d_1^0, p_i, y_{i,t-2}) \geq \Pr(y_{i,t-1} = d_1^0 \,|\, p_i, y_{i,t-2}). \tag{3.6}$$

Our assumptions imply inequality (3.6) holds when $\kappa = \kappa_0$. We can calculate the difference $SU_{d^1,i,t} - SU_{d^1,i,t-1}$ for any κ value. If (3.6) is violated when replacing d_1^0 by $d_1(\tilde{\kappa})$ then

$\tilde{\kappa} \neq \kappa_0$. Moreover, if $y_{i,t-2} \neq d_1(\tilde{\kappa})$, then $\tilde{\kappa} \geq \kappa_0$. I.e. if a significant number of people switch out of d_1, κ_0 cannot be too large.

Note that the slackness in (3.6) is due to the fact that it ignores the variance in $\Pr(y_{i,t-1} = d_1^0 | p_i, y_{i,t-2}, \lambda_i)$ conditional only on $(p_i, y_{i,t-2})$. This conditional variance, in turn, depends on the variance of the λ_i. The λ_i are explained by both the observable and the unobservable determinants of utility, and the richer the set of observable characteristics that the analyst can condition on, the lower the conditional variance of the λ_i in the data, and the more powerful this inequality. This motivates our decision to form moments from cells with common observable characteristics and $y_{i,t-2}$ in the empirical analysis which follows.

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_0^1 with D_0 in equation (3.6), though then the inequality is required to hold for each $y_{i,t-1} \in D_0$. Second, when $s < t-1$, $y_{i,t-1}$ only enters the inequality in the lemma through the conditional probability for $y_{i,t}$. This implies that $y_{i,t-1}$ can be allowed to take values different than $y_{i,t}$ provided the inequality holds for those values of $y_{i,t-1}$. When $s < t-1$ the next condition defines the set of values $y_{i,t-1}$ can take as D_1.

Condition 3.4. *Given $t > s$ and choice sets $D_0, D_1 \subset \mathcal{D}$, for all $d' \in D_1$,*

$$\min_{d \in D_0} [SU_{d,i,t}(d', p_i; \lambda_i, \kappa_0) - SU_{d,i,s}(y_{i,s-1}, p_i; \lambda_i, \kappa_0)] \;\geq\; \max_{c \notin D_0} [SU_{c,i,t}(d', p_i; \lambda_i, \kappa_0) - SU_{c,i,s}(y_{i,s-1}, p_i; \lambda_i, \kappa_0)].$$

This condition ensures that structural utility differences for the choices in D_0 are larger at time t for any value of the lagged dependent variable in D_1 than at time s with a lagged dependent value of $y_{i,s-1}$.

Theorem 3.5. *Suppose Assumption 3.2 holds.*

(a) For $s = t-1$, for any choice set $D_0 = D_1$ satisfying Condition 3.4,

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

(b) For $s < t-1$, for any choice sets D_0 and D_1 satisfying Condition 3.4,

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

Note that with $s = t-1$ the amount of inequalities Theorem 3.5(a) generates will depend on κ_0, with a minimum of $D-1$. If, as in Theorem 3.5(b), $s < t-1$ the fact that Condition 3.4 allows the set D_1 of time t lagged values to be distinct from the set D_0, it leads to potentially even more inequalities, though again the actual number will depend on κ.

Selecting Moments. The fact that Theorem 3.5 generates many inequalities raises the question of which inequalities to use. We provide more detail on the importance of, and the procedure for, selecting moments in the next subsection where we can illuminate the arguments by reference to our data. We conclude this section with an explanation of a moment selection procedure for the simple case illustrated in inequality (3.6).

As in equation (2.1), the inequalities in Theorem 3.5 compare changes in switching costs to changes in price. The economics inducing the consumer to make its choices underlies our approach to choosing among inequalities. To reject the inequality in (3.6), and hence rule out particular values of κ, we require the contrapositive

$$\Pr(y_{i,t} = d_1^0 \,|\, y_{i,t-1} = d_1^0, p_i, y_{i,t-2}) < \Pr(y_{i,t-1} = d_1^0 \,|\, p_i, y_{i,t-2}).$$

When might this occur? Recall that the distribution of $\epsilon_{i,t}$ in the two periods is the same, so the answer lies in the economics underlying $\{SU_{d,i,t}(y_{i,t-1}, p_i; \lambda_i, \kappa) - SU_{d,i,t-1}(y_{i,t-2}, p_i; \lambda_i, \kappa)\}_d$. Moreover, we can condition on the initial choice and prices. So consider those who chose c in $t-2$ and say there was a sharp increase $p_{c,t-1} - p_{d_1^0,t-1}$. This should induce a fraction of them to choose d_1^0 in period $t-1$. Then, in period t, the relative price of d_1^0 fell. This should induce a significant fraction of them to shift back. The structural utility difference for choosing d_1^0 in period $t-1$ instead of c is $-p_{d_1^0,t-1} - \kappa_0 + p_{c,t-1} + \lambda_{i,d_1^0} - \lambda_{i,c}$, while the structural utility difference for chosing c in period t instead of d is $-p_{c,t} - \kappa_0 + p_{d_1^0,t} + \lambda_{i,c} - \lambda_{i,d_1^0}$. If we add the inequalities derived from these differences together and ignore the error terms, we get the upper bound $\Delta\Delta p_{i,t,t-1}^{d_1^0,c} > 2\kappa_0$.

More formally, with two choices $\mathcal{D}_t = \mathcal{D}_{t-1} = \{c, d\}$, any κ value that satisfies $2\kappa \geq (p_{d,t} - p_{d,t-1}) - (p_{c,t} - p_{c,t-1})$ implies $\Delta\Delta SU^{d,c}_{i,t,t-1} \geq 0$. The theorem then implies

$$Pr(y_{i,t} = d | y_{i,t-1} = d, p) \geq Pr(y_{i,t-1} = d | y_{i,t-2} = c, p).$$

But if, in the data indicate that $Pr(y_{i,t} = d|y_{i,t-1} = d, p) < Pr(y_{i,t-1} = d|y_{i,t-2} = c, p)$; i.e. some who went from d to c in $t-1$ switch back in t, then we know $2\kappa_0 < \Delta p_{d,t,t-1} - \Delta p_{c,t,t-1}$ (our "contrapositive").

3.3 Empirical Analysis: Massachusetts Health Insurance

We analyze health insurance plan choices made by enrollees in the Commonwealth Care ("CommCare") 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.[10] For more detail on the data and the CommCare program see Shepard (2022); Finkelstein, Hendren, and Shepard (2019); McIntyre, Shepard, and Wagner (2021).

Forming Inequalities. We want to capture switching costs that are not induced by changes in the individual's choice environment, just by prices, and this requires the choice set to be the same in the two periods we compare. We therefore remove comparisons for

[10]In a few 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.

individuals who changed regions (there are five in the data), or who faced different plan offerings in the comparison periods, and use separate inequalities for each possible pair of income groups in years t and s. We distinguish income groups because subsidies—and therefore post-subsidy premiums—vary across five income groups (0-100% of poverty, and four 50% of poverty groups from 100-300%). Lower-income groups both pay lower premiums overall and have narrower premium differences across plans. This generates substantial price variation that we can use to estimate κ_0. 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 time. This is a sensible assumption given the nature of plans in 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. 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. To account for this, we treat Network as two different plans, one before and one after 2012, and apply the rules above with that understanding. There were no other major changes in the networks of the plans during our study period. However, one plan enters mid-sample (Celticare in 2010), and one plan (Fallon) exits several areas in 2011.

To form the sample analogues of the inequalities in Theorem 3.5 we form cells with the same observed characteristics and $y_{i,s-1}$. The observed characteristics of a cell are denoted by x_i and are defined by the Cartesian product of: a) couple of years, b) region c) plan availability and d) income group.[11] So the λ_i represent differences in tastes among consumers with the same x_i and $y_{i,s-1}$. Recall that though the ratio of price sensitivity to switching cost is held constant, the price sensitivities themselves can vary in an arbitrary way and are not identified, and so not included, in the nonparametric analysis.

[11] As noted the more finely we condition on observable variables, the smaller we expect the within variance in the inequalities generated by the λ_i to be.

Table 1 provides summary statistics on the data used, which is constructed from all the cells defined above that have more than 20 members. We then sum the inequalities generated by these cells across regions and plan availabilities to obtain our groups, and use these in estimation. There are 242 groups defined by couple of years, income (the dimensions in which price could differ in all sample years), and prior choice. This generates between 8,926 and 23,644 inequalities depending on the value of κ. These inequalities use 75,000 comparisons of the choices made by the same individuals in different time periods.

Table 1: Summary Statistics for the Nonparametric Estimator

(s, t)	Number of Members	Number of Groups	Number of Members Above Cutoff	Number of Groups Above Cutoff	Minimum Number of Moments	Maximum Number of Moments
(2009, 2010)	19,550	96	17,349	66	1,494	3,671
(2010, 2011)	13,989	96	13,181	76	3,181	8,748
(2012, 2013)	47,266	120	44,438	100	4,251	11,225
Total	**80,805**	**312**	**74,968**	**242**	**8,926**	**23,644**

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 20 members, and the min and max number of moment inequalities (which depends on the value of κ being tested).

Economics and the Choice of Inequalities. There are benefits and costs to increasing the number of groups used. As we increase the number of inequalities by increasing the number of groups we (weakly) tighten the identified interval for κ. On the other hand Theorem 3.5 generates thousands of inequalities and all but two of them will be slack. The ability of test statistics to reject values of κ (weakly) decreases with the number of slack moments, and testing procedure designed to mitigate this problem are likely to be challenged by the number of inequalities in Table 1. So we turn to the economics underlying our choice model to select inequalities.

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.[12] Since the groups condition on time and income, they determine the prices used when forming an inequality as well as the initial choice, and as noted, the comparison of price changes to switching behavior underlies our estimators.

Figure 1 provides the average prices paid by consumers (i.e. after subsidy) by year and plan.[13] 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 is the year that the market regulators introduced a new rule that changed the nature of competition in the market.

Throughout, enrollees with incomes below the poverty line were fully subsidized. Prior to 2012 these enrollees were fully subsidized regardless of the plan they chose, but from 2012 on these new enrollees were only allowed to choose between the two lowest-priced plans (though below-poverty individuals already in the market prior to 2012 retained free choice among plans). 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.

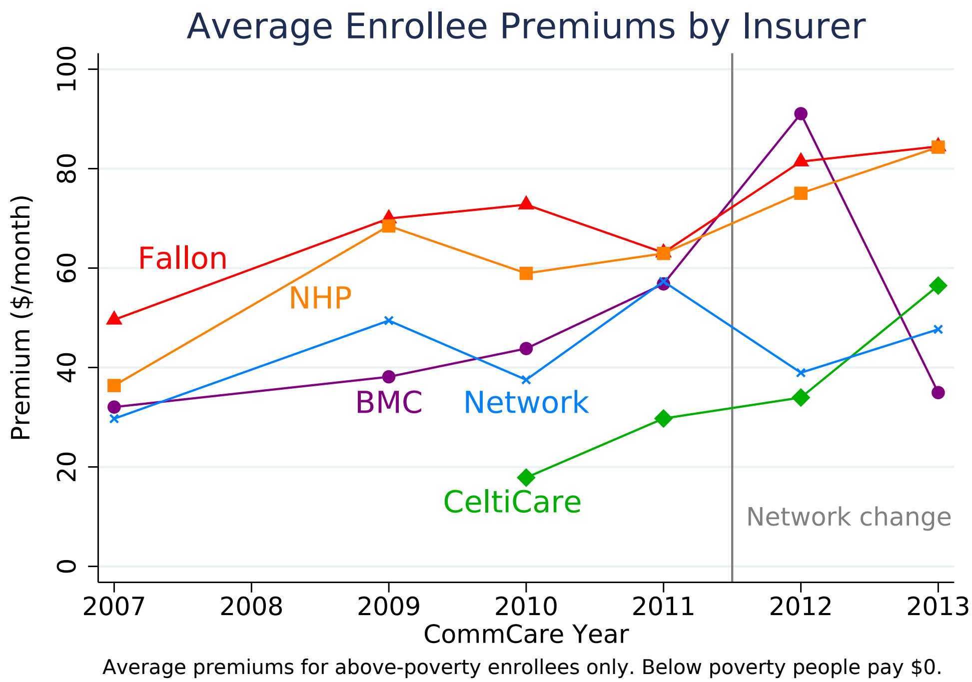
The different plans reacted differently to the change in rules. Boston Medical Center's plan (BMC), the plan with the largest share, increased its price sharply in 2012, essentially ceding the market for full-subsidy new enrollees to the other plans, but then lowered its price by an even greater amount in 2013. This generated the two largest price changes in Figure 1. As noted there were no major changes in BMC's network or 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 move. 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

[12]This assumes that the plan-specific effects are constant over time; but regulation insured that all 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; see Appendix A.

[13]Prices set in 2007 were locked in from 2007-08, which is why there are no separate points shown for 2008.

Figure 1: Choices and Premiums



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" (partly 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. The fraction who switch plans in 2012 is 14.5%; BMC, the plan with the largest price increase, loses 19% 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, about 36% (= 6.8%/19%) switch back in 2013, while only 4% (= 0.5%/12.3%) of those who switched out of other plans in 2012 switched back in 2013.

Two Sets of Results. We begin with the results that use all possible inequalities, as this illustrates the power problems likely when using a number of inequalities as large as those listed in Table 1. In our main results, which are presented next, we chose those groups which are most likely to generate bounds.

For an upper bound we know we need two switches, so we look for groups subject to a

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	111,226	36,235	74,991
Leave Market before 2013	76,007	24,812	51,195
Stay in Market 2011-13	35,219	11,423	23,796
Switching Rates (among stayers in market)			
Switch Plans from 2011-2012	14.5%	19.0%	12.3%
Switch in 2012, Switch Back in 2013	2.5%	6.8%	0.5%
Switch in 2012, Do Not Switch Back 2013	11.9%	12.1%	11.8%

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.

sharp relative price rise, followed by a relative price fall. Figure 1 makes it clear that the price changes most likely to induce this behavior are the changes in the relative prices of BMC from 2011 to 2013. So, to maximize the probability of getting an upper bound, we look to maximize $\Delta\Delta p^{d,c}_{i,t,t-1} + \Delta\Delta p^{c,d}_{i,t-1,t-2}$, where $c = BMC$ and $t = 2013$. If this is similar for two or more plan d's we enlarge D_0 to increase precision (assuming the conditions of Theorem 3.5 are satisfied).

For a lower bound we need $sw^{d,c}_{i,t,s} > 0$, which requires conditioning the $y_{i,t}$ choice on a plan which differs from the choice in $y_{i,s}$. There are three possible sequences of choices which could generate $sw^{d,c}_{i,t,s} > 0$, but two of them involve the agent choosing three different plans, and we do not have cells that satisfy this and our minimum cell-size condition.[14] The remaining sequence is $(y_{i,t} = d, y_{i,t-1} = c, y_{i,s} = c, y_{i,s-1} = d)$, and we require $p_{d,t} - p_{c,t} < p_{d,s} - p_{c,s-1}$ to get a positive lower bound. In the contrapositive here the probability of switching into d was smaller despite the relative price being lower, which indicates significant switching costs. To ensure we get a lower bound, we use the large price movements between 2012 and 2013, but this time $d = BMC$ and we chose the c to minimize $p_{d,t} - p_{c,t} - (p_{d,s} - p_{c,s})$.[15]

[14] If each of (d, c, e) represent distinct choices then the other two choice sequences were $(y_{i,t}, y_{i,t-1}, y_{i,s}, y_{i,s-1}) \in \{(d, c, c, e), (d, e, c, d)\}$.

[15] Recall that there are groups who could switch between the 2012 and 2013 open enrollment periods due to a change in family size or income. Consequently these groups could be conditioning on different choices

The Estimates. All estimates provided in this section are obtained as follows. We divide the negative part of each inequality used by its estimated standard errors, stack them, and compute the inequality with the largest negative value for each candidate κ. As suggested by Armstrong (2014), this becomes the sample value of the test statistic for that κ. The simulated value of the test statistics at $\alpha = .05$ for the given κ were obtained once without any adjustment for slack moments, and once using the adjustment proposed in Romano, Shaikh, and Wolf (2014, henceforth RSW), and the variance-covariance required for these calculations was obtained via a bootstrap.

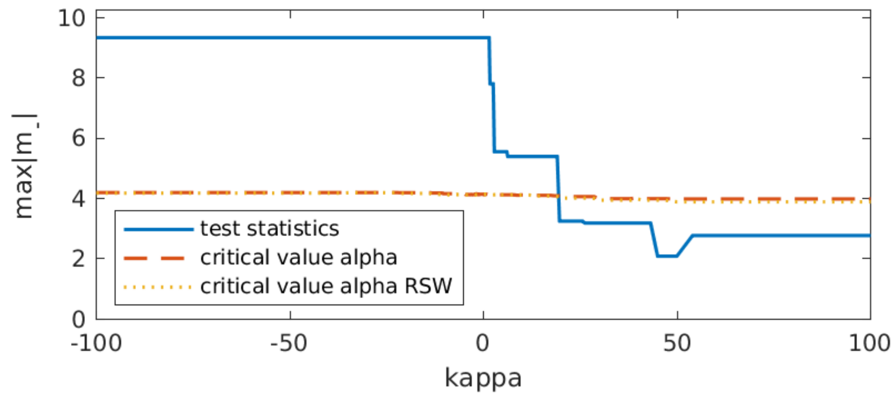
The results are presented in figures with κ values on the x-axis and the value of the test statistics on the y-axis. The blue line in each figure provides the value of the test statistic that the data generates. The long-dashed red line provides the five percent critical value for the κ values obtained from a test statistic that does not use a correction for slack moments, and the dotted red line is the test statistic when we use the RSW correction. Acceptable values of κ are all values where the blue line is lower than the red lines.

Figure 2 provides the results when we use the full set of inequalities listed in table (1). The blue line crosses the red lines at κ = \$19.6 and since it remains below it thereafter \$19.6 becomes $\underline{\hat{\kappa}}$, the lower bound of confidence set κ_0 at $\alpha = .05$. Though there is no upper bound, notice that the blue line is always noticeably above zero, indicating that there are inequalities that violate our conditions at high values of κ, but with this many moments our test statistics do not generate the power required to reject at those values.

Next we estimate using only the exogenously selected inequalities described above. Figure 3 panels (a) to (d) relate to the two groups which satisfy the conditions for yielding upper bounds. Panel (a) graphs prices facing a particular income group in the Boston area, and panel (b) graphs the prices for an income group in Western Massachusetts. These two panels demonstrate that there are groups with price changes in those years that satisfy the condition for the upper bound, that is for which $(p_{d,i,t} - p_{c,i,t}) - (p_{d,i,t-1} - p_{c,i,t-1}) > 0$ (with choice $c =$ BMC). Panels (c) and (d) show the corresponding test statistics and critical values for these cells. It shows sharp increases in the test statistics from the sample at κ values greater then

in 2012 and 2013 without the choice conditioned on in 2013 being the open enrollment choice in 2012. This enables us to look at a 2013 choice which does not condition on the 2012 choice, and make use of the large movements in prices between 2012 and 2013.

Figure 2: Nonparametric Estimates of Switching Costs (κ_0)



Note: The nonparametric estimator restricts the comparisons to individuals with the same choice sets at time s and t. Cell level moments are constructed, and then aggregated into groups. The lower bound identified is at $19.6.

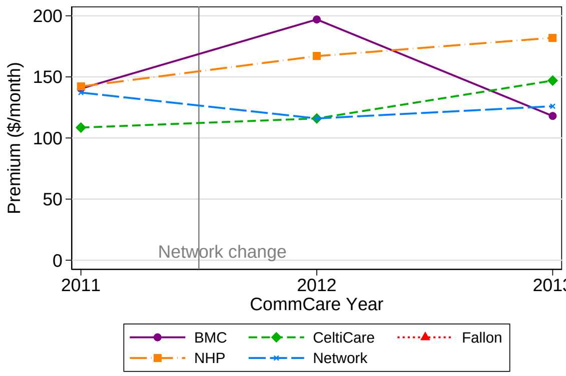
$48.5 and $54 respectively. Panels (e) and (f) show test statistics and critical values based on the two cells that underlie the lower bound in Figure 2.

Panel (g) provides the test statistics when we use the inequalities generated by all four of the selected groups. When we put the four groups which satisfy our exogenous selection criteria together we obtain the confidence interval $\kappa_0 \in [\underline{\hat{\kappa}} = 19.6, \hat{\bar{\kappa}} = 57]$ and a sample value of the criterion which is zero between 19.6 and 48.5. So though our estimates indicate that there are switching costs, they indicate that the switching costs are not higher than $57. This is about an average month's premium for an individual, and as discussed in the introduction, is substantially less than switching costs estimated on this data previously. We now consider whether parametric estimates are consistent with these nonparametric values, and then consider their implications.

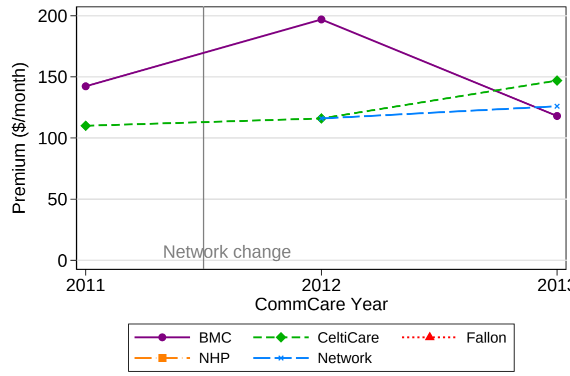
4 Parametric Models

We want to consider the sources of the difference between the bounds obtained from our nonparametric procedure and the estimates from prior research. Of particular interest is our upper bound and the implications of the difference between it and the noticeably higher point estimates of κ obtained in prior studies of health insurance. To investigate further we

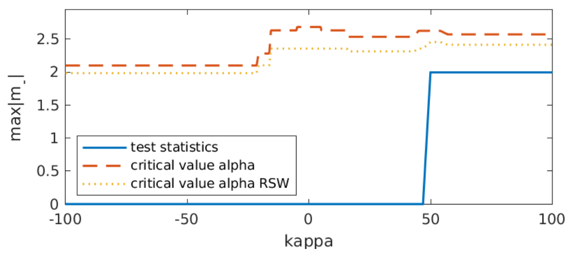
Figure 3: Average premiums and nonparametric estimators in two selected cells



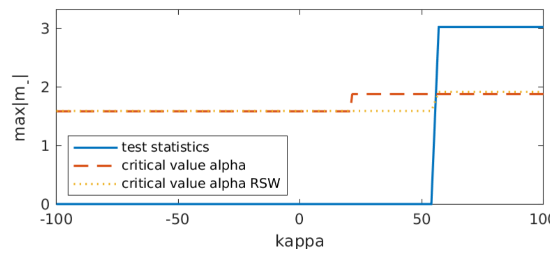
(a) Upper bound cell #1



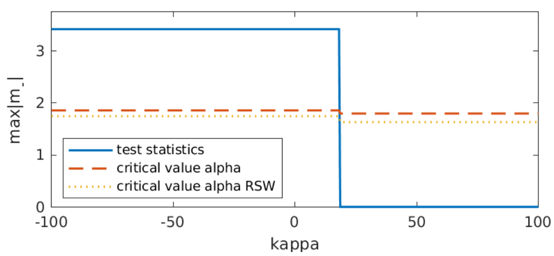
(b) Upper bound cell #2



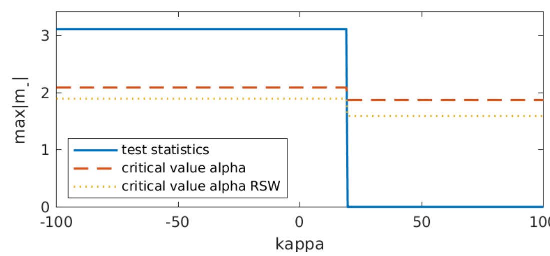
(c) Upper bound cell #1



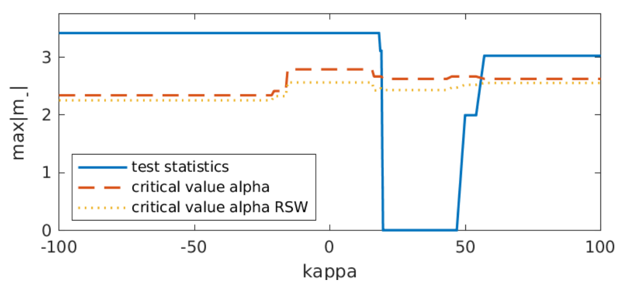
(d) Upper bound cell #2



(e) Lower bound cell #3



(f) Lower bound cell #4



(g) Combining all four cells

Note: The figures show groups with price changes that generate upper and lower bounds, which come from BMC's large price changes from 2011-2013. The groups generating bounds all have a lagged 2011 choice of BMC. They are defined defined by a given region and income group path (and associated choice set) from 2012-13. Upper bound cell #1 (panels (a), (c)) has income 250-300% of poverty in 2012-13, in Boston with choice set: {BMC, CeltiCare, NHP, Network}. Upper bound cell #2 (panels (b), (d)) have income 250-300% of poverty in 2012-13, in Western MA with choice set: {BMC, CeltiCare, Network}. Lower bound cell #1 (panel (e)) has income 100-150% of poverty in 2012-13, in Western MA with choice set: {BMC, Network}. Lower bound cell #2 (panel (f)) has income 100-150% of poverty in 2012 and 150-200% of poverty in 2013, in Western MA with choice set: {BMC, Network}.

compare our nonparametric results to the results we obtain when we use parametric distributions for disturbances like those used in prior research. We begin by developing parametric models that allow for state dependence and fixed effects, and then, for comparison, estimate an assortment of models that do not allow for fixed effects.

Parametric models require a distributional assumption for the random utility disturbances and a functional form for β_i. We begin by assuming only that β_i can be written as a function of observed variables x_i that do not vary over time, or $\beta_i = \beta(x_i)$. In addition the shape restrictions in a parametric model imply we can identify separate coefficients on price and the lagged dependent variable, so we re-write the utility function in equation (1.1) as

$$U_{d,i,t} = (-p_{d,it}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq d\}\delta_0)\,\beta(x_i) + \lambda_{d,i} + \epsilon_{d,i,t}, \tag{4.1}$$

where $(\beta(x_i), \delta_0, \gamma_0) > (0,0,0)$. However we remain focused on estimating the tradeoff between price and switching costs, i.e. on $\kappa_0 \equiv \delta_0/\gamma_0$.

We begin by showing that parametric models generate a transparent set of inequalities that will be available regardless of the exact form of the distribution of $\epsilon_{d,i,t}$. To use these inequalities, the researcher has to specify a particular functional form for the disturbance distribution and this can generate additional inequalities. Prior results on our data used a logistic distribution. So in going to the data we add inequalities that are available for that case.

General Parametric Inequalities. Given the model in equation (4.1), if Assumption 2.1 holds and $(d,c) \in \mathcal{D}_t \cap \mathcal{D}_s$ with $d \neq c$, the logic that led to equation (3.1), now leads to

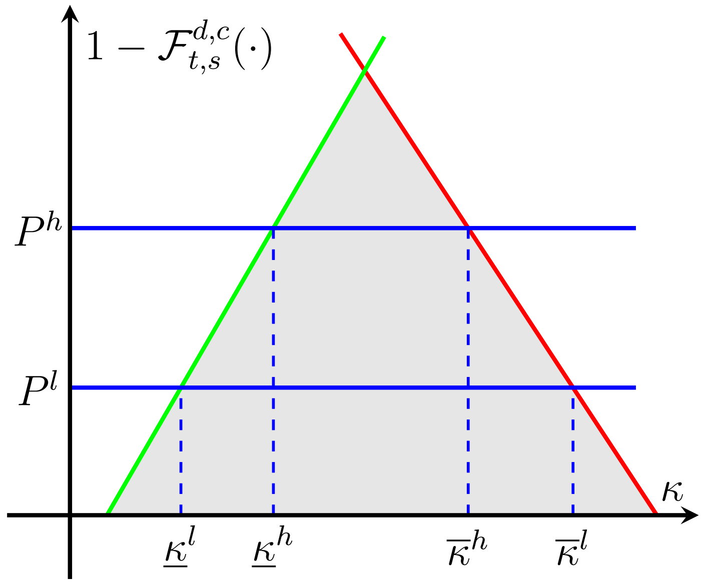
$$\Pr(y_{i,t} = d, y_{i,s} = c \,|\, p_i, y_{i,s-1}, x_i) \leq\; 1 - \mathcal{F}^{d,c}_{t,s}\Big((\Delta\Delta p^{d,c}_{i,t,s} + \kappa_0 sw^{d,c}_{i,t,s})\gamma_0\beta(x_i)\Big). \tag{4.2}$$

where $\mathcal{F}^{d,c}_{t,s}(\cdot)$ is the distribution of $\Delta\Delta\epsilon^{d,c}_{i,t,s}$, and we have substituted $\kappa_0 \times \gamma_0$ for δ_0.

Figure 4 illustrates the intuition behind the bounds on κ_0 that this inequality generates. The left hand side of this inequality is a number we estimate from the data. The right hand side depends on κ_0. If $sw^{d,c}_{i,t,s} > 0$, the boundary of the inequality will be decreasing in κ and all values smaller than the boundary will be acceptable, producing the red line in the figure.

If $sw^{d,c}_{i,t,s} < 0$ the boundary of the inequality will increase in κ and we will accept all values to the right of the boundary. An empirical probability that intersects the resulting triangle twice, produces an upper (our $\overline{\kappa}$) and a lower ($\underline{\kappa}$) bound to κ_0. Notice that the higher is $Pr(y_{i,t} = d, y_{i,s} = c \,|\, p_i, y_{i,s-1})$ the tighter are the bounds we derive.

Figure 4: Identified set for κ_0



More formally each inequality generated by equation (4.2) generates a line which divides the (γ, δ) plane into acceptable and non-acceptable half-spaces. The slope and/or quadrant of the acceptable half-space differs with

- $sw^{d,c}_{i,t,s} \in \{-2, -1, 0, 1, 2\}$, and
- the sign of $\Delta\Delta p^{d,c}_{i,t,s}$ (greater than or less than zero).

This seems to generate twelve different cases. However, recall that if $s < t - 1$ and we condition on $y_{i,s-1}$, we can arrive at $(y_{i,t} = d, y_{i,s} = c)$ from different intermediate $y_{i,t-1}$ states. If $y_{i,t-1} = r$ generates an $sw^{d,c}_{i,t,s} = 1$ and equation (4.2) holds for a particular value of κ, it will also hold for values of $y_{i,t-1}$ which generate $sw^{d,c}_{i,t,s} = 2$. This implies that

$$\sum_{r:sw^{d,c}(r,y_{i,s-1})\geq \mathsf{y}} \Pr(y_{i,t} = d, y_{i,t-1} = r, y_{i,s} = c|p_i, y_{i,s-1}, x_i) \leq 1 - F^{d,c}_{i,t,s}\big((\Delta\Delta p^{d,c}_{i,t,s}\gamma_0 + \mathsf{y}\delta_0)\beta(x_i)\big). \tag{4.3}$$

Moreover by using the sum, instead of each individual inequality separately, we elevate the line in figure Figure 4 and generate sharper bounds for κ_0. An analogous argument implies that if the inequality holds for $sw_{i,t,s}^{c,d} = -2$ it also holds for $sw_{i,t,s}^{c,d} = -1$. So there are only six cases to consider, and Appendix B considers each case with explanatory graphs.[16] The graphs shows that only four of the six generate restrictions in the appropriate quadrants so we only use those four cases in our empirical work.[17]

The Magic of Logits. Next we explore the implications of assuming a Gumbel (logistic) distribution for ϵ.

Assumption 4.1. *Assumption 2.1 holds 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.*

Assumption 4.1 yields the traditional logit form for the choice probabilities,

$$\begin{aligned} &\mathcal{P}_{d,i,t|y_{i,t-1}} \equiv \Pr(y_{i,t} = d \,|\, p_i, y_{i,t-1}, x_i, \lambda_i) \\ &= \frac{\exp[(-p_{d,i,t}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq d\}\delta_0)\,\beta(x_i) + \lambda_{d,i}]}{\sum_r \exp[(-p_{r,i,t}\gamma_0 - \mathbf{1}\{y_{i,t-1} \neq r\}\delta_0)\,\beta(x_i) + \lambda_{r,i}]} \equiv \frac{\mathcal{N}_t(d, y_{i,t-1})e^{\lambda_{d,i}}}{\mathcal{M}_t(y_{i,t-1}, \lambda_i)}. \end{aligned}$$

So the ratio of the probability of choosing d at t and c at $t-1$, to the probability of choosing c at t and d at $t-1$ (conditional on $y_{i,t-2}$) is

$$\frac{\mathcal{P}_{d,i,t|c}}{\mathcal{P}_{d,i,t-1|y_{i,t-2}}}\frac{\mathcal{P}_{c,i,t-1|y_{i,t-2}}}{\mathcal{P}_{c,i,t|d}} = \frac{\mathcal{N}_t(d, y_{i,t-1} = c)}{\mathcal{N}_{t-1}(d, y_{i,t-2})}\frac{\mathcal{N}_{t-1}(c, y_{i,t-2})}{\mathcal{N}_t(c, y_{i,t-1} = d)} \times \frac{\mathcal{M}_t(y_{i,t-1} = d, \lambda_i)}{\mathcal{M}_t(y_{i,t-1} = c, \lambda_i)}.$$

Only the last term in this expression depends on the fixed effects and

$$exp(\delta_0\beta(x_i)) \geq \frac{\mathcal{M}_t(y_{i,t-1} = d, \lambda_i)}{\mathcal{M}_t(y_{i,t-1} = c, \lambda_i)} \geq exp[-\delta_0\beta(x_i)].$$

The ratio of the odds of choosing $(y_{i,t-1} = c, y_{i,t} = d)$ to $(y_{i,t-1} = d, y_{i,t} = c)$ is bounded

[16] This presumes $Pr(y_{i,t} = d, y_{i,s} = c \,|\, p_i, y_{i,s-1}) < .5$, as is the case in our data. If the median of $\mathcal{F}_{t,s}^{d,c}(\cdot) = 0$, as it will for our distributional assumption, $\mathcal{F}_{t,s}^{d,c}(\Pr(\cdot|\cdot))^{-1}$ differs in sign according as $\Pr(\cdot|\cdot) \lesseqgtr 1/2$, and this will double the numer of cases. Appendix B considers this case also.

[17] These are the inequalities for the following cases: $\Delta\Delta sw > 0$ and $\Delta\Delta p \geq 0$ or ≤ 0; $\Delta\Delta sw < 0$ and $\Delta\Delta p \geq 0$; and $\Delta\Delta sw = 0$ and $\Delta\Delta p \geq 0$.

by functions that are independent of the $\{\lambda_i\}_i$.[18] Rearranging terms we get the inequalities in the theorem that follows for choice probabilities at t and $t-1$. A more detailed argument shows that similar inequalities are valid for the odds ratio at t and $t-2$.[19]

Theorem 4.2. *Suppose Assumption 4.1 holds, $s \in \{t-1, t-2\}$, and $(d,c) \in \mathcal{D}_t \cap \mathcal{D}_s$. Then,*

$$\begin{aligned} exp\left[\gamma_0\left(\Delta\Delta p_{i,t,s}^{c,d}\right)\beta(x)\right] &\leq \frac{\Pr(y_{i,t}=d, y_{i,s}=c \,|\, p_i, y_{i,s-1}=c, x_i=x)}{\Pr(y_{i,t}=c, y_{i,s}=d \,|\, p_i, y_{i,s-1}=c, x_i=x)} \\ &\leq exp\left[\left(2\delta_0+\gamma_0\left(\Delta\Delta p_{i,t,s}^{c,d}\right)\right)\beta(x)\right]. \end{aligned}$$

It is worth pointing out that the logit assumption guarantees meaningful upper and lower bounds on κ_0.[20]

4.1 Parametric Empirical Results

Table 3 provides summary statistics for the data and inequalities used in the parametric analysis. The only difference between the sample used for the nonparametric analysis (described in section 3.3) and that used in the parametric analysis, is that in the parametric analysis we keep groups who face different plan offerings in the comparison periods (recall that our assumptions ruled this out for the nonparametric analysis). This increases the size of the data set considerably. The number of inequalities generated by the parametric analysis is, on the other hand, much smaller than in the nonparametric analysis.

We begin with the results from models that allow for state dependence but do not allow for flexible fixed effects. This will enable a comparison of our nonparametric results to the parametric results from the point identified models. We then proceed to the parametric

[18]This finding is reminiscent of Chamberlain (1980) who derived a conditional likelihood that did not depend on the $\{\lambda_i\}_i$ for the multinomial logit panel case with no lagged dependent variable.

[19]See Appendix C. Theorem 4.2 does not exhaust the additional inequalities available when the disturbance distribution is logistic. Additional inequalities are stated as Theorem C.2 in Appendix C and cover the cases: (i) $y_{i,s-1} = r \notin \{c,d\}$; and (ii) $s < t-2$. In different applications, these additional inequalities could be quite useful, but here they are relegated to the appendix as our data does not have groups of sufficient size to exploit them.

[20]More precisely, the general inequalities when none of the probabilities are greater than one half need to produce upper and lower bounds for δ but need only produce an upper bound for γ. A lower bound for γ is needed to produce an upper bound to $\kappa \equiv \delta/\gamma$. Under the logit assumption, both log odds ratios have positive probability, so the first inequality in Theorem 4.2 provides upper [lower] bound information on γ_0 when $\Delta\Delta p_{i,t,s}^{c,d}$ is positive [negative]. The second inequality yields lower bound information on δ_0.

Table 3: Summary Statistics for the Parametric Estimator

(s, t)	Number of Members	Number of Groups	Number of Members Above Cutoff	Number of Groups Above Cutoff	Number of Moments
(2009, 2010)	59,322	100	32,738	69	248
(2009, 2011)	39,955	100	24,740	78	296
(2010, 2011)	59,629	100	34,300	83	217
(2012, 2013)	69,441	125	43,138	99	522
Total	**228,347**	**425**	**134,916**	**329**	**1,283**

Notes: The table shows sample statistics for the parametric estimator, analogous to Table 1 for the nonparametric estimator. The minimum cell size cutoff is 20 members.

models that allow for fixed effects, and ask whether these are consistent with our nonparametric results. All results, both from the comparison models and from the models of section 4, assume that the distribution of the disturbances is logistic. Note that this generates a distribution for the double difference of disturbances (for $\Delta\Delta\epsilon_{i,t,s}^{d,c}$) that is analytic, simplifying computation.[21]

4.2 State Dependence Without Fixed Effects

Table 4 summarizes the results from a number of specifications. The estimate of the switching cost is always obtained as the ratio of the lagged dependent variable coefficient to the price coefficient.[22]

The first three columns of the table present results from specifications in which the individual-specific fixed effects used in the inequality analysis are replaced with increasingly detailed interactions of individual characteristics with plan dummies. Column (1) has no plan

[21] That distribution and its density are

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

[22] Its standard error is obtained from a Taylor expansion (i.e., the Delta method), which in this context should be accurate as all the price coefficients are two or more orders of magnitude greater than their standard errors. As in all discrete choice models the comparison models require a normalization. We normalize the disturbance to have a standard Gumbel distribution (with variance of $\pi^2/3$). So both the coefficient of price and of the lagged dependent variable should be thought of as the variable's coefficient divided by this standard error.

Table 4: Multinomial Logit Estimation

	Simple (1)	Plan Dummies (2)	Detailed Plan Dum. (3)	Detailed Plan Dum. + Network (4)	Plan Dum. + Random Effects (5)	Include New Enr (6)	New Enr + Random Effects (7)
Normalize $\epsilon_{i,d,t}$ to EV1							
Switching Cost (δ)	-4.086	-4.196	-4.156	-4.120	-4.480	-3.974	-4.478
	(0.006)	(0.007)	(0.007)	(0.007)	(0.014)	(0.006)	(0.010)
Price Coefficient (β)	-0.036	-0.049	-0.051	-0.053	-0.051	-0.036	-0.041
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Hospital Network Utility	—	—	—	0.137	—	—	—
				(0.007)			
Prev. Used Hospitals Covered	—	—	—	0.804	—	—	—
				(0.019)			
Prev. Used x Partners Hosp.	—	—	—	0.974	—	—	—
				(0.026)			
Normalize β to 1							
Switching Cost ($\kappa = \delta/\beta$)	114.03	86.35	80.76	78.13	87.88	110.79	110.17
	(0.55)	(0.34)	(0.32)	(0.30)	(0.35)	(0.43)	(0.44)
Plan Dummies	—	Yes	Yes	Yes	Yes	Yes	Yes
Plan x (Area, Age-Sex, Illness)	—	—	Yes	Yes	—	—	—
Plan Random Effects	—	—	—	—	Yes	—	Yes
N Parameters	2	7	249	252	11	7	12
N Individuals x Years	2,623,699	2,623,699	2,623,699	2,623,699	2,623,699	3,832,629	3,832,629

dummies; column (2) has simple plan dummies; and column (3) interacts each plan dummy with 20 age-sex groups, 38 geographic areas ("service areas" determined by the state), and with three chronic illness groups (where the three sets of interactions are additively separable). After excluding interactions with no observations, this generates 247 dummy variables (which with the price and switching variables, implies 249 total model parameters).

Column (4) further adds three variables designed to measure individual-specific aspects of the hospital network that our cell-specific variables do not capture. These variables, taken from Shepard (2022), 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 (Capps et al., 2003), and two variables which measure the share of a patient's previously used hospitals that are 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). Shepard (2022) finds these network variables to be highly statistically significant, and they remain so after we include the 247 interactions with the plan dummies.

The switching cost estimate declines monotonically as we add interactions, from \$114.03 (0.55) to \$78.13 (0.30), where here and below the numbers in parentheses are standard errors. Including more detailed plan dummies and network variables *does* improve the model's ability to capture unobserved preference heterogeneity. But notably, these estimates of κ_0 are all substantially larger than the upper bound of \$57 generated by the nonparametric results. This suggests that there is still substantial unobserved heterogeneity even with the very detailed specification of column (4).

Next we replace the fixed effects in the inequality analysis with random effects. That is we interact the plan dummies with agent-specific independent normal random variables that are held constant over the period the individual is observed, and use simulated maximum likelihood to estimate. We begin the random effect analysis by allowing for random effects conditional on the first observed choice. So this analysis assumes both that: (i) the within group variance in the plan specific effects is normal with variances that vary by plan (but not by group), and (ii) is uncorrelated with the initial observed choice. This mimics what researchers have done in related problems when they do not have sufficient information on the actual initial choices of individuals. The results from this specification are provided in column (5) of the table. The random effects model generates a switching cost of \$87.88 (0.35), and estimated dummy variables for the plan and standard errors for the random effects which are both highly significant for all but the smallest plan (Fallon). The t-values for the standard errors varied from eight to over fifty.

We are in the enviable position of knowing the first time a consumer enters the Massachusetts exchange. So provided we are willing to assume that any pre-exchange health choices of these individuals does not influence their behavior on the exchange, we can implement an "initial conditions estimator" that allows for normal random draws on preferences for the exchange's plans that are known to the consumers before making their first choice. Column (7) provides the simulated maximum likelihood estimates for this specification.[23] Since column (7) adds the first choice to the switching choices analyzed in columns (1) to

[23] We also tried to use the estimator suggested by Honoré and Kyriazidou (2000), an estimator which does allow for both fixed effects and switching costs. However, their restrictions left us with data on 36 individuals and 144 choices, which was not sufficient to obtain estimates with reasonable precision.

(5), it uses a different data set than those columns did. So for comparison column (6) uses the column (7) data in a model without random effects; i.e. column (6) mimics column (2) but uses the data set used in column (7).

When we include initial choices and simple plan dummies (but no random effects) in column (6), our estimated switching cost is \$110.79 (0.43), noticeably larger than in the analogous specification without initial choices (column (2)). When we also include random effects in column (7), the plan dummies and the estimated standard deviations of the random effects for all plans are estimated to be even larger than those from column (5), with t-values for the estimated variances of the plan specific random effects for all plans (including Fallon) now ranging from twelve to over one hundred. Perhaps more surprising is that the estimates of κ_0 in both columns (6) and (7) are quite similar at \$110.79 (0.43) and \$110.17 (0.44).

Part of the reason that the initial condition estimator generates relatively large values of κ_0 may be that some consumers had experience in making health insurance choices before entering CommCare. This might have generated priors when entering the program that plans differed in their coverage, out of pocket payments, etc. After entering they would have learned that regulation requires these features to not vary across plans. As a result they become more price sensitive in subsequent periods.[24]

We conclude that models that do not allow for individual by product fixed effects generate estimates of κ_0 that lie somewhere between \$78 and \$114. This accords well with published work on the CommCare data which contains estimates of about \$100 (see Shepard (2022)). Recall from Figure 1 that average monthly premiums ranged from \$20 to \$90. So the models without fixed effect generate switching costs which are four to five times the average monthly premium for the lowest cost plan and about equal to the average monthly premium for the highest cost plan. These estimates are noticeably larger than the nonparametric upper bound obtained from our analysis. There is a question of how much of the difference can be attributed to the logit assumption and how much to the absence of fixed effects.

[24]For a model with fixed effects that explicitly allows for Bayesian learning see Aguirregabiria et al. (2021).

4.3 Parametric Estimators That Allow For Fixed Effects.

This subsection uses the inequalities from section 4, and the sample described in Table 3 (also used in columns (1) to (5) of Table 4), to estimate bounds on κ_0. This generates over 1,200 inequalities from about 330 groups with an average size of over 400 individuals.

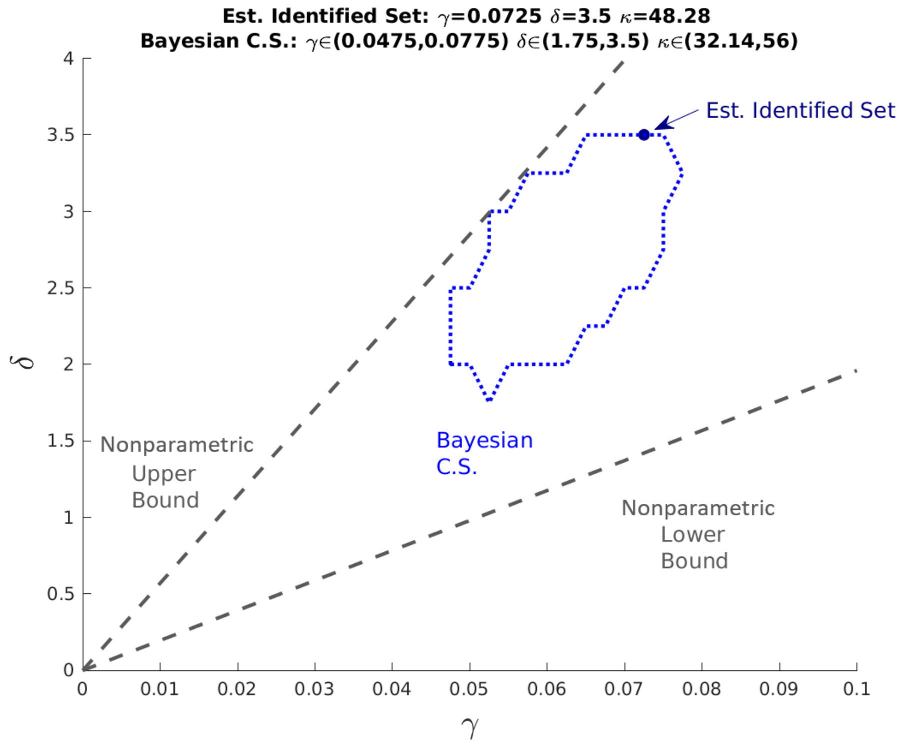
When we used the estimation algorithm described at the beginning of section 4 for the current specification, the simulated value of the test statistic obtained from the normal approximation to the distribution of the moments often implied probabilities that were negative, rendering the assumptions underlying that asymptotic approximation inappropriate. We present the point estimate from that estimation algorithm but do not want to rely on its simulated test statistics for inference. Instead we use the Bayesian approach proposed by Kline and Tamer (2020) with the implementation in Chamberlain and Imbens (2003). This combines an uninformative prior with the data to generate a multinomial posterior distribution for the probabilities.[25] We then take draws from this posterior, calculate the (possibly set-valued) estimate of the parameters that minimizes the sup-norm of the negative part of the inequalities for each draw, and then find a conservative 95 percent confidence set for γ_0, δ_0, and separately for $\kappa_0 \equiv \delta_0/\gamma_0$.

The results are plotted in (γ, δ) space in Figure 5. The point estimate from minimizing the largest of the negative parts of the moments is given by the dark blue dot. The 95% confidence sets for γ_0, δ_0 and linear combinations of the two are obtained from the 2.5% and 97.5% quantiles of the distribution of their lower and upper bounds found from the posterior draws. The accepted (γ_0, δ_0) combinations are given by the area interior to the shape produced by the blue dots in the figure. The (γ_0, δ_0) combinations that generated our lower and upper nonparametric bounds for κ_0 are given by the dashed grey lines.

The "point estimate" of κ_0 from the moment minimization problem was $\hat{\kappa}$=\$48.3. Recall

[25] Treating the choice probabilities for each cell as a multinomial distribution, Chamberlain and Imbens (2003) show that the Dirichlet distribution with parameters set to the observed frequencies is the posterior distribution for the multinomial distribution with uninformative Dirichlet prior. Since the parameter identified set is a simple transformation of the cell probabilities, we follow Kline and Tamer (2020) and form a credible set for the identified set by straightforward simulation from the Dirichlet posterior. Given the large number of inequalities, when a simulated draw of probabilities generates an empty identified set, we conservatively include the parameter value(s) that minimize the criterion based on the worst violation of the inequalities. Kline and Tamer (2020) show the asymptotic connection to a frequentist confidence interpretation of the resulting intervals.

Figure 5: Identified set and Bayesian confidence set for γ_0 and δ_0.



Notes: Cell-level moments based on the general parametric and the magic of logit inequalities are constructed, and then aggregated into groups. The area inside the dashed blue curve is the Bayesian confidence set for the estimates of (γ, δ). The dashed gray lines represent the upper and lower bounds on κ $(= \delta/\gamma)$ from the nonparametric estimator in Section 3.

that Figure 3 showed a large but statistically insignificant jump in the sample test statistic of the nonparametric results at \$48.5. So the parametric assumptions seem to be consistent with, but more powerful than, the nonparametric results.

To get a more precise quantification of the power generated by the parametric assumptions we use the 95% credible interval produced by the Bayesian bootstrap. This generates estimates of $\hat{\underline{\kappa}} = \$32$ and of $\hat{\overline{\kappa}} = \$56$. Recall that the lower bound from our nonparametric estimates was \$20, so the parametric assumptions lead to a considerably tighter lower bound than the nonparametric analysis. However, the upper bound, which is the bound of particular interest, is almost identical to the upper bound of \$57 we obtained from the nonparametric results.

Moreover, as should be the case if our assumptions are an adequate approximation to reality, the confidence set from the parametric analysis that allows for individual-by-product specific fixed effects lies within the bounds obtained from the nonparametric analysis that allows for fixed effects. On the other hand the entire confidence set lies well below any of the point estimates obtained from the comparison models in section 4.2 that do not allow for fixed effects. So the comparison models seem to overestimate the switching cost by a considerable amount. We turn next to an investigation of whether this difference is likely to influence the economic implications of the estimated models.

4.4 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 4, the other uses the γ_0 estimate from the relevant comparison model but restricts δ to equal $\gamma\hat{\kappa}$ where $\hat{\kappa} = \$48.28$, as in Figure 5. 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 5 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 premiums), was $58.4 per month in 2011. In 2012 that average increased

Table 5: Counterfactual Comparisons

Specification	2011 market shares	2012 status-quo	2012 counterfactual	2012 *% diff*	2013 status-quo	2013 counterfactual	2013 *% diff*
Market shares without imposing κ							
Plan FE	0.357	0.289	0.321	*11.0*	0.266	0.304	*14.2*
Plan × Region FE	0.357	0.289	0.320	*10.6*	0.266	0.298	*12.2*
Plan FE + RE	0.357	0.282	0.324	*15.1*	0.289	0.311	*7.6*
Market shares imposing κ							
Plan FE	0.357	0.186	0.306	*64.1*	0.399	0.326	*-18.3*
Plan × Region FE	0.357	0.205	0.313	*52.8*	0.381	0.318	*-16.6*
Plan FE + RE	0.357	0.183	0.305	*66.8*	0.410	0.329	*-19.7*
Premium	58.4	91.1	62.9		41.5	65.3	

Note: Table shows a counterfactual comparison of BMC market shares among current enrollees above 100% FPL. The top panel shows observed market shares in 2011, and then predicted market shares under status-quo premium and counterfactual premium, as well as their percentage difference in 2012 and 2013. We include results based on two FE specifications and one random coefficient specification. "Imposing κ" indicates whether we restrict the switching cost coefficient. The bottom panel shows the average BMC premium under status-quo and counterfactual in 2011-2013.

to \$91.1, and in 2013 it fell to \$41.5; 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 \$63 in 2012 and \$65 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 \$91.1 to \$63 leads to a prediction of an 11% to 15% increase in share when we use the parameters estimated by the comparison models, but a prediction of a dramatic 53% to 67% increase in share when we constrain $\hat{\kappa} = 48.28$. In 2013 when the counterfactual average price was \$65 compared to the actual average price of \$41.5 , the estimates from the comparison models predict an 8 to 14% *higher* share from the *higher* counterfactual price. In contrast when we use $\hat{\kappa} = 48.28$ the higher counterfactual price in 2013 generates a two period prediction of a 17 to 20% *lower* share than the prediction from the status quo prices.

Recall that this is the prediction for 2013 which conditions on 2011 shares and the counterfactual prices in both 2012 and 2013. The comparison models do predict the shares fall from 2012 to 2013 (by 1 to 2%). However because the comparison models' estimates of κ_0 are so high, this decrease is more than offset by the comparison model's increased share in 2012. That is, 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 What Have We Learned?

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

Our empirical results indicate that health insurance estimates of state dependence that do not allow for very flexible unobserved heterogeneity seem to seriously bias estimates of switching costs upwards; in our data by a factor of 37-100 percent. We found this bias regardless of whether the comparison model (without individual-by-product fixed effects) allows for a rich set of plan interactions, random effects conditional on the initial choice, or random effects known prior to the initial choice. 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 counterfactual, the reversal of what seems to be a failed pricing experiment by the largest insurer, illustrated that the difference in estimated switching costs matters. The comparison models' predicted a one year share change of 10-15% while when we use our estimate of κ we find a share change of 55-65%. Moreover the analogous predicted differences for the share change over the two years that includes the insurer's policy reversal actually differ in sign; so longer-term predictions using the comparison models' κ estimates can be particularly problematic.

Our methodological results on estimators that allow for both state dependence and fixed effects depend on what the researcher is willing to assume on the distribution of the disturbances. If one does not want to assume a parametric distribution for ϵ and switching probabilities are less than a half, then finite positive bounds for κ_0, the ratio of price sensitivity to switching costs, are obtained by employing Assumption 3.2, the stationarity assumption. If switching probabilities are greater than a half, then nonparametric revealed preference inequalities can be used to bound κ_0 either with or without adding the inequalities from the stationarity assumption. If we are willing to make a *parametric* assumption on the ϵ distribution, then we can obtain tighter bounds. Moreover if the specified distribution is the logistic distribution, then there will be positive finite upper and lower bounds that are exceptionally easy to compute.

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, 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, 512–521.

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

Armstrong, T. B. (2014): "Weighted KS statistics for Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 181, 92–116.

Bonhomme, S. (2012): "Functional differencing," *Econometrica*, 80, 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, 2472–2508.

Brot-Goldberg, Z. C., T. Layton, B. Vabson, and A. Y. Wang (2021): "The behavioral foundations of default effects: theory and evidence from Medicare Part D," Tech. rep., National Bureau of Economic Research.

Brown, Z. Y. and J. Jeon (2020): "Endogenous information and simplifying insurance choice," Tech. rep., Mimeo, University of Michigan.

Capps, C., D. Dranove, and M. Satterthwaite (2003): "Competition and market power in option demand markets," *RAND Journal of Economics*, 737–763.

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

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

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

Conley, T. G. and C. R. Udry (2010): "Learning About a New Technology: Pineapple in Ghana," *American Economic Review*, 100, 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, 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, 435–445.

——— (2010): "State dependence and alternative explanations for consumer inertia," *The RAND Journal of Economics*, 41, 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, 38–64.

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

Fox, J. (2007): "Semiparametric Estimation of Multinomial Discrete-Choice Models using a Subset of Choices," *The RAND Journal of Economics*, 38, 1002–1019.

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

Heckman, J. (1978): "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence," *Annales de l'inséé*, Apr–Sep, 227–269.

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, 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, 877–905.

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

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

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

Honoré, B. E. and M. Weidner (2020): "Moment Conditions for Dynamic Panel Logit Models with Fixed Effects," Tech. rep.

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, 192–226.

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

Khan, S., F. Ouyang, and E. Tamer (forthcoming): "Inference on Semiparametric Multinomial Response Models," *Quantitative Economics*.

Khan, S., M. Ponomareva, and E. Tamer (2020): "Identification of Dynamic Binary Response Models," Tech. rep., Working Paper.

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, 1123–1167.

Manski, C. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

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

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

Pakes, A. and J. Porter (2016): "Moment Inequalities for Multinomial Choice with Fixed Effects," Tech. rep., National Bureau of Economic Research.

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

Romano, J. P., A. M. Shaikh, and M. Wolf (2014): "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 82, 1979–2002.

Samuelson, P. A. (1938): "A Note on the Pure Theory of Consumer's Behaviour," *Economica*, 5, 61–71.

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

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

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

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

Tebaldi, P., A. Torgovitsky, and H. Yang (2019): "Nonparametric estimates of demand in the california health insurance exchange," Tech. rep., National Bureau of Economic Research.

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

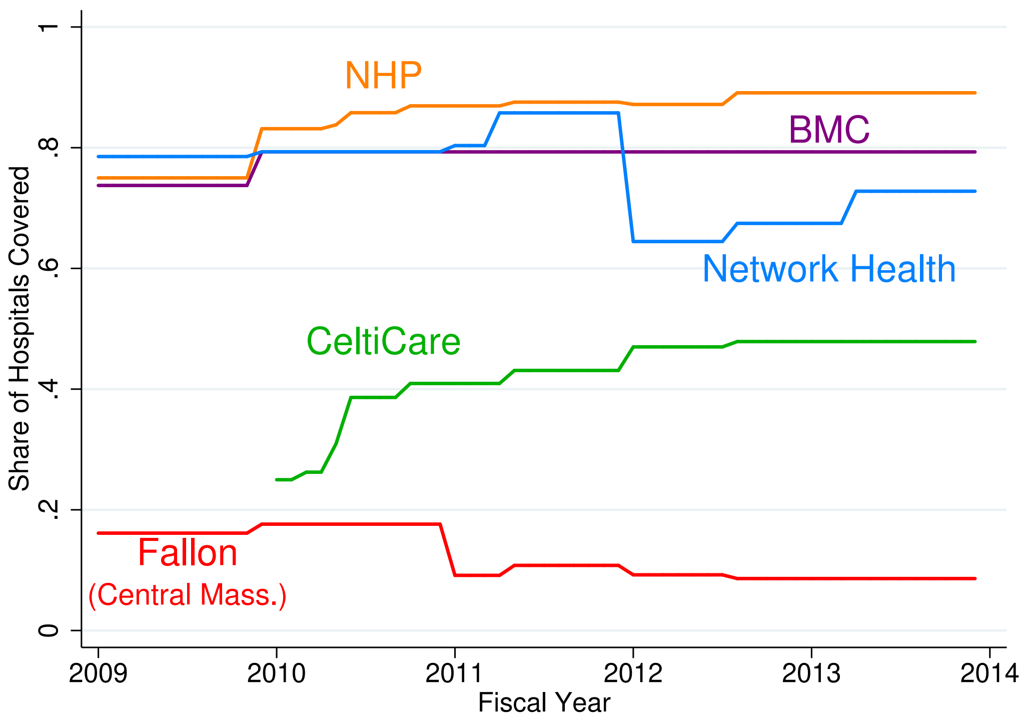
Torgovitsky, A. (2019): "Nonparametric inference on state dependence in unemployment," *Econometrica*, 87, 1475–1505.

Yan, J. (2013): "A Smoothed Maximum Score Estimator for Multinomial Discrete Choice Models," University of Wisconsin Working Paper.

Appendix: For Online Publication

A Plan Hospital Networks

Figure 6: Hospital Coverage in Massachusetts Exchange Plans



NOTE: The graph shows the shares of Massachusetts hospitals covered by each CommCare plan, where shares are weighted by hospital bed size in 2011. Fallon's hospital coverage share is much lower than other plans largely because it mainly operates in central Massachusetts and therefore does not have a statewide network. The large decline in Network Health's network size in 2012 reflects its dropping of the Partners Healthcare System and several other providers from its network.

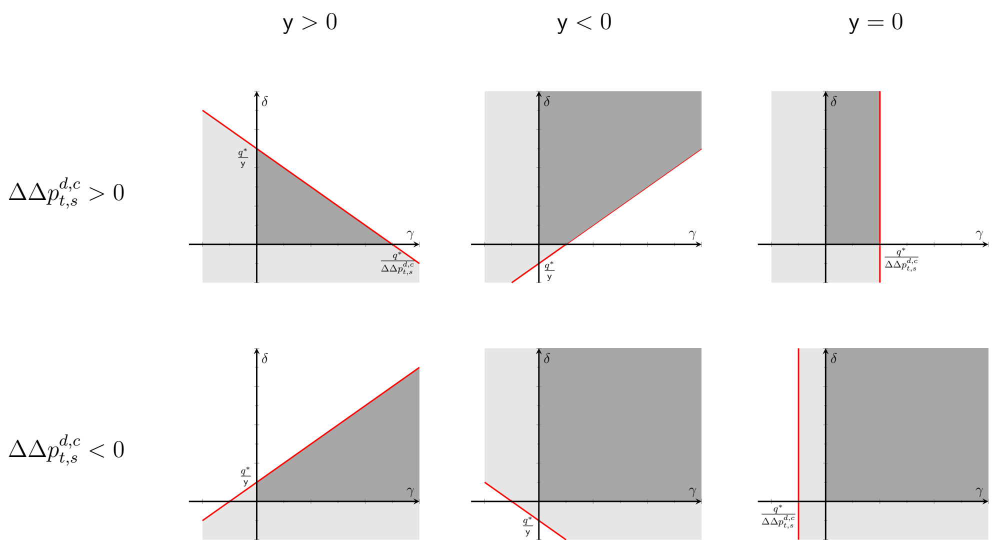
B Revealed Preferences Cases

Below we graphically display the information on the parameters (γ_0, δ_0) contained in the equation (4.3). Given y, define

$$q^* = \frac{(F^{d,c}_{t,s})^{-1}\left(1 - \sum_{r: sw^{d,c}(r, y_{i,s-1}) \geq \mathsf{y}} \Pr(y_{i,t} = d, y_{i,t-1} = r, y_{i,s} = c | z, y_{i,s-1})\right)}{\beta(x_i)}$$

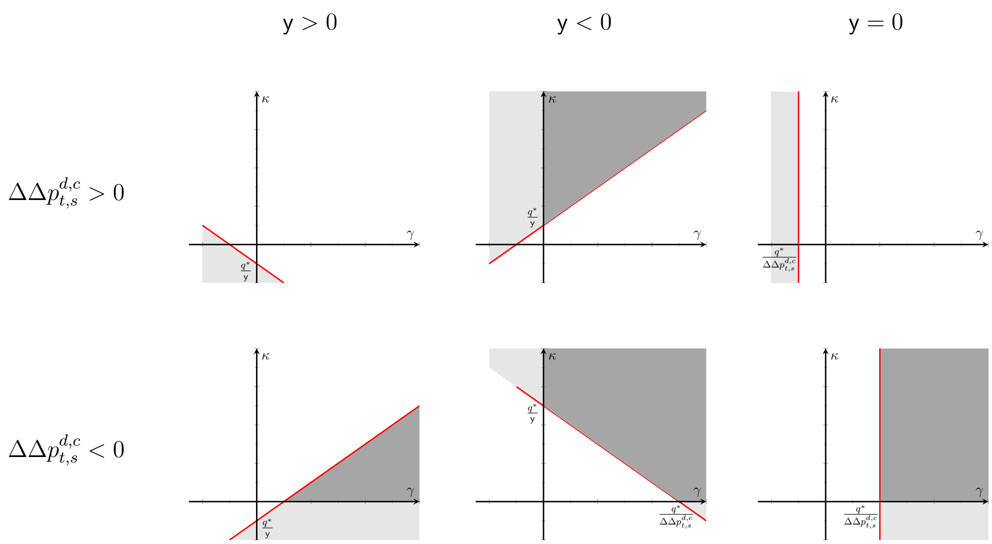
Figure 7 [8] considers the case where $q^* > 0$ [< 0]. If $F^{d,c}_{t,s}(0) = 0.5$, then the case $q^* > 0$ [< 0] corresponds to $\sum_{r: sw^{d,c}(r, y_{i,s-1}) \geq \mathsf{y}} \Pr(y_{i,t} = d, y_{i,t-1} = r, y_{i,s} = c | z, y_{i,s-1}) \ < \ 0.5$ [$>$ 0.5].

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.

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



C Proofs

Proof of Lemma 3.3:

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

$$\begin{aligned}&(-p_{c,i,s} - \mathbf{1}\{y_{i,s-1} \neq c\}\kappa_0)\,\beta_i - (-p_{d,i,s} - \mathbf{1}\{y_{i,s-1} \neq d\}\kappa_0)\,\beta_i + (\lambda_{c,i} - \lambda_{d,i})\\ &\quad \geq\ (-p_{c,i,t} - \mathbf{1}\{y_{i,t-1} \neq c\}\kappa_0)\,\beta_i - (-p_{d,i,t} - \mathbf{1}\{y_{i,t-1} \neq d\}\kappa_0)\,\beta_i + (\lambda_{c,i} - \lambda_{d,i})\end{aligned}$$

Hence,

$$\begin{aligned}&\Big\{\varepsilon_{i,s}\,\Big|\,\varepsilon_{d,i,s} \geq \max\nolimits_{c\notin D_0}\ (-p_{c,i,s} - \mathbf{1}\{y_{i,s-1} \neq c\}\kappa_0)\,\beta_i\\ &\qquad\qquad - (-p_{d,i,s} - \mathbf{1}\{y_{i,s-1} \neq d\}\kappa_0)\,\beta_i + (\lambda_{c,i} - \lambda_{d,i}) + \varepsilon_{c,i,s}]\Big\}\\ &\subseteq \Big\{\varepsilon_{i,t}\,\Big|\,\varepsilon_{d,i,t}\ \geq \max\nolimits_{c\notin D_0}\ \ [(-p_{c,i,t} - \mathbf{1}\{y_{i,t-1} \neq c\}\kappa_0)\,\beta_i\\ &\qquad\qquad - (-p_{d,i,t} - \mathbf{1}\{y_{i,t-1} \neq d\}\kappa_0)\,\beta_i + (\lambda_{c,i} - \lambda_{d,i}) + \varepsilon_{c,i,t}]\Big\}\\ &= \Big\{\varepsilon_{i,t}\,\Big|\ (-p_{d,i,t} - \mathbf{1}\{y_{i,t-1} \neq d\}\kappa_0)\,\beta_i + \lambda_{d,i} + \varepsilon_{d,i,t}\ \geq\\ &\qquad\qquad \max\nolimits_{c\notin D_0}\ \ [(-p_{c,i,t} - \mathbf{1}\{y_{i,t-1} \neq c\}\kappa_0)\,\beta_i + \lambda_{c,i} + \varepsilon_{c,i,t}]\Big\}\end{aligned}$$

So,

$$\begin{aligned}&\Pr(y_{i,t} \in D_0\,|\,p_i, y_{i,t-1}, \beta_i, \lambda_i)\\ &\quad=\ \Pr\Bigg(\bigcup_{d\in D_0}\Big\{\varepsilon_{i,t}\,\Big|\ (-p_{d,i,t} - \mathbf{1}\{y_{i,t-1} \neq d\}\kappa_0)\,\beta_i + \lambda_{d,i} + \varepsilon_{d,i,t}\\ &\qquad\qquad \geq\ \max_{c\notin D_0}\ [(-p_{c,i,t} - \mathbf{1}\{y_{i,t-1} \neq c\}\kappa_0)\,\beta_i + \lambda_{c,i} + \varepsilon_{c,i,t}]\Big\}\ \Bigg|\,\beta_i, \lambda_i\Bigg)\\ &\quad\geq\ \Pr\Bigg(\bigcup_{d\in D_0}\{\varepsilon_{i,s}\,\big|\ (-p_{d,i,s} - \mathbf{1}\{y_{i,s-1} \neq d\}\kappa_0)\,\beta_i + \lambda_{d,i} + \varepsilon_{d,i,s}\\ &\qquad\qquad \geq\ \max_{c\notin D_0}\ [(-p_{c,i,s} - \mathbf{1}\{y_{i,s-1} \neq c\}\kappa_0)\,\beta_i + \lambda_{c,i} + \varepsilon_{c,i,s}]\Big\}\ \Bigg|\,\beta_i, \lambda_i\Bigg)\\ &\quad=\ \Pr(y_{i,s} \in D_0\,|\,p_i, y_{i,s-1}, \beta_i, \lambda_i)\end{aligned}$$

In the second and third probabilities, the terms $p_{i,t}$, $p_{i,s}$, $y_{i,t-1}$, and $y_{i,s-1}$ denote the realized value of the price variable and lagged dependent variable from the conditioning statement.

⊠

Proof of Theorem 3.5:

(a) The supposition of Lemma 3.3 is satisfied for D_0 and $y_{i,t-1} = d'$ for any $d' \in D_0$. Hence,

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

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

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

(b) $s < t - 1$.

The supposition of Lemma 3.3 is satisfied for D_0 and $y_{i,t-1} = d'$ for any $d' \in D_1$. Hence,

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

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

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

⊠

Derivation of Equation (4.2):

$$
\begin{aligned}
A \equiv \Big\{ & (\varepsilon_{i,t}, \varepsilon_{i,s}) \,|\, (-p_{d,i,t} - \mathbf{1}\{y_{i,t-1} \neq d\}\kappa_0)\gamma_0\beta(x_i) + \lambda_{d,i} + \varepsilon_{d,i,t} \geq \\
& \qquad \max_{d' \neq d} \left(-p_{d',i,t} - \mathbf{1}\{y_{i,t-1} \neq d'\}\kappa_0\right)\gamma_0\beta(x_i) + \lambda_{d',i} + \varepsilon_{d',i,t}, \\
& (-p_{c,i,s} - \mathbf{1}\{y_{i,s-1} \neq c\}\kappa_0)\,\gamma_0\beta(x_i) + \lambda_{c,i} + \varepsilon_{c,i,s} \geq \\
& \qquad \max_{c' \neq c} \left(-p_{c',i,s} - \mathbf{1}\{y_{i,s-1} \neq c'\}\kappa_0\right)\gamma_0\beta(x_i) + \lambda_{c',i} + \varepsilon_{c',i,s} \Big\} \\
\subset \ \Big\{ & (\varepsilon_{i,t}, \varepsilon_{i,s}) \,|\, (-p_{d,i,t} - \mathbf{1}\{y_{i,t-1} \neq d\}\kappa_0)\,\gamma_0\beta(x_i) + \lambda_{d,i} + \varepsilon_{d,i,t} \geq \\
& \qquad (-p_{c,i,t} - \mathbf{1}\{y_{i,t-1} \neq c\}\kappa_0)\,\gamma_0\beta(x_i) + \lambda_{c,i} + \varepsilon_{c,i,t}, \\
& (-p_{c,i,s} - \mathbf{1}\{y_{i,s-1} \neq c\}\kappa_0)\,\gamma_0\beta(x_i) + \lambda_{c,i} + \varepsilon_{c,i,s} \\
& \qquad \geq (-p_{d,i,s} - \mathbf{1}\{y_{i,s-1} \neq d\}\kappa_0)\,\gamma_0\beta(x_i) + \lambda_{d,i} + \varepsilon_{d,i,s} \Big\} \\
\subset \ \Big\{ & (\varepsilon_{i,t}, \varepsilon_{i,s}) \,|\, \Delta\Delta\varepsilon^{d,c}_{i,t,s} \geq \left(\Delta\Delta p^{d,c}_{i,t,s} + \kappa_0 s w^{d,c}_{i,t,s})\right)\gamma_0\beta(x_i) \Big\},
\end{aligned}
$$

which implies

$$
\begin{aligned}
& \Pr(y_{i,t} = d, y_{i,s} = c \,|\, p_i, y_{i,s-1}, x_i, \lambda_i) \leq \Pr((\varepsilon_{i,t}, \varepsilon_{i,s}) \in A \,|\, p_i, y_{i,s-1}, x_i, \lambda_i) \\
& \quad \leq \ \Pr(\Delta\Delta\varepsilon^{d,c}_{i,t,s} \geq \left(\Delta\Delta p^{d,c}_{i,t,s} + \kappa_0 s w^{d,c}_{i,t,s})\right)\gamma_0\beta(x_i) \,|\, p_i, y_{i,s-1}, x_i, \lambda_i)
\end{aligned}
$$

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

Derivation of Equation (4.3):

Take $s < t-1$.

Define

$$
B^{d,c}_{t,s}(\mathsf{p}, \mathsf{y}, x_i) = \left\{ (\varepsilon_{i,t}, \varepsilon_{i,s}) \,|\, \Delta\Delta\varepsilon^{d,c}_{i,t,s} \geq (\mathsf{p}\gamma_0 + \mathsf{y}\delta_0)\beta(x_i) \right\}
$$

and note that since $\mathcal{F}^{d,c}_{t,s}$ denotes the c.d.f. of the conditional distribution of $\Delta\Delta\varepsilon^{d,c}_{i,t,s}$

$$
\Pr((\varepsilon_{i,t}, \varepsilon_{i,s}) \in B^{d,c}_{t,s}(\mathsf{p}, \mathsf{y}, x_i) \,|\, p_i, y_{i,s-1}, x_i) = 1 - \mathcal{F}^{d,c}_{t,s}((\mathsf{p}\gamma_0 + \mathsf{y}\delta_0)\beta(x_i))
$$

where p, y are constant values or functions of the conditioning set.

Using the same argument as above,

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

Note that if $\delta_0 \geq 0$ and $\mathsf{y}' \geq \mathsf{y}$, then $B^{d,c}_{t,s}(\mathsf{p}, \mathsf{y}', x_i) \subset B^{d,c}_{t,s}(\mathsf{p}, \mathsf{y}, x_i)$. It follows that for any r such that $sw^{d,c}(r, y_{i,s-1}) \geq \mathsf{y}$, $A(r) \subset B^{d,c}_{t,s}(\Delta\Delta p^{d,c}_{i,t,s}, \mathsf{y}, \beta_i)$. And so,

$$\bigcup_{r: sw^{d,c}(r, y_{i,s-1}) \geq \mathsf{y}} A(r) \subset B^{d,c}_{t,s}(\Delta\Delta p^{d,c}_{i,t,s}, \mathsf{y}, x_i).$$

Then,

$$\begin{aligned}
& \sum_{r: sw^{d,c}(r, y_{i,s-1}) \geq \mathsf{y}} \Pr(y_{i,t} = d, y_{t-1} = r, y_{i,s} = c \,|\, p_i, y_{i,s-1}, \beta_i, \lambda_i) \\
\leq \; & \sum_{r: sw^{d,c}(r, y_{i,s-1}) \geq \mathsf{y}} \Pr(\{(\varepsilon_{i,t}, \varepsilon_{i,s}) \in A(r)\} \cap \{y_{i,t-1} = r\} \,|\, p_i, y_{i,s-1}, \beta_i, \lambda_i) \\
\leq \; & \Pr\left((\varepsilon_{i,t}, \varepsilon_{i,s}) \in \bigcup_{r: sw^{d,c}(r, y_{i,s-1}) \geq \mathsf{y}} A(r) \,|\, p_i, y_{i,s-1}, \beta_i, \lambda_i \right) \\
\leq \; & \Pr((\varepsilon_{i,t}, \varepsilon_{i,s}) \in B^{d,c}_{t,s}(\Delta\Delta p^{d,c}_{i,t,s}, \mathsf{y}, \beta_i) \,|\, p_i, y_{i,s-1}, \beta_i, \lambda_i) \qquad \text{(C.1)}
\end{aligned}$$

The equation follows by integrating out both sides of the inequality with respect to the conditional distribution of λ_i. ⊠

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

$$e^{-\delta_0 \beta(x)} \leq \frac{\mathcal{M}_t(d, \lambda_i)}{\mathcal{M}_t(c, \lambda_i)} \leq e^{\delta_0 \beta(x)}$$

(b) Let $\mathcal{S}_{t,s}(d, c, \lambda_i) = \sum_{r \in \mathcal{D}_{t-1}, r' \in \mathcal{D}_{s+1}} \frac{\exp[-\delta_0(\mathbf{1}\{r \neq d\} + \mathbf{1}\{c \neq r'\})\beta(x) - \gamma_0 p_{r',i,s+1}\beta(x) + \lambda_{r',i})}{\mathcal{M}_t(r, \lambda_i)}$

$\cdot \Pr(y_{i,t-1} = r \,|\, p_i, y_{i,s+1} = r', x_i = x, \lambda_i)$. *For any* c, d,

$$e^{-2\delta_0\beta(x)} \;\leq\; \frac{\mathcal{S}_{t,s}(d,c,\lambda_i)}{\mathcal{S}_{t,s}(c,d,\lambda_i)} \;\leq\; e^{2\delta_0\beta(x)}$$

Proof of Lemma C.1:

(a) For any c, d, r, $-1 - \mathbf{1}\{c \neq r\} \;\leq\; -\mathbf{1}\{d \neq r\} \;\leq\; 1 - \mathbf{1}\{c \neq r\}$, and so

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

Result (a) follows.

(b) For any c, d, r, r',
$-2 - \mathbf{1}\{r \neq c\} - \mathbf{1}\{d \neq r'\} \;\leq\; -\mathbf{1}\{r \neq d\} - \mathbf{1}\{c \neq r'\} \;\leq\; 2 - \mathbf{1}\{r \neq c\} - \mathbf{1}\{d \neq r'\}$.
Hence,

$$\begin{aligned} & e^{-2\delta_0\beta(x)}\mathcal{S}_{t,s}(c,d,\lambda_i) \\ &= \textstyle\sum_{r\in\mathcal{D}_{t-1}, r'\in\mathcal{D}_{s+1}} \frac{e^{\delta_0(-2-\mathbf{1}\{r\neq c\}-\mathbf{1}\{d\neq r'\})\beta(x)} e^{-\gamma_0 p_{r',i,s+1}\beta(x)+\lambda_{r',i}}}{\mathcal{M}_t(r,\lambda_i)} \\ &\qquad \cdot \Pr(y_{i,t-1} = r \,|\, p_i, y_{i,s+1} = r', x_i = x, \lambda_i) \\ &\leq \textstyle\sum_{r\in\mathcal{D}_{t-1}, r'\in\mathcal{D}_{s+1}} \frac{e^{\delta_0(-\mathbf{1}\{r\neq d\}-\mathbf{1}\{c\neq r'\})\beta(x)} e^{-\gamma_0 p_{r',i,s+1}\beta(x)+\lambda_{r',i}}}{\mathcal{M}_t(r,\lambda_i)} \\ &\qquad \cdot \Pr(y_{i,t-1} = r \,|\, p_i, y_{i,s+1} = r', x_i = x, \lambda_i) \\ &= \mathcal{S}_{t,s}(d,c,\lambda_i) \\ &\leq \textstyle\sum_{r\in\mathcal{D}_{t-1}, r'\in\mathcal{D}_{s+1}} \frac{e^{\delta_0(2-\mathbf{1}\{r\neq c\}-\mathbf{1}\{d\neq r'\})\beta(x)} e^{-\gamma_0 p_{r',i,s+1}\beta(x)+\lambda_{r',i}}}{\mathcal{M}_t(r,\lambda_i)} \\ &\qquad \cdot \Pr(y_{i,t-1} = r \,|\, p_i, y_{i,s+1} = r', x_i = x, \lambda_i) \\ &\leq e^{2\delta_0\beta(x)}\mathcal{S}_{t,s}(c,d,\lambda_i) \end{aligned}$$

and (b) follows. ⊠

The following theorem extends the results in Theorem 4.2.

Theorem C.2. *Suppose Assumption 4.1 holds, and* $(d,c) \in \mathcal{D}_t \cap \mathcal{D}_s$*. Let* $\tau = \mathbf{1}\{r=c\}$*, and* $\Lambda = \begin{cases} 1 & \text{if } s = t-1 \text{ or } t-2 \\ 3 & \text{if } s < t-2 \end{cases}$
Then, for $s \leq t-1$ *and* $r \neq d$*,*

$$\Pr(y_{i,t}=c, y_{i,s}=d \,|\, p_i, y_{i,s-1}=r, x_i=x)\; e^{(\tau-\Lambda)\delta_0\beta(x)} e^{\gamma_0\left(\Delta\Delta p^{c,d}_{t,s}\right)\beta(x)}$$

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

$$\leq \Pr(y_{i,t}=c, y_{i,s}=d \,|\, p_i, y_{i,s-1}=r, x_i=x)\; e^{(\tau+\Lambda)\delta_0\beta(x)} e^{\gamma_0\left(\Delta\Delta p^{c,d}_{t,s}\right)\beta(x)}$$

Remark C.3. The case $r = d$ is implied by the case $r = c$.

Proof of Theorem C.2:

(i) $s = t-1$.

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

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

So,

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

By Lemma C.1,

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

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

(ii) $s=t-2$.

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

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

Hence,

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

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

(iii) $s < t-2$.

$$
\begin{aligned}
&\Pr(y_{i,t} = d, y_{i,s} = c \,|\, p_i, y_{i,s-1}, x_i = x, \lambda_i) \\
&\quad = \textstyle\sum_{r \in \mathcal{D}_{t-1}, r' \in \mathcal{D}_{s+1}} [\Pr(y_{i,t} = d \,|\, p_i, y_{i,t-1} = r, x_i = x, \lambda_i) \\
&\qquad\qquad \cdot \Pr(y_{i,t-1} = r \,|\, p_i, y_{i,s+1} = r', x_i = x, \lambda_i) \\
&\qquad\qquad \cdot \Pr(y_{i,s+1} = r' \,|\, p_i, y_{i,s} = c, x_i = x, \lambda_i) \\
&\qquad\qquad \cdot \Pr(y_{i,s} = c \,|\, p_i, y_{i,s-1}, x_i = x, \lambda_i)] \\
&\quad = \Bigg[\textstyle\sum_{r \in \mathcal{D}_{t-1}, r' \in \mathcal{D}_{s+1}} \frac{e^{-\delta_0(\mathbf{1}\{r \neq d\} + \mathbf{1}\{c \neq r'\})\beta(x)} e^{-\gamma_0 p_{r',i,s+1}\beta(x) + \lambda_{r',i}}}{\mathcal{M}_t(r, \lambda_i)} \\
&\qquad\qquad \cdot \Pr(y_{i,t-1} = r \,|\, p_i, y_{i,s+1} = r', x_i = x, \lambda_i) \Bigg] \\
&\qquad\qquad \cdot \frac{e^{-\gamma_0 p_{d,i,t}\beta(x)} e^{(-\gamma_0 p_{c,i,s} - \delta_0 \mathbf{1}\{y_{i,s-1} \neq c\})\beta(x)} e^{\lambda_{d,i} + \lambda_{c,i}}}{\mathcal{M}_{s+1}(c, \lambda_i) \mathcal{M}_s(y_{i,s-1}, \lambda_i)}
\end{aligned}
$$

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

Using the notation from Lemma C.1(b),

$$
\begin{aligned}
&\frac{\Pr(y_{i,t} = d, y_{i,s} = c \,|\, p_i, y_{i,s-1}, x_i = x, \lambda_i)}{\Pr(y_{i,t} = c, y_{i,s} = d \,|\, p_i, y_{i,s-1}, x_i = x, \lambda_i)} \\
&= \frac{e^{-\gamma_0(p_{d,i,t} + p_{c,i,s})\beta(x)} e^{-\kappa_0 \mathbf{1}\{c \neq y_{i,s-1}\}} \mathcal{M}_{s+1}(d, \lambda_i)\, \mathcal{S}_{t,s}(d, c, \lambda_i)}{e^{-\gamma_0(p_{c,i,t} + p_{d,i,s})\beta(x)} e^{-\kappa_0 \mathbf{1}\{d \neq y_{i,s-1}\}} \mathcal{M}_{s+1}(c, \lambda_i)\, \mathcal{S}_{t,s}(c, d, \lambda_i)} \\
&= \exp\left[-\gamma_0 \left(\Delta\Delta p_{t,s}^{d,c}\right) \beta(x)\right] \exp\left[-\delta_0 \left(\mathbf{1}\{y_{i,s-1} \neq c\} - \mathbf{1}\{y_{i,s-1} \neq d\}\right) \beta(x)\right] \\
&\quad \cdot \frac{\mathcal{M}_{s+1}(d, \lambda_i)\, \mathcal{S}_{t,s}(d, c, \lambda_i)}{\mathcal{M}_{s+1}(c, \lambda_i)\, \mathcal{S}_{t,s}(c, d, \lambda_i)}
\end{aligned}
$$

Now apply both parts of Lemma C.1 and integrate out λ_i. The result for $s < t-2$ follows. ⊠