Using Subjective Expectations Data to Allow for Unobserved Heterogeneity in Hotz-Miller Estimation Strategies*

Juan Pantano
Washington University
in St. Louis

Yu Zheng
City University
of Hong Kong

December 6, 2013

Abstract

We introduce a novel approach to allow for unobserved heterogeneity in two-step structural estimation strategies for discrete choice dynamic programming models (i.e strategies that avoid full solution methods). We contribute to the literature by adopting a *fixed effects* approach: rather than identifying an unobserved heterogeneity *distribution*, we actually reveal the true unobserved type of each observation in a first step. We do so by exploiting the tight link between the conditional choice probabilities that are derived from the economic model and subjective self-reported assessments about future choice probabilities such as those commonly elicited in major surveys. We uncover the unusual power of ideal expectations data to identify unobserved types for different classes of models. A single self-report suffices in models with a monotonicity property whereas two self-reports allow for identification in more general models. Of more empirical relevance, we show that our results hold when we allow these subjective future choice probabilities to be elicited in less than ideal circumstances, such as, for example, when self-reports display substantial "heaping" at focal reference values.

*We have benefited from valuable comments and suggestions by Dan Ackerberg, Victor Aguirregabiria, Peter Arcidiacono, Irene Brambilla, Moshe Buchinsky, Federico Bugni, Maria Casanova, Flavio Cuhna, Chris Flinn, Eric French, Ahu Gemici, Donna Gilleskie, Jin Hahn, Phil Haile, Han Hong, Joe Hotz, J.F. Houde, Clement Joubert, John Kennan, Maurizio Mazzoco, Rosa Matzkin, Costas Meghir, Pedro Mira, Bob Miller, Alvin Murphy, Salvador Navarro, Ariel Pakes, Seth Sanders, Matt Shum, Chris Taber, Melissa Tartari, Kevin Thom, Wilbert van der Klaauw, Matt Wiswall, Basit Zafar and participants at several conferences including the 2009 LACEA/LAMES Meetings in Buenos Aires, the 2010 MOOD Meetings in Rome, the 2010 World Congress of the Econometric Society in Shanghai, the 2012 SED Meetings in Cyprus, the UCLA proseminar in econometrics and seminars at Tulane, Duke, UW-Madison, Universidad de San Andres, UNC-Chapel Hill, Yale, NYU and University of Milan. All errors remain our own.

1 Introduction

Progress on structural estimation within applied microeconomics has been limited, given the difficulty of implementation in "frontal" or "full solution" strategies, i.e. strategies that solve the complicated optimization and/or equilibrium problem at each trial of the structural parameter vector in the estimation routine.[1] The work of Hotz and Miller (1993) shows how to estimate the structural parameters of a discrete choice dynamic programming model without solving the optimization problem even once. The Hotz-Miller strategy has generated some methodological work on estimation of structural models that builds upon this initial insight[2]. However, an inherent problem in the Hotz-Miller type of strategy exploited by these papers is that, because of its very own nature, it cannot accommodate permanent sources of unobserved heterogeneity.[3] The first step recovers equilibrium behavior policies from the data, and as such, these can only be recovered based on observables. On the other hand, the more computationally intensive "frontal strategies" can handle permanent unobserved heterogeneity by integrating out the unobserved types in the likelihood function using a finite mixture.[4]

Given its computational simplicity but its limitation regarding the handling of unobserved heterogeneity, in recent years there have been some efforts directed towards generalizing the Hotz-Miller approach to allow for unobserved heterogeneity.[5] In this paper we explore the potential use of expectations data

[1] Within the full solution paradigm, Rust (1987) and Keane and Wolpin (1994,1997) provided substantial computational savings that stimulated most of the empirical research to date with this type of models. See Keane and Wolpin (2009), Todd and Wolpin (2009), and Keane, Todd and Wolpin (2010) for surveys of a substantial number of applications using full solution methods in development, labor, consumer behavior and other fields in applied microceconomics. More recently, Su and Judd (2012) proposed a novel, promising approach (MPEC) to alleviate the computational burden associated with estimation by recasting the problem in a constrained optimization framework. See also Dube, Fox and Su (2012)

[2] See Hotz, Miller, Sanders and Smith (1994) to extend the original estimator to deal with the "Data Curse of Dimensionality" and for possible generalizations to allow for continuous choices and states. See Aguirregabiria and Mira (2002) for a recursive implementation of Hotz-Miller that improves small sample properties and for convergence to Full Information Maximum Likelihood. See Altug and Miller (1998) for a consistent account of aggregate shocks. See Jofre-Bonet and Pesendorfer (2003) for dynamic auctions. See Gayle and Golan (2011) for estimation of dynamic signaling models. See Bajari, Hong, Krainer and Nekipelov (2010) for similar ideas applied to estimation of static games. See Aguirregabiria and Mira (2007), Bajari, Benkard and Levin (2007), Pesendorfer and Schmidt-Dengler (2008) and Pakes, Ostrovsky and Berry (2008) for dynamic discrete games, Choo and Siow (2005) for the use of Hotz-Miller approaches in facilitating estimation of a dynamic two-sided matching game and Gayle, Golan and Soytas (2013) for estimation of Barro-Becker models

[3] This important limitation was noted early on by Eckstein and Wolpin (1989) among others.

[4] This is the approach taken by Wolpin (1984), van der Klaauw (1996), Keane and Wolpin (1997), Eckstein and Wolpin (1999) , Carro and Mira (2006), Mira (2007), Arcidiacono, Khwaja and Ouyang (2012) where some of the parameters are allowed to vary by type. Blau and Gilleskie (2008) and Liu, Mroz and van der Klaauw (2009) use a related, factor structure approach. Alternative approaches to handle unobserved heterogeneity, which still require DP solutions have been advanced by Ackerberg (1999, 2009) and Bajari, Fox, Kim and Ryan (2009). Whether discrete or continuous, parametric or non-parametric, all of the above are "random effects" approaches in the sense that only the probability of an observation being of a given type is contemplated.

[5] Buchinsky, Hahn and Hotz (2005) propose a clustering approach that is similar to ours in the sense of being essentially a fixed effects approach. Houde and Imai (2006) and Arcidiacono and Miller (2011) suggest alternative estimation strategies in a random effects context and allow for the unobserved heterogeneity to transition in systematic ways over time. Kasahara and Shimotsu (2008, 2009) and Hu and Shum (2009) focus on estimation and identification of related dynamic discrete choice models with unobserved types. Imai, Jain, and Ching (2009) and Norets (2009) provide Bayesian alternatives for estimation

such as, for example, subjective assessments of future choice probabilities to allow for estimable unobserved heterogeneity in these two-step estimation strategies for dynamic structural models.[6] We show that while requiring a particular type of data, our strategy can be an interesting alternative in the toolkit of microeconometricians if and when such data is available. In that sense, we think of our approach as complementary to the above literature. Our aim is to expand the toolkit that empirical researchers have when it comes to estimating dynamic structural models in computationally feasible ways. Our explicit use of elicited subjective expectations distinguishes our contribution from these other approaches taken in the literature. We will be focusing on single agent models, as the availability of expectations data seems more widespread in areas more amenable to single agent applications. However our idea can be applied to multiple agent contexts, in particular to dynamic discrete games. Indeed, much of the literature that builds upon the Hotz-Miller strategy to estimate dynamic games is now being generalized to allow for game and/or player level unobserved heterogeneity.[7]

We first characterize the power of expectations data to identify and estimate these models with the computational simplicity of a Hotz-Miller type of approach while, at the same time, allowing for unobserved heterogeneity, assuming that expectations are precisely elicited. For example, we first assume we have the ideal scenario in which there is no "heaping" or "focal measurement error" in Self-Reported Choice Probabilities (SR-CPs from now on).[8] Second, we show that when the use of more realistic, *focal*, subjective expectations data is contemplated in real applications, most of our results from the "ideal" case hold. Finally, we characterize how a modified version of our "linking technology" can alleviate some of the problems created by focal, reference point-based SR-CPs.

In addition to the theoretical insight, several datasets already include this kind of questions so our estimation strategy can be readily applied in a variety of settings. In the U.S. alone, all the major longitudinal surveys such NLSY or HRS include this type of questions. Looking ahead, however, the insights from our proposed estimation strategy are also informative about questionnaire design. In particular, about how these SR-CPs should be elicited to add the most value in a computationally feasible structural

---

of these models. See Aguirregabiria and Mira (2010) and Arcidiacono and Ellickson (2011) for reviews of this line of work.

[6] We focus our exposition on expectations about future choice probabilities because they are more widely available. Other questions may elicit probabilistic assessments about the future value of some state variables and could also be used to identify types with our method.

[7] See Aguirregabiria and Mira (2007) , Aguirregabiria, Mira and Roman (2007), Arcidiacono and Miller (2011), Siebert and Zulehner (2008), Hu and Shum (2008), Blevins (2009) and Kasahara and Shimotsu (2012). Aguirregabiria and Mira (2010) provide a comprehensive overview of structural estimation in the context of dynamic discrete choice models using full solution and non-full solution methods. Their review covers single agent and multiple agent models. See also Bajari, Hong and Nekipelov (2010) and Aguirregabiria and Nevo (2013) for more recent developments in the multiple agent literature.

[8] By "focal measurement error" we mean the systematic tendency of respondents to report round numbers (focal points) when assesing their future choice probabilities.

estimation strategy.[9]

Finally, it is worth mentioning that there exist two strands of literature on the use of expectations data that are somewhat, but not directly related to our work: a) Relaxing Rational Expectations. This is a strand of literature that uses expectations data in a more direct but still very important manner. The basic idea is to leverage data on expectations to be more flexible about the modelling of expectations. Key contributions here are Manski (2004) and Attanasio (2009). b) Using expectations data in estimation strategies for structural models that do not exploit the Hotz-Miller inversion. In this approach, like in ours, the expectations data are directly linked to the expectations used in the optimization problem. See Wolpin and Gonul (1985), Wolpin (1999), van der Klaauw and Wolpin (2008) and van der Klaauw (2011) for important contributions. In these cases, it is shown that these data are similar to revealed choice data and their use can provide more efficient estimators. These are important gains in estimator efficiency, but the contribution of such expectations data in those contexts is somewhat different than the one explored here.

The rest of the paper is organized as follows: The next section presents an extremely simple machine replacement example. We will use this example throughout the paper to fix ideas. Section 3 adds unobserved heterogeneity to the set up and discusses alternative conditions under which the use of expectations data succeeds in identifying such heterogeneity. Section 4 provides Montecarlo experiments that describe the performance of our estimation strategy. Conclusions follow.

2 Example: Estimating a Simple Dynamic Structural Model of Machine Replacement Decisions

Consider a simplified capital replacement problem similar to that in Rust (1987). Firms each use one machine to produce output in each period. These machines age, becoming more likely to breakdown, and in each time period the firms have the option of replacing the machines. Let x_t be the age of the machine at time t and let the current period profits from using a machine of age x_t be given by:

$$\Pi\left(x_t, d_t, \varepsilon_{0t}, \varepsilon_{1t}\right) = \begin{cases} \theta_1 x_t + \varepsilon_{0t} & \text{if } d_t = 0 \\ R + \varepsilon_{1t} & \text{if } d_t = 1 \end{cases} \tag{1}$$

where $d_t = 1$ if the firm decides to replace the machine at t, R is the net cost of a new machine, and

[9] Recently, scholars have begun eliciting subjective expectations in their own data collection efforts. See, among others, Delavande (2008), Zafar (2011a,2001b), Mahajan and Tarozzi (2011), Giustinelli (2012), Wiswall and Zafar (2012), Arcidiacono, Hotz and Kang (2012), Arcidiacono, Hotz, Maurel and Romano (2013) and Cuhna et al. (2013).

$(\varepsilon_{0t}, \varepsilon_{1t})$ are time specific shocks to the utilities/profits from replacing and not replacing. Let's assume that $(\varepsilon_{0t}, \varepsilon_{1t})$ are i.i.d. across replacement choices, firms and time periods, and, while not required for the implementation of our methods below, let's further assume that they follow a type I extreme value distribution. We consider a model with stochastic aging in which

$$x_{t+1} = \begin{cases} \min\{5, x_t + 1\} & \text{with probability } \pi_f & \text{if } d_t = 0 \\ x_t & \text{with probability } 1 - \pi_f & \text{if } d_t = 0 \\ 1 & \text{with probability } 1 & \text{if } d_t = 1 \end{cases} \tag{2}$$

Note that in this very simple model the state space only has 5 points and therefore full-solution methods can easily be used to estimate the model. We do this for illustrative purposes, but it should be kept in mind that the method we propose below can deal with more realistic state spaces in which standard full solution methods cannot be used or can only be used at substantial computational cost. Estimation is standard, and can proceed using either Rust (1987) nested fixed point algorithm or Hotz-Miller (1993) two-step estimator, among other alternatives. The Hotz-Miller strategy avoids the solution of the complicated dynamic structural model. The associated optimization problem is not solved even once. However, one is able to recover the structural parameters and can, after estimation, solve the model at those parameters if needed for, say, baseline simulation of artificial data and/or counterfactual policy experiments.

3 Adding Unobserved Heterogeneity

We now modify the machine replacement example to allow for heterogeneity in the structural parameters capturing age related maintenance costs θ_{1k} and machine replacement costs R_k. We consider the case of finite, time-invariant discrete types. We index types by $k = 1, ..., K$. In this setup, a standard estimation strategy would proceed by integrating out unobserved heterogeneity in the likelihood function, treating types as discrete random effects in the population. Alternatively, a modification of the Hotz-Miller strategy, exploiting subjective probabilities of future choices, can be used to estimate the structural parameters allowing for unobserved heterogeneity and without solving the dynamic program. In the remaining of this section we consider this possibility in detail.

Suppose we have available self-reported probabilities of next period machine replacement for each firm after the current period replacement decisions have been made.

Let

$$p_i^{SR}\left(d_{t'+1}=1|x_{t'},d_{t'},k\right) \tag{3}$$

denote the 1-period ahead self-reported probability of choosing $d=1$ (replacement choice) at time $t'+1$, elicited at t' from the technician in charge of machine maintenance at firm i, of unobserved type k, who, in addition is at the observed state $x_{t'}$ and who has recently made choice $d_{t'}$.Note that $(x_{t'},d_{t'})$ are the actual current state and decision of this agent at the time of the question (t'), and they are observed in the data. We do not require the agent to report one-period-ahead replacement $(d_{t'+1}=1)$ probabilities from the perspective of alternative, hypothetical states he could've been in and choices he could've made at time t'.

If the model in question featured deterministic transitions for its observed state variables, it would be clear which state point next period the SR-CP is giving choice information about. In models with stochastic transitions we need a more detailed "theory of self-report" that specifies what goes through the respondent's mind between the time she listens to the question and the time she provides the answer. Our theory of self report is the following: We assume the question is asked at time t after x_t has been realized and d_t has been chosen. Upon listening to the question "what's the probability that you will set $d_{t+1}=1$?" respondents use the solution to the dynamic programming problem to calculate the implied CCPs, $p\left(d_{t+1}=1|x_{t+1},k\right)$ at each feasible state next period, x_{t+1}. Note that there will be many probabilities, especially when the state space is large. After computing these, however, they need to provide a single answer. One reasonable way forward is to assume that respondents then report the average of these CCPs using the one-period ahead transition probability for the state variables, $f_x\left(x'|d,x\right)$ as weights. In other words, the question elicits the one-period-ahead "expected CCP". Formally,

$$\begin{aligned} \text{SR-CP} &= E\left[CCP|x,d,k\right] \qquad (4)\\ p^{SR}\left(d_{t+1}=1|x_t,d_t,k\right) &= E_{x_{t+1}|x_t,d_t,k}\left[\Pr\left(d_{t+1}=1|x_{t+1},k\right)\right] \\ &= \sum_{x_{t+1}}\Pr\left(d_{t+1}=1|x_{t+1},k\right)f_x\left(x_{t+1}|d_t,x_t,k\right) \end{aligned}$$

In some cases, depending on the specific wording, the question may elicit a different object, but in what follows, and unless noted otherwise, we assume that subjective expectation questions elicit the expected CCP.

Assumption SR-E[CCP]: The subjective probability questions elicit the expected CCP.

3.1 Estimation using Hotz-Miller with Precise Subjective Choice Probability Data

Throughout this section we assume that these probabilities are elicited with great precision. For future reference we establish this feature of the data in the following assumption

Assumption SR-Precise: The subjective probabilities are elicited with precision.

The basic intuition can be grasped in the context of our machine replacement example. Presumably if we have two firms A and B with machines in the same state in the current period $x_{At} = x_{Bt} = x_t$, *and these two firms make the same choice*, $d_{At} = d_{Bt} = d_t$, but report different probabilities of replacement tomorrow,

$$p^{SR}\left(d_{A,t+1} = 1 | x_{At}, d_{At}, k_A\right) \neq p^{SR}\left(d_{B,t+1} = 1 | x_{Bt}, d_{Bt}, k_B\right) \tag{5}$$

or

$$p^{SR}\left(d_{A,t+1} = 1 | x_t, d_t, k_A\right) \neq p^{SR}\left(d_{B,t+1} = 1 | x_t, d_t, k_B\right)$$

it must be the case that there is something unobserved by the econometrician but known by the technician in charge of machine maintenance in each firm that induces the difference in the self-reports. In other words, the unobserved state k is different for the two firms, $k_A \neq k_B$. Therefore, differences in self-reports are informative about underlying unobserved heterogeneity.

To be more specific, imagine two agents who are of the same type. They face the same prospects regarding their state variables next period. Moreover, they also face a common distribution of idiosyncratic error terms next period $f\left(\varepsilon_{t+1}\right)$. Hence, they will provide the same report about the probability of making the choice next period. However, observations that are of different unobserved types will report a different probability.[10]

This implies that under assumption SR-Precise, there is a one-to-one correspondence between SR and type at a particular state-choice combination. As a result, the number of types, K can be readily identified by counting the number of different $p^{SR}\left(d_{t+1} = 1 | x, d, k\right)$ elicited at a given state-choice cell, (x, d).[11] To simplify the exposition and without loss of generality let's assume there are only two types $k = 1, 2$. Then at any time t, the set of observations i with a common observable state x_t *and a common*

[10] In this world of precise elicitation, different types having the same 1-period ahead choice probability is a measure-zero event if the choice is feasible next period and the utility of the choice depends on the type.

[11] Note that it is important to consider future choice probability elicitation at particular state-choice combinations, not just particular states. The reason is that, among observations with the same state at t, x_t, those who make different choices will induce different probability distributions for the state variables next period, and then, even if they are of the same type, they will end up reporting different future choice probabilities. By focusing on those who are at the same state *and* make the same current choice, we avoid this problem.

current choice d_t must be either type $k = 1$ or type $k = 2$. We should then see two, and only two, different values of SR-CPs for each observed state-choice combination. Essentially, self-reported probabilities allow us to "reveal" type membership.

Of course, it is impossible for all agents to visit a common state and make the same decision at the time of the question, in which case assigning types would be immediate. Hence, we always identify types for sub-groups of agents who happen to be at the same state and have made the same decision at the time of the survey. We show in the next section that with two survey questions asked at two points in time, we are able to leverage on the transitivity of type memberships and consolidate information from both surveys in such a way so that the type membership of all agents can be identified.[12] We term the technology of integrating information of type memberships from different survey waves "the linking technology".

Then, after uncovering the unobserved type, estimation methods such as those proposed by Hotz and Miller (1993) or Hotz, Miller, Sanders and Smith (1994) apply directly, treating type as a time-invariant, observed discrete state. Moreover, for the purposes of identification, the model can be reduced to one without unobserved heterogeneity. Then identification results such as those in the work of Magnac and Thesmar (2002) apply directly.[13]

3.1.1 Linking Technology and Type Revelation

Suppose we have elicitations of subjective probabilities from the agents at two points in time. For example, this could be done by using the answers to a subjective assessment of the future choice probability in two different survey waves of a longitudinal study. If we have two (or more) self-reports for each individual, then we can identify a rich heterogeneity structure. In particular, we can link two observations who belong to the same type, even though they never reach a common state-choice combination at any time of a survey. More specifically, to establish a direct link between two observations, we need the self-reports from both of these observations to be elicited at a common state-choice cell and we need the self-reported probabilities to be the same. Call such two observations Alice and Bob. Apart from the elicitation that is used to link the two directly, Alice's and Bob's subjective probabilities are elicited again in a subsequent

[12] For example, if agent A is identified to have the same type as agent B from the first survey and agent B is identified to have the same type as agent C from the second survey, then A and C must have the same type even though A and C may not visit any common state-choice combination at either of the two surveys. All that is needed for A and B to be at a common state-choice cell at the time of the first survey and B and C to be at a common state-choice cell at the time of the second survey.

[13] Magnac and Thesmar (2002) do consider identification of a model with correlated fixed effects without relying on expectations data. However, the structure of unobserved heterogeneity they focus on is somewhat different than the one considered here. Kasahara and Shimotsu (2009) derive conditions on panel length and covariate heterogeneity required for identification of finite mixtures in dynamic discrete choice models without the use of subjective expectations data.

wave. In general, at the time of this subsequent elicitation, Alice and Bob will have arrived to two different state-choice combinations. Let's say Bob reaches a new state-choice combination which can be used to link him to Carl, since Carl happens to be at that state-choice combination and reports the same future probability of making a particular choice. Then Carl is identified to have the same type as Bob, and by transitivity he has the same type as Alice, even though Carl and Alice never reached a common state-choice combination during either elicitation. Mathematically, this is done by representing type membership as an equivalence relation and relying on the transitivity of the equivalence relation to group agents. Once a single agent is proven to relate to another agent, he is proven to relate to the entire equivalent class of the latter. We give the formal definitions of type revelation and linking technology below.

With the linking technology, we prove that under a weak assumption of absence of "isolated islands" in the space of the state-choice combinations, the "linking technology" can identify the type of every observation in the sample. The "isolated islands" are sets of state-choice combinations, (x, d) in which the pairs of self-reports of individuals are all contained and have no connections to other regions of the state-choice space. In other words, we can classify a group of agents whose self-reports are elicited at state-choice combinations within an isolated island into types, but there is no way to establish a link between groups across islands. As long as there is enough movement of the agents from one state-choice combination to another over the two questions, the No-islands assumption will hold.[14]

Finally, note that the "linking technology" with (at least) two self-reports places no restriction on the structure of heterogeneity in the data. In particular, the rank order of type-specific expected CCPs may differ across the state-choice space. Consider, for example, the possibility that a given type might have the highest expected future CCP elicited at some state-choice combination and the lowest at some other state-choice combination. We can link two observations of this type whose self-reports are elicited in these two state-choice combinations, even though it wouldn't be obvious how to do this by merely comparing the rankings of these reported probabilities.

Definition 1 (*Revelation of types*) *A revelation of types is defined by an equivalence relation $\sim$ on the set of observations $I = \{1, 2, ...N\}$. Call the cardinality of the quotient set $I/\sim$ the revealed number of types and denote it by M.*

By the Fundamental Theorem of Equivalence Relations, an equivalence relation $\sim$ on a set, partitions that set. Let the pair of self-reports be elicited at t' and t'' for all observations.

[14]Alternatively, observations in the isolated islands can be discarded provided that suitable assumptions about their representativeness hold.

Definition 2 (*Linking Technology*) *Define a binary relation, R, in the following way: $\forall i, j \in \{1, 2, ...N\}$,*

$$i \ R \ j$$

iff

$$\{\widetilde{p}_i^{SR}(x_{it'}, d_{it'}), \widetilde{p}_i^{SR}(x_{it''}, d_{it''})\} \cap \{\widetilde{p}_j^{SR}(x_{jt'}, d_{jt'}), \widetilde{p}_j^{SR}(x_{jt''}, d_{jt''})\} \neq \emptyset.$$

The linking technology is a relation $\sim$ on $\{1, 2, ..., N\}$: $\forall i, j \in \{1, 2, ...N\} = I$,

$$i \sim j$$

iff $\exists$ a subset of observations $\{i_1, i_2, ...i_n\} \subseteq I$, such that

$$i \ R \ i_1 \ R \ i_2 \ R ... R \ i_n \ R \ j.$$

The linking technology defines an equivalence relation. It is easily checked that $\sim$ satisfies reflexivity, symmetry and transitivity.

Assumption SR-No Islands: Define Σ^k to be the set of all state-choice cells at which a type k observation makes a self-report in the data. Then, $\forall (x, d), (x', d') \in \Sigma^k, \exists$ observations m and n of type k, with m reporting at (x, d), and n reporting at (x', d'), and $m \sim n$.

Lemma 1 *Under Assumption SR-Precise and SR-No Islands, the linking technology recovers the true number of types and type membership for each observation.*

Proof. *See Appendix A.* ■

While the above lemma is quite powerful, it is worth acknowledging its limitations. Even in the ideal world of precise elicitation, it still can happen that two observations (i, j) reporting the same expected CCP at a common state-choice combination in period t' end up reporting differently at another common state-choice combination in period t''. A potential reason, except for stochastic singularity, is that the true DGP features time-varying unobserved heterogeneity. To accommodate this, one would need to start thinking about modelling unobserved heterogeneity in a time-varying fashion, which is out of the scope of this paper and therefore a limitation of this approach. However, within the framework of this paper, we interpret the empirical appearance of this seemingly inconsistent reporting behavior as a form of measurement error. Indeed, in the next section, we move to the more empirically relevant case in which elicitations are no longer precise. In that world, whenever a pair of individuals report the same expected future CCP at time t' but different ones at t'', our modified linking technology under bunching will immediately "flag" this pair of individuals as belonging to different types. We will then rationalize the fact that they reported the same expected CCP at t' (despite being of different types) as a result of two different expected CCPs being rounded-off into a common value, thus creating bunching of the two different types.

3.2 Estimation using Hotz-Miller with "Focal" Subjective Choice Probability Data

Unfortunately, in many contexts the SR-CPs are not as clean as we assumed them to be in the previous section. While people may take more care in thinking about these probabilities when making actual choices, it is likely that they exercise less care when quickly computing these probabilities in a few seconds when answering to the interviewer.[15] In particular, there is likely to be substantial "heaping" at common reference points like 0, 0.10, 0.50, 0.90 and 1. See Walker (2003), Hill, Perry and Willis (2004) and Blass, Lach and Manski (2010). Surprisingly, there is no heaping at 0.33 and 0.66 which a priori appear to be good focal points when the probability reflects 1 out 3 or 2 out of 3 odds. Interestingly, respondents seem to be more precise when reporting probabilities close to the boundaries. For example, it is not uncommon to observe self-reports of 0.01, 0.02, 0.98 and 0.99. It is understandable that respondents care more about distinguishing 0 from 0.01 or 0.99 from 1 than 0.50 from 0.51 or 0.49. We accommodate these empirical regularities of probability self-reporting behavior in our discussion below.

Therefore in this section we characterize to what extent the results derived in the previous sections hold in the more realistic case in which Assumption SR-Precise does not hold. We will work with a set of H "focal" or "reference" values, h, that account for most of the self-reported probabilities. With a little abuse of notation, let H also denote the cardinality of the set H.

Focal SR-CPs may lead to "bunching" which may create uncertainty in the identification of the types. Say, for example, we have two observations of different types at the same (x, d). For simplicity, consider self-reports about 1-period ahead E[CCP] . Say under assumption SR-Precise type 1 reports 68% while type 2 reports 72%. Now, in a more realistic scenario in which SR-Precise no longer holds, we will have both types reporting 70%.

Following the notation in the previous sections, let $p_i^{SR}(d_{t+1} = 1 | x_{it}, d_{it}, k_i)$ be a self-reported choice probability that satisfies SR-Precise. In this section, we want to focus on the case in which the SR-CPs are probabilities that are rounded-off to the nearest focal point. We add an F to the self-report probability

[15] See Karni (2009) for a formalization of truthful elicitation of probabilities.

notation to emphasize it is now a focal self-report: $p_i^{SRF}(x_{it}, d_{it}, k_i)$. Formally,

$$p_i^{SRF}(x_{it}, d_{it}, k_i) = \arg\min_{h \in H} |p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i) - h| \tag{6}$$

where $p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i)$ is an Expected CCP. Actually, if $p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i) = E\left[CCP\right]$ we need to account for an additional layer of round-off in the underlying CCPs, which we then denote FCCPs:[16]

$$\begin{aligned} E\left[FCCP\right] &= \sum_{x_{i,t'+1}} \text{FCCP}\left(x_{i,t'+1}, k_i\right) f\left(x_{i,t'+1}|x_{it'}, d_{it'}, k_i\right) \\ &= \sum_{x_{i,t'+1}} \left[\arg\min_{h \in H} |\text{CCP}\left(x_{i,t'+1}, k_i\right) - h|\right] f\left(x_{i,t'+1}|x_{it'}, d_{it'}, k_i\right) \\ &= \sum_{x_{i,t'+1}} \left[\arg\min_{h \in H} |\Pr\left(d_{i,t'+1} = 1|x_{i,t'+1}, k_i\right) - h|\right] f\left(x_{i,t'+1}|x_{it'}, d_{it'}, k_i\right) \end{aligned} \tag{7}$$

We assume all observations follow this "rounding" procedure. Since k_i is unobserved, from the econometrician's point of view, the SRs can be associated with states and actions only: $\widetilde{p}_i^{SRF}(x_{it}, d_{it}) = p_i^{SRF}(x_{it}, d_{it}, k_i)$.

Definition 3 (Bunching) *Two self-reports are said to be bunched at (x, d) for **observations** i and j of different types, if $p_i^{SRF}(x, d, k_i) = p_j^{SRF}(x, d, k_j)$ and $k_i \neq k_j$. Two SRs are said to be bunched at (x, d) for **types** k and k', if $p^{SRF}(x, d, k) = p^{SRF}(x, d, k')$.*

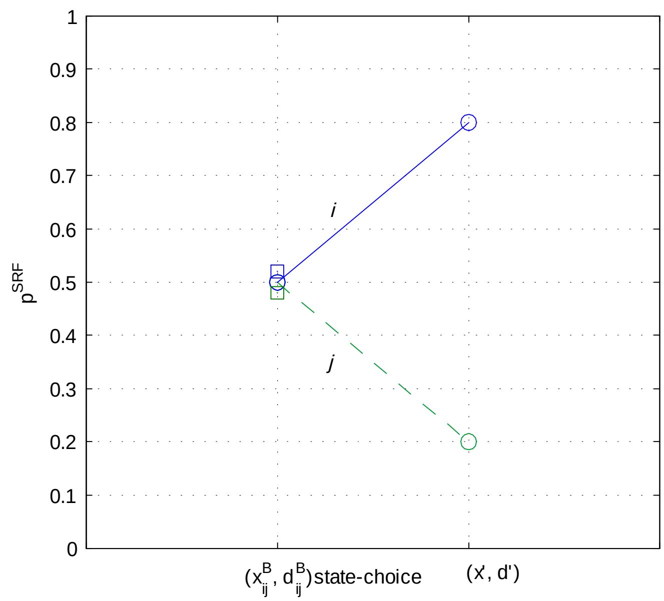
Note that "Bunching" is defined both, for observations and for types. When focal self-reports generate bunching in the data, some variation of our basic linking technology works under some additional assumptions.

Assumption B1 (Immediate Detection of Bunching Observations) If a pair of SRs by two observations i and j who belong to different types, bunch at the state-choice (x, d), then their other SRs must be elicited at another common state-choice (x', d'), at which the two types' focal SRs differ:

$$\widetilde{p}_i^{SRF}(x', d') \neq \widetilde{p}_j^{SRF}(x', d'). \tag{8}$$

[16] This additional layer of rounding off corresponds to the idea that an additional source of discrepancy between the theoretical $E\left[CCP\right]$ and the self-report resides in the respondent's inability to exactly compute the value function "off the top of her head". This inability induces computation of FCCPs, rather than CCPs at each feasible state point next period. Then, a second layer of rounding is introduced when the average of these rounded CCPs is itself rounded off when the answer is provided to the interviewer. Note that this assumption only introduces some limited rationality at the self-report stage. Behavior continues to be rational.

Figure 1: Illustration of Immediate Detection of Bunching Observations



Assumption B1 essentially makes sure that all bunchings of a pair of observations can be detected immediately. It will be relaxed later in the sense that we will not require immediate detection of bunching observations, but will require detection of bunching types.

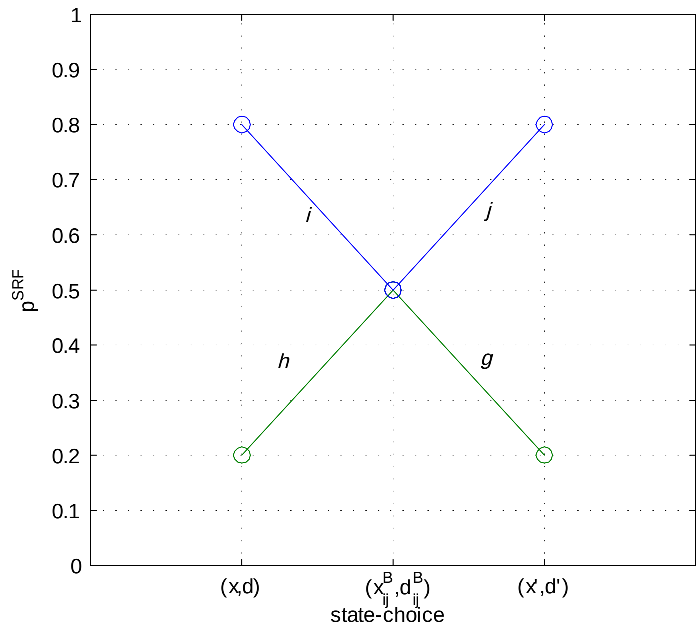
Definition 4 (Bunching state-choice for $\{i, j\}$) *The bunching state-choice for $\{i, j\}$ is the state-choice (x, d) at which their SRs bunch. Denote it by (x^B_{ij}, d^B_{ij}).*

Assumption B1 guarantees that whenever there are two observations i and j of different types reporting at a bunching state-choice, the bunching of different types is immediately detected. Assumption B1, implies two observations i and j can bunch at most at one state-choice cell.

In Figure 1, the squares mark the precise SRs, which are rounded-off to the nearest focal points, marked by circles. Whenever there is bunching of two different precise SRs, we include the square-marked precise SRs for illustrative purposes. As is evident from the figure, observations i and j have the same focal self-reports at the state-choice (x^B_{ij}, d^B_{ij}).

However, Assumption B1 is not enough to identify the types. Consider the following example in Figure 2. We can easily see that i and h are of different types. Same applied to j and g . But there is no way of telling whether the observations should be grouped as $\{j, i\}$ and $\{g, h\}$ or $\{i, g\}$ and $\{h, j\}$. In light of this, we make Assumption B2, which bridges the two SRs by the same type.

Figure 2: Problem Without Assumption B2



Assumption B2 (Bridging Bunchings) For all observations i and j who belong to the same type, but the singleton intersection of whose SRs is at (x^B_{ih}, d^B_{ih}) for some h, there exists another observation l of the same type as i and j, who has SRs in the two non-bunching state-choice cells.

Figure 3 illustrates how observation l bridges a bunching.

Definition 5 (Linking Technology under Bunching) *Define a binary relation, R^B, in the following way:* $\forall i, j \in \{1, 2, ...N\}$,

$$i \ R^B \ j$$

iff the following conditions are met:

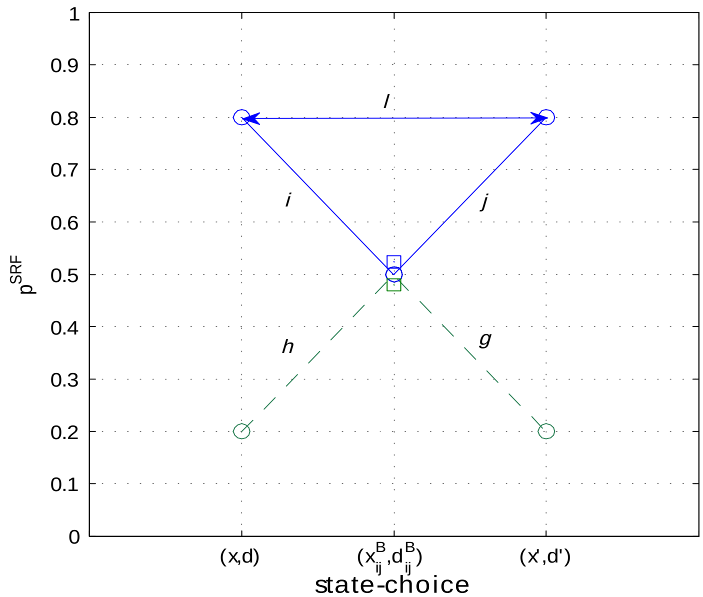
1. *The pairs of self reports for i and j are such that*

$$\{\widetilde{p}_i^{SRF}(x_{it'}, d_{it'}), \widetilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \cap \{\widetilde{p}_j^{SRF}(x_{jt'}, d_{jt'}), \widetilde{p}_j^{SRF}(x_{jt''}, d_{jt''})\} \neq \emptyset,$$

2. *if $\exists$ observation h,*

$$\{\widetilde{p}_i^{SRF}(x_{it'}, d_{it'}), \widetilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \cap \{\widetilde{p}_j^{SRF}(x_{jt'}, d_{jt'}), \widetilde{p}_j^{SRF}(x_{jt''}, d_{jt''})\} = \{\widetilde{p}_i^{SRF}(x^B_{ih}, d^B_{ih})\}$$

Figure 3: Illustration of Bridging the Bunching



then

$$\begin{aligned} & \exists l, \{\widetilde{p}_l^{SRF}(x_{lt'}, d_{lt'}), \widetilde{p}_l^{SRF}(x_{lt''}, d_{lt''})\} \\ = \quad & \{\widetilde{p}_i^{SRF}(x_{it'}, d_{it'}), \widetilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \triangle \{\widetilde{p}_j^{SRF}(x_{jt'}, d_{jt'}), \widetilde{p}_j^{SRF}(x_{jt''}, d_{jt''})\}, \end{aligned}$$

where $\triangle$ *denotes the set difference.*

The linking technology under bunching is a relation $\sim^B$ *on* $\{1, 2, ..., N\}$*:* $\forall i, j \in \{1, 2, ...N\} = I$,

$$i \sim^B j$$

iff $\exists$ *a subset of observations* $\{i_1, i_2, ...i_n\} \subseteq I$*, such that*

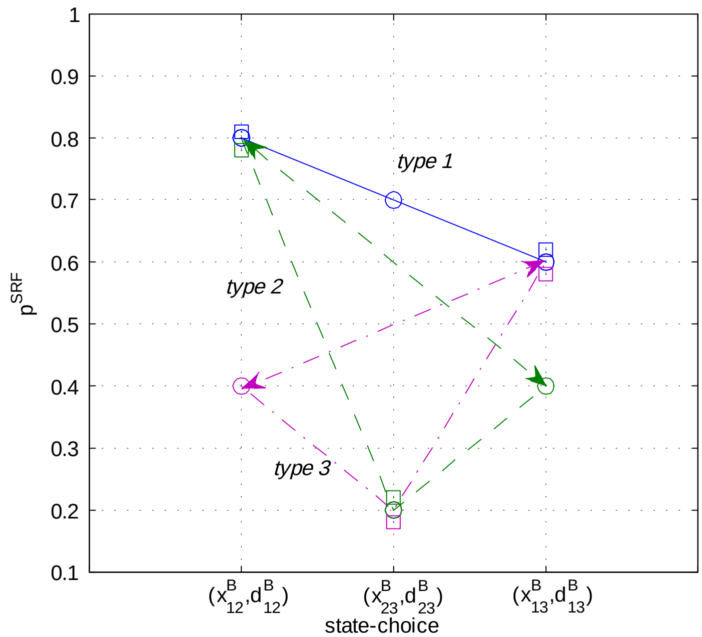
$$i \ R^B \ i_1 \ R^B \ i_2 \ R^B ... R^B \ i_n \ R^B \ j.$$

It can be easily proved that the linking technology under bunching also defines an equivalence relation.

Lemma 2 *Under Assumptions B1, B2 and SR-No Islands, the linking technology under bunching recovers the true types exactly.*

Proof. See Appendix A. ■

Figure 4: Identification of the Number of the Types



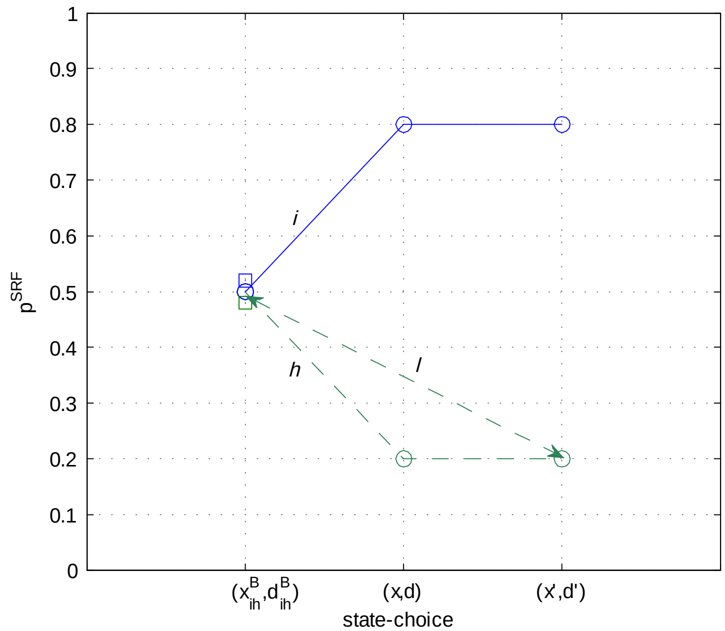
Note that now the number of types is only identified after the partition. In particular, it is not identified by counting the number of different SRs in each state-choice cell. Consider Figure 4. The partition identifies 3 types, though at each state-choice cell, there are only 2 different SRs. With a slight abuse of notation, $(x^B_{kk'}, d^B_{kk'})$ here denotes the bunching state-choice cell for type k and type k'. The arrows indicate "bridges".

Assumption B3 (Detection of Bunching Types) If two types, k and k', bunch at the state-choice (x, d), then $\exists$ two observations i of type k and j of type k', and another state-choice (x', d') s.t.

$$\begin{cases} \widetilde{p}^{SRF}_i(x,d) = \widetilde{p}^{SRF}_j(x,d) \\ \widetilde{p}^{SRF}_i(x',d') \neq \widetilde{p}^{SRF}_j(x',d') \end{cases}$$

Assumption B3 is weaker than Assumption B1. Assumption B1 ensures that whenever two observations of different types bunch, their other SRs reveal the bunching to the researcher. Assumption B3 only requires that whenever two types bunch, some observations' SRs reveal the bunching of the types. Figure 5 gives an example which satisfy Assumption B3 but not Assumption B1. Consider the observation l in the figure. Assumption B1 would require the existence of another observation linking $p^{SRF}(x^B_{ih}, d^B_{ih}) = 0.5$ and $p^{SRF}(x', d') = 0.8$ for immediate detection of bunching types. Nevertheless,

Figure 5: SRFs Allowed under Assumption B3 but not B1



Assumption B3 is satisfied as long as the observations i and h reveal the bunching of two types at (x_{ih}^B, d_{ih}^B).

With Assumption B3 replacing Assumption B1, the linking technology under bunching now cannot guarantee to recover the exact type of each observation. For example, types of i and h in Figure 6 are not distinguishable. Assumption B4 deals with this issue.

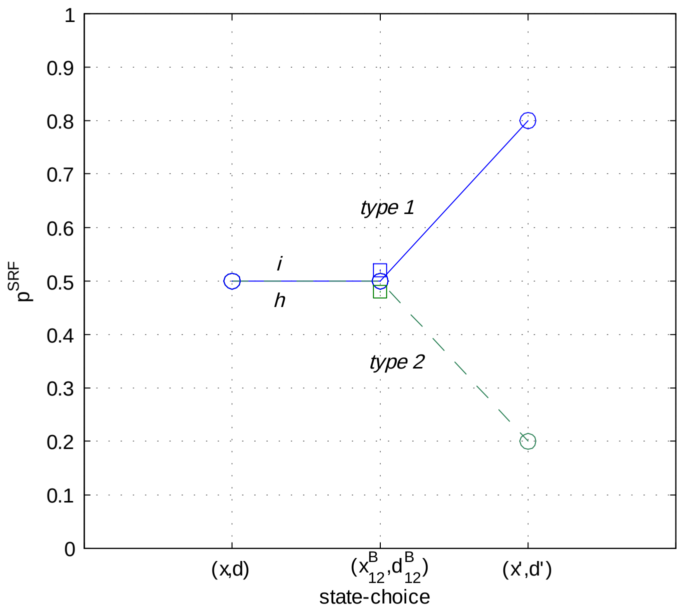
Assumption B4 (No observations with two "bunched" self-reports) Every observation i has at least one self-report elicited at a state-choice in which there is no bunching.

The following proposition establishes one of the most important results in this paper.

Proposition 1 *Under Assumptions B2,B3,B4 and SR-No Islands, the linking technology under bunching recovers the true types.*

Proof. Given Lemma 2, the critical step is to restore the identification of bunching state-choice cells under Assumption B3 (Detection of Bunching Types), which is weaker than Assumption B1 (Detection of Bunching Observations). Consider an observation i, whose SRs involve one SR in a bunching state (x, d) where her type bunches with another type. By Assumption B3, this bunching of types is detectable by two observations that do not necessarily involve i. There are thus two possibilities. One, $\exists$ an

Figure 6: Non-identification of Types



observation u, who bunches with i at (x_{iu}^B, d_{iu}^B), but differentiates itself at another state-choice (x', d'), as is depicted in Figure 7. Two, while i's other SR is at (x', d'), there are two observations u and v, who reveal the bunching of the types at some other state-choice (x'', d''), as is in Figure 8.

Now consider observations i and j, who are of the same type. We want to show that $i \sim^B j$. In the first case, by Lemma 2, we have $i \sim^B j$. In the second case, by Assumption B3, there exist two observations u and v that reveal the bunching of types at (x, d), that is, $(x, d) = (x_{uv}^B, d_{uv}^B)$. By Assumption B2, there exists some observation l that bridges j and v and there exists some other observation w that bridges v and i. The linking technology under bunching gives that $j \sim^B v$ and $v \sim^B i$. By the transitivity of the equivalence relation, $j \sim^B i$.

Now comes the other direction that $j \sim^B i$ implies $k_j = k_i$. It suffices to show that $\forall m, m\ R^B\ n$ implies $k_m = k_n$. Suppose not. Since $m\ R^B\ n$, let the common state-choice cell at which m and n made a common SR be (x, d). By Assumption B3, $\exists$ two observations m' and n' and another state-choice (x', d') s.t.

$$\begin{cases} \widetilde{p}_{m'}^{SRF}(x, d) = \widetilde{p}_{n'}^{SRF}(x, d) \\ \widetilde{p}_{m'}^{SRF}(x', d') \neq \widetilde{p}_{n'}^{SRF}(x', d') \end{cases}.$$

Assumption B2 identifies through bridging that $m' \sim^B m$ and $n' \sim^B n$. Hence, $m' \sim^B n'$. Contradiction.

However, for all pairs of observations whose SRs are identical at two of their bunching state-choice cells (observations ruled out in Assumption B4), their types are not identified. Recall Figure 6. Assumption

Figure 7: Immediate Detection of i-u Bunching

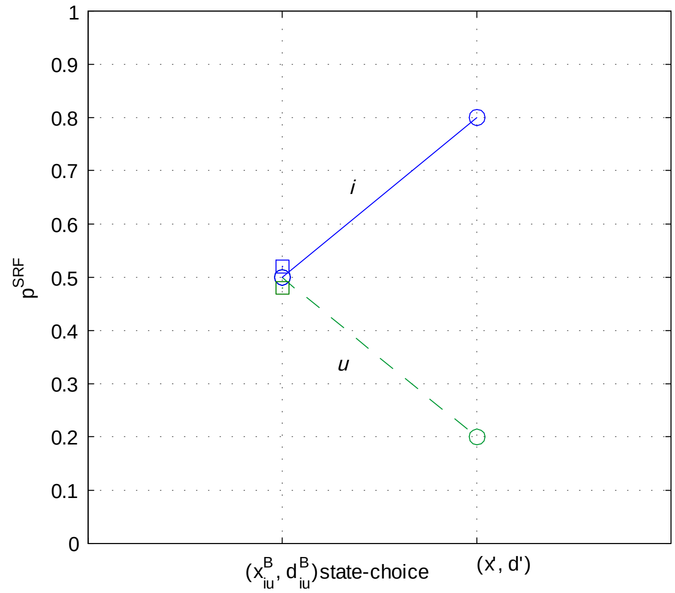
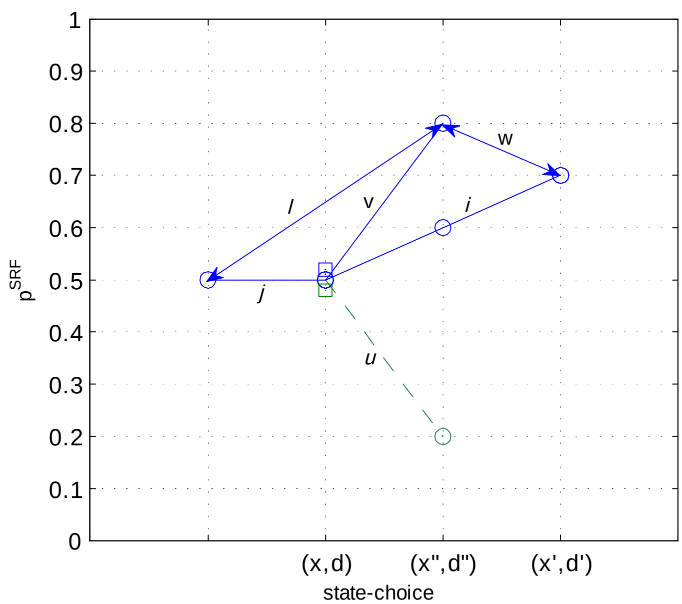


Figure 8: Detection of i-u Bunching Using v



B3 nevertheless indicates which two types these two observations may belong to. ■

In practice, we can write a computer algorithm that implements the linking technology to determine the type of those observations whose two SRs do not bunch with those of another type simultaneously.[17]

Finally, if we are willing to assume type-invariant transitions (i.e. $f(x'|d,x,k)$ does not depend on k) , we can relax Assumption B4. For those observations whose types are indeterminate, we will impute their types by finding the conditional probability of being a particular type given the observation's history of states and choices and its pair of bunching state-choices. Let i be such an observation of type k, whose SRs are $\{\widetilde{p}_i^{SRF}(x_{t'},d_{t'}),\widetilde{p}_i^{SRF}(x_{t''},d_{t''})\}$, where $(x_{t'},d_{t'})$ and $(x_{t''},d_{t''})$ are two bunching state-choice cells for types k and k'. Below we describe the procedure used to impute i's type.

First, we use the subsample where types can be correctly revealed to form a system of equations in terms of CCPs for each type and solve for the CCPs for each type. There are in general $|X| \times K$ equations and unknowns. Note that unlike the situation under SR-Precise, now even with 1-period ahead SRs the system will be non-linear. In the case of expected CCPs, the non-linearity is introduced by the double rounding-off. A typical equation of such a system will then look like

$$
\begin{aligned}
& p_i^{SRF}(x_{it'},d_{it'},k_i) && (9)\\
= \;& p_i^{SRF}\left(d_{i,t'+1}=1|x_{it'},d_{it'},k_i\right)\\
= \;& \arg\min_{h\in H}|p_i^{SR}(d_{t+1}=1|x_{it},d_{it},k_i)-h|\\
= \;& \arg\min_{h\in H}\left\{\left|\sum_{x_{i,t'+1}}\left[\mathrm{FCCP}\left(x_{i,t'+1},k_i\right)\right]f\left(x_{i,t'+1}|x_{it'},d_{it'}\right)-h\right|\right\}\\
= \;& \arg\min_{h\in H}\left\{\left|\sum_{x_{i,t'+1}}\left[\arg\min_{b\in B}|\Pr\left(d_{i,t'+1}=1|x_{i,t'+1},k_i\right)-h|\right]f\left(x_{i,t'+1}|x_{it'},d_{it'}\right)-h\right|\right\}
\end{aligned}
$$

Note that in general the above system may not have a unique solution. Therefore we work with an approximate problem that essentially disregards the two layers of rounding-off. Given the set of focal points H, the bias introduced by the approximation will be bounded.

Once we solve the above system, we compute the conditional probability of i being type k given i's history of choices and states for every "problematic" observation i (i.e. every observation whose pair of self-reports does not provide enough information to uncover its type).

[17] The algorithm is described in detail in a supplementary Appendix available upon request.

To that end note that for problematic observations we have

$$\begin{aligned}
\Pr\left(k \left| \{x_t, d_t\}_{t \neq t', t''}\right.\right) &= \frac{\Pr\left(\{x_t, d_t\}_{t \neq t', t''} |k\right) \Pr(k)}{\Pr\left(\{x_t, d_t\}_{t \neq t', t''}\right)} \\
&= \frac{\Pr\left(\{x_t, d_t\}_{t \neq t', t''} |k\right) \Pr(k)}{\sum_{k'=1}^{K} \Pr\left(\{x_t, d_t\}_{t \neq t', t''} |k'\right) \Pr(k')}
\end{aligned} \tag{10}$$

In the RHS, $\Pr\left(\{x_t, d_t\}_{t \neq t', t''} |k\right)$ can be computed using type k's CCPs and the estimates of the transition probabilities of the states. $\Pr(k)$ is estimated using, for example, the following equation

$$\begin{aligned}
\Pr(d_t = 1 | x_t = 5) &= \Pr(d_t = 1 | x_t = 5, k = 1) \Pr(k = 1) \\
&\quad + \Pr(d_t = 1 | x_t = 5, k = 2) [1 - \Pr(k = 1)]
\end{aligned} \tag{11}$$

where $\Pr(d_t = 1 | x_t = 5)$ is estimated by simple frequency from the data and $\Pr(d_t = 1 | x_t = 5, k = 1)$ and $\Pr(d_t = 1 | x_t = 5, k = 2)$ are computed using the type specific CCPs for each type. Given that obtaining such CCPs is not feasible, we work with approximate CCPs which solve the approximate system of equations described above.[18] Among all those problematic observations who have the same SRs and the same remaining history for $t \neq t', t''$ as $i's.$, we then assign their types such that with probability $p(k | \{x_t, d_t\}_{t \neq t', t''})$, they are of type k.[19]

3.3 Single Self-Report and The Monotonicity Property

An empirically relevant case is that in which the researcher only has one self-report about the agent's subjective probability of taking a given action in the future. In this case, an important class of models can still be identified. In particular, if the unobserved heterogeneity is characterized by a monotonicity property, one can link observations via a single self-report.

Assumption Monotonicity: the structure of unobserved heterogeneity is such that for unobserved types $k = 1, ..., K$,

$$\Pr(d = 1 | x, k = 1) < \Pr(d = 1 | x, k = 2) <, ..., < \Pr(d = 1 | x, k = K) \text{ for all } x \tag{12}$$

[18] Alternatively, the denominator in the RHS, $\Pr\left(\{x_t, d_t\}_{t \neq t', t''}\right)$, could be obtained by counting the proportion of observations who have this particular history of states and choices.

[19] This procedure can be readily extended to the case where there are more than two types bunching at the state-choice cells at the time of the SRs.

In words, the monotonicity assumption implies that types can be ranked according to their propensity to make choice $d = 1$ and this ranking holds across the entire state space. For example, in our machine replacement model, if types vary only according to replacement costs R, then monotonicity holds. It is easy to show that then the expected CCPs elicited at any state choice combination obey the same rank

$$E\left[CCP_{t+1}|x_t, d_t, k=1\right] < E\left[CCP_{t+1}|x_t, d_t, k=2\right] <, ..., < E\left[CCP_{t+1}|x_t, d_t, k=K\right] \text{ for all } (x_t, d_t) \tag{13}$$

Thus, under monotonicity and with precise elicitation, it suffices to group observations according to the rank of their self-reported future expected CCPs at the state-choice combination in which they were elicited. That is, regardless of the state-choice combination at which individuals were located when the question was asked, those who belong in a common type k, will report the k^{th} highest expected future CCP from the perspective of that particular state-choice combination. Of course if every individual is at the same state-choice when reporting the future choice probability then the grouping into types would be immediate. But note also that individuals could be at different state-choice combinations when reporting their subjective probability of a future choice. Under monotonicity, the rank of their reports is preserved across different state-choice combinations. Moreover, note that since $(\varepsilon_0, \varepsilon_1)$ are distributed in R^2, in a large enough sample, at least some observations from each type end up visiting every state so there are no rank reversals. That is, the ranking of an individual self-report would be the same no matter at which state-choice combination is being elicited.

For situations in which a single self-report is available and assumption SR-Precise does not hold we take a stand on the number of unobserved types, K, we can identify. Then we partition the unit interval into K ranges or segments and classify observations according to which segment their (rounded) self-reported probability belongs to. We must then make sure that for each individual the self-reported expected CCP is elicited at a state-choice combination in which each and every type is being "expressed". That is, each individual must report her expected CCP at a state-choice combination where at least one observation from every type is also reporting her own expected CPP. That way we guarantee that this individual's rank k is correctly assessed.

4 Montecarlo Experiments

In this section we do not discuss the precise data case because its empirical implementation is less feasible given that most subjective assessments of future choice probabilities have focal measurement error. We instead focus on the more realistic case in which there is focal measurement error in SR-CPs. We analyze two cases : a) a case in which this particular form of noise in the self-reports is innocuous and b) the more general case in which it leads to bunching. We analyze a general model without monotonicity.

Consider the model in the machine replacement example of Section 2. Again, note that we purposefully work with a simple toy model to be able to assess timing gains relative to a full-solution approach. However the method works equally well if we have a realistic state space that prevents estimation via full-solution. When the state space gets large it is likely that we will run into a "Data Curse of Dimensionality" in the sense that we will not have enough data to estimate the first-stage CCPs non-parametrically, even if we do not condition on type. This is not a limitation of our method, but one shared with the original Hotz-Miller (1993) estimator. However, there exist well known generalizations of the original Hotz-Miller strategy that preserve the initial insight while at the same time solving the "Data Curse of Dimensionality". For example, after unraveling the types we could use the estimator advanced by Hotz, Miller, Sanders and Smith (1994) that combines the alternative representation of the value function with a forward path simulation approach to greatly diminish the data requirements of the original Hotz-Miller strategy.[20] One can also work with models featuring stochastic finite dependence. As shown by Arcidiacono and Miller (2011) these models require a much smaller number of CCPs to be estimated.

We consider the simplest case in which there are $K = 2$ types. We simulate data on $N = 100,000$ firms and $T = 10$ periods using that model as underlying DGP with the following parameters:

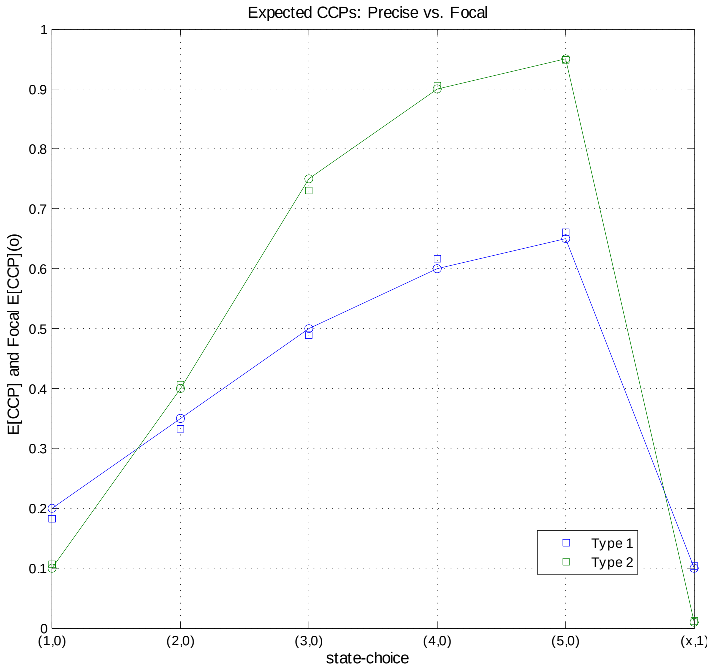
$$\text{Type 1:} \quad (\theta_{11}, R_1) = (-0.4, -3)$$
$$\text{Type 2:} \quad (\theta_{12}, R_2) = (-1.2, -7)$$

We compute each simulated firm's subjective assessment of its the expected CCP next period given its current state and choice using our theory of self-report, including the rounding off procedure that generates focal measurement error as described in Section 3.2. The simulated self-reports are taken in periods $t = 6$ and $t = 9$. In Figure 9 we see that despite the measurement error induced by focal self-reports, no type-bunching occurs. The squares point to the location of the precise E[CCP]s, the ones that would be elicited in the ideal case without "heaping" in focal values. The circles show the corresponding "focal" E[CCP]s

Since no type bunching occurs, the linking technology quickly establishes the number of types and type membership, and Hotz-Miller proceeds with type as an extra state variable. Table 1 describes the

[20] Only states visited with positive probability in the sample at hand (as opposed to all feasible states conceptually possible in the model) are used in this estimation strategy.

Figure 9: Focal Self-Reports That Do Not Lead to Bunching



results of the Montecarlo simulations and illustrates that our linking technology allows quick and precise estimation of the unobserved heterogeneity in the structural model.[21] The mean estimate over the R=500 repetitions is virtually the same as the truth. The standard deviation of the Montecarlo distribution is very small.[22]

Table 1

	Truth	Full Solution		Hotz-Miller	
		Mean	SD	Mean	SD
θ_{11}	-0.4	-0.4000	0.0058	-0.3999	0.0058
R_1	-3.0	-2.9997	0.0198	-2.9995	0.0198
θ_{12}	-1.2	-1.2012	0.0269	-1.2008	0.0268
R_2	-7.0	-7.0071	0.0951	-7.0057	0.0949
Avg. Time	-	11 minutes		30 seconds	

[21] Convergence of the entire algorithm takes on average aproximately half a minute. Almost all of the time is spent in the Hotz-Miller step. Indeed, preliminary type revelation and linking only takes about half a second. The montecarlo was run in a standard desktop using MATLAB.

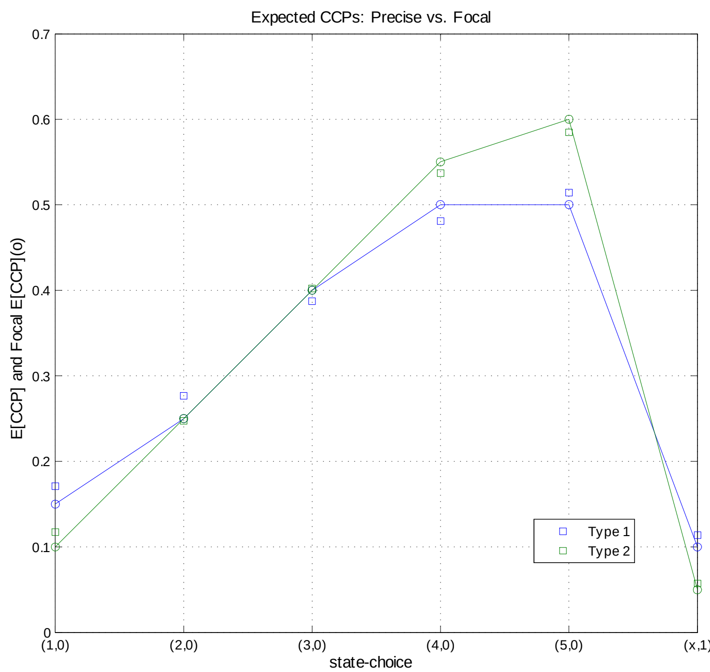
[22] Standard Deviations for the montecarlo distribution of estimates are computed for each parameter as follows: $\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\theta_r-\overline{\theta})^2}$ where $\overline{\theta}=\frac{1}{R}\sum_{r=1}^{R}\theta_r$

We now modify our DGP to generate a more complex situation. The parameters are now:

$$\text{Type 1:} \quad (\theta_{11}, R_1) = (-0.27, -2.65)$$
$$\text{Type 2:} \quad (\theta_{12}, R_2) = (-0.40, -3.75)$$

In Figure 10 we can see that focal self-reports now lead to bunching in state-choice combinations $(x, d) = (2, 0)$ and $(x, d) = (3, 0)$. Again, the squares point to the location of the precise E[CCP]s. The nearby circles show the corresponding "focal" E[CCP]s that respondents actually provide.

Figure 10: Focal Self-Reports That Lead to Bunching



We consider four estimation strategies for this case. In Table 2 we show the Montecarlo results for each of these.

1. **Naive:** In this strategy, we just drop from the sample those observations whose type cannot be determined. Column 2 shows the mean estimates. While the maintenance cost, θ_1 is estimated very precisely for both types, there is a small bias in the estimates of replacement costs R_1 and R_2. In both cases we tend to underestimate replacement costs. This makes sense. Since the two bunching state choice combinations $(2, 0)$ and $(3, 0)$ involve non-replacement decisions, when we discard observations we tend to disproportionately eliminate from the sample observations that

do not replace machines. Therefore the sample becomes more dominated by observations that do replace machines. The structural parameter estimates rationalize this behavior in the data by making machine replacement decisions less costly than they really are.

2. **Infeasible A:** In this case we pretend we know each observation's type. Then we estimate

$$p(k\,|\{(x_t, d_t)\}_{t=t',t''}, \{x_t, d_t\}_{t\neq t',t''})$$

by simple frequency and assign types to "problematic" observations such that they (as a group) are consistent with this estimated probability. Here we are back to the scenario of our first Montecarlo without bunching. Not surprisingly the performance is excellent.

3. **Infeasible B:** Here we no longer pretend we know each observation's type but instead claim we know the precise CCPs. Then we compute $p\left(k\,|\{(x_t, d_t)\}_{t=t',t''}, \{x_t, d_t\}_{t\neq t',t''}\right)$ using the Bayesian update described above and again assign types to "problematic" observations.[23] Again results are extremely good.

4. **Feasible:** Our feasible estimation strategy follows the same protocol as Infeasible B, but now using the approximate type-specific CCPs derived from the approximate system of equations based on focal E[CCP]s. The performance here is also excellent and virtually the same as the one achieved by Infeasible B, which uses the (usually unavailable) precise CCPs.

Table 2

	Truth	"Naive"		Infeasible A		Infeasible B		Feasible	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
θ_{11}	-0.27	-0.2735	0.0016	-0.2707	0.0016	-0.2713	0.0016	-0.2709	0.0015
R_1	-2.65	-2.6227	0.0088	-2.6538	0.0085	-2.6566	0.0085	-2.6561	0.0082
θ_{12}	-0.40	-0.4006	0.0024	-0.3985	0.0022	-0.3968	0.0022	-0.3977	0.0022
R_2	-3.75	-3.7024	0.0138	-3.7414	0.0134	-3.7304	0.0131	-3.7329	0.0134
Avg. Time	-	24.1 seconds		27.4 seconds		30.2 seconds		76.6 seconds	

[23] In the actual implementation there is a trade-off when choosing how much information to condition on when doing the Bayesian update. If we condition on all the history $\{x_t, d_t\}_{t\neq t',t''}$,the number of observation in each cell might be very small so in practice it might be better to condition on a subset of the available history.

5 Conclusions

We have introduced a new approach to allow for unobserved heterogeneity in two-step, CCP-based estimation strategies for discrete choice dynamic programming models such as those pioneered by Hotz and Miller (1993). Our strategy exploits the availability of expectations data. Since subjective expectations data about future choice probabilities integrate the future temporary idiosyncratic shocks, they are extremely powerful and they become a valuable resource to identify and estimate unobserved heterogeneity. We believe that if and when such data is available, our approach should be attractive given that identification requires mild assumptions and estimation can proceed with very light data requirements. Indeed, the method can be implemented with only one or two unconditional self-reports about expected future choice probabilities per respondent. Our Montecarlo experiments show that the computational burden is essentially the same as that of the (already fast) original Hotz-Miller estimator. The method can be applied in combination with variants of the original Hotz-Miller estimator that reduce its onerous data requirements in models with rich state spaces. Our focus has been on single agent models of dynamic discrete choice. Extensions might be possible to other contexts discussed in the Introduction, as long as subjective expectations data is available to supplement traditional data on observed choices and states.

Appendix A: Proofs of Lemmas

Lemma 1 Under Assumption SR-Precise and SR-No Islands, the linking technology recovers the true number of types and type membership for each observation.

Proof. First, we establish that under Assumption SR-Precise, the linking technology implies that for all $i \sim j$, $k_i = k_j$. By definition of the linking technology, $i \sim j$ iff $\exists$ a possibly empty subset of observations $\{i_1, ...i_n\} \subseteq I$, such that $i \; R \; i_1 \; R...R \; i_n \; R \; j$. By transitivity of " = ", it is enough to show that $\forall m \; R \; n$, $k_m = k_n$. By definition, $m \; R \; n$ iff

$$\{\widetilde{p}_m^{SR}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SR}(x_{mt"}, d_{mt"})\} \cap \{\widetilde{p}_n^{SR}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SR}(x_{nt"}, d_{nt"})\} \neq \emptyset$$

Since under Assumption SR-Precise the probability of two different types having exactly the same report on the same state-choice cell is zero, it holds that $m \; R \; n \Rightarrow k_m = k_n$. Hence, $k_i = k_{i_1} = ... = k_{i_n} = k_j$.

Second, we need to show that $\forall i, j$ with $k_i = k_j$, $i \sim j$. Suppose not. Consider observations i, j, who are of type k, but $i \nsim j$. Let $\Sigma^{[i]}$ ($\Sigma^{[j]}$) be the set of all state-choice cells at which the equivalent class $[i]$ ($[j]$) gives SRs. Then, $\Sigma^{[i]} \cap \Sigma^{[j]} = \emptyset$. Otherwise, $\exists (x, d) \in \Sigma^{[i]} \cap \Sigma^{[j]}$ and $\exists$ observations $i' \in [i], j' \in [j]$, who share a common SR at (x, d) and this implies that $i' \; R \; j'$. Hence, $i \sim i' \; R \; j' \sim j$, contradicting our assumption that $i \nsim j$. So $\Sigma^{[i]} \cap \Sigma^{[j]} = \emptyset$. But it further contradicts Assumption SR-No Islands, because $\Sigma^{[i]} \cap \Sigma^{[j]} = \emptyset$, together with $i \nsim j$, implies that $\forall (x, d) \in \Sigma^{[i]} \subseteq \Sigma^k, (x', d') \in \Sigma^{[j]} \subseteq \Sigma^k$, there does not exist two observations m and n, such that m gives a SR at (x, d) and n at (x', d') and $m \sim n$. ■

Lemma 2 Under Assumptions B1, B2 and SR-No Islands, the linking technology under bunching recovers the true types exactly.

Proof. First, we want to prove that under Assumption B1, the linking technology under bunching implies that for all $i \sim^B j$, $k_i = k_j$. Definition of linking under bunching gives $i \sim^B j$ iff $\exists$ a possibly empty subset of observations $\{i_1, ...i_n\} \subseteq I$, such that $i \; R^B \; i_1 \; R^B ... R^B \; i_n \; R^B \; j$. By transitivity of " = ", it is enough to show that $\forall m \; R^B \; n$, $k_m = k_n$. Consider such $m \; R^B \; n$. They must satisfy

1. $\{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \cap \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\} \neq \emptyset$,

2. if $\exists$ observation h,

$$\{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \cap \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\} = \{\widetilde{p}_m^{SRF}(x_{mh}^B, d_{mh}^B)\}$$

then

$$
\begin{aligned}
& \exists l, \{\widetilde{p}_l^{SRF}(x_{lt'}, d_{lt'}), \widetilde{p}_l^{SRF}(x_{lt''}, d_{lt''})\} \\
= \; & \{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \triangle \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\},
\end{aligned}
$$

where $\triangle$ denotes the set difference.

Proceed by contradiction. Suppose that $k_m \neq k_n$. Then by Assumption B1, m and n bunching at the state-choice cell of their common SR is immediately detected: $\{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \cap \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\} = \{\widetilde{p}_m^{SRF}(x_{mn}^B, d_{mn}^B)\}$. In this case, n qualifies as the observation h in the second condition, so

$$
\begin{aligned}
& \exists l, \{\widetilde{p}_l^{SRF}(x_{lt'}, d_{lt'}), \widetilde{p}_l^{SRF}(x_{lt''}, d_{lt''})\} \\
= \; & \{\widetilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \widetilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \triangle \{\widetilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \widetilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\},
\end{aligned}
$$

contradiction to the non-existence of such an observation who reports two different probabilities at one state-choice cell. Hence, $k_m = k_n$ and $k_i = k_{i_1} = k_{i_n} = k_j$.

Second, to show that for any pair of observations of the same type, they must belong to the same equivalence class, we proceed by contradiction. Consider two observations i and j of the same type k, but $i \nsim^B j$. Define $\Sigma^{[i]}, \Sigma^{[j]}$ as in the proof of Lemma 1. Now for any $(x, d) \in \Sigma^{[i]} \subseteq \Sigma^k$ and any $(x', d') \in \Sigma^{[j]} \subseteq \Sigma^k$, by Assumption SR-No Islands, $\exists$ two observations m and n of type k, with m reporting at (x, d) and n at (x', d'), and $m \sim^B n$. Since $(x, d) \in \Sigma^{[i]}$, $\exists$ observation $i' \in [i]$ who reports at (x, d) and some other $(x'', d'') \in \Sigma^{[i]}$. Obviously, i' and m share a common SR at (x, d). If (x, d) is a bunching state-choice cell, Assumption B1 immediately identifies this and Assumption B2 makes sure that $\exists$ an observation l that bridges i' and $m's$ non-bunching SRs. Linking technology under bunching implies $i' \sim^B m$. If (x, d) is not a bunching state-choice cell, the linking technology under bunching directly gives $i' \sim^B m$. By the same argument, $\exists j' \in [j]$ such that $j' \sim^B n$. Therefore, $i \sim^B i' \sim^B m \sim^B n \sim^B j' \sim^B j$. Therefore, $[i] \cap [j] = \{i', j', m, n\} \neq \emptyset$. A contradiction to the definition of equivalence class. ■

References

[1] Ackerberg, D. (1999) Importance Sampling and the Method of Simulated Moments, mimeo.

[2] Ackerberg, D. (2009) A New Use of Importance Sampling to Reduce Computational Burden In Simulation Estimation, Quantitative Marketing and Economics.

[3] Arcidiacono, P. (2005), Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?, Econometrica, Vol. 73, No. 5, September, 1477–1524

[4] Arcidiacono, P. and Ellickson, P. (2011) Practical Methods for Estimation of Dynamic Discrete Choice Models, Annual Reviews.

[5] Arcidiacono, P., Khwaja, A. and L. Ouyang (2012) Habit Persistence and Teen Sex: Could Increased Access to Contraception have Unintended Consequences for Teen Pregnancies?", Journal of Business and Economic Statistics, Vol. 30, No. 2 (April), 312-325

[6] Arcidiacono, P. and R. Miller (2011), Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity, Econometrica, Vol. 7, No. 6, November, 1823-1868.

[7] Arcidiacono, P., Hotz, V. J., Kang, S. (2012) "Modeling College Major Choice using Elicited Measures of Expectations and Counterfactuals", Journal of Econometrics, Vol. 166, No. 1 (January), 3-16

[8] Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. "Recovering Ex-Ante Returns and Preferences for Occupations using Subjective Expectations Data", working paper.

[9] Aguirregabiria, V. and Mira, P. (2002), Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models, Econometrica

[10] Aguirregabiria, V. and Mira, P. (2007), Sequential Estimation of Dynamic Discrete Games, Econometrica

[11] Aguirregabiria, V. and Mira, P. (2010), Dynamic Discrete Choice Structural Models: A Survey. Journal of Econometrics.

[12] Aguirregabiria, V. Mira and Roman (2007), An Estimable Dynamic Model of Entry, Exit, and Growth in Oligopoly Retail Markets, American Economic Review

[13] Aguirregabiria, V. and Nevo, A. (2012) “Recent Development in Empirical IO: Dynamic Demand and Dynamic Games” Advances in Economics and Econometrics: Theory and Applications. Tenth World Congress of the Econometric Society

[14] Altug, S. and R. Miller (1998), The Effect of Work Experience on Female Wages and Labor Supply, Review of Economic Studies

[15] Attanasio, O. P. (2009): Expectations and Perceptions in Developing Countries: Their Measurement and Their Use" American Economic Review, Papers and Proceedings.

[16] Bajari, P., Hong, H., Krainer, J. and D. Nekipelov (2009), Estimating Static Models of Strategic Interactions, Journal of Business & Economic Statistics, October, Vol. 28, No. 4

[17] Bajari, P., Hong, and D. Nekipelov (2010), Game Theory and Econometrics: A Survey of Some Recent Research, mimeo, University of Minnesota.

[18] Bajari, P., Benkard, L. and J. Levin (2007), Estimating Dynamic Models of Imperfect Competition, Econometrica

[19] Bajari, P., Fox, J., Kim, K. and Ryan, S. (2009), A Simple Nonparametric Estimator for the Distribution of Random Coefficients, mimeo.

[20] Blass, Lach and C. Manski (2010), "Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability", International Economic Review, Vol. 51, No. 2, May

[21] Blau, D. and D. Gilleskie (2008) The Role of Retiree Health Insurance in the Employment Behavior of Older Males. International Economic Review.

[22] Blevins, J. (2009) “Sequential MC Methods for Estimating Dynamic Microeconomic Models", Duke University, mimeo.

[23] Bresnahan, T., Stern, S. and M. Trajtenberg (1997), Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s, RAND

[24] Buchinsky, M., Hahn, J. and V. J. Hotz (2005), Cluster Analysis: A Tool for Preliminary Structural Analysis.

[25] Carro, J. and P. Mira (2006), A Dynamic Model of Contraceptive Choice of Spanish Couples. Journal of Applied Econometrics 21, 955-980.

[26] Choo, E. and Siow, A. (2005): Dynamic Two-Sided Matching Games, mimeo

[27] Cunha, F., Elo, I and Culhane, J. (2013) "Eliciting Maternal Expectations About the Technology of Skill Formation", NBER working paper 19144.

[28] Delavande, A. (2008), "Pill, Patch or Shot? Subjective Expectations and Birth Control Choice", International Economic Review, 2008, Vol. 49, 3.

[29] Dube, J.P. Fox, J. and C. Su (2012), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation" Econometrica, Vol. 80 (5), 2231-2267, 2012.

[30] Gayle, G. and Golan, L. (2011), Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968-1997, Review of Economic Studies, 1-41.

[31] Gayle, G., Golan L and Soytas, M. (2013), What Accounts for the Racial Gap in Time Allocation and Intergenerational Transmission of Human Capital ?, mimeo, Washington University in St. Louis.

[32] Giustinelli, P. (2012) "Group Decision Making with Uncertain Outcomes: Unpacking Child-Parent Choice of the High School Track", mimeo, University of Michigan.

[33] Eckstein, Z. and K. I. Wolpin (1989) "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey" Journal of Human Resources, Fall, 24, 4

[34] Eckstein, Z. and K. I. Wolpin (1999) "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities" Econometrica, 67(6): 1295-339

[35] Heckman, J., and B. Singer, (1984) A method for minimizing the impact of distributional assumptions in economic models for duration data. Econometrica 52, 271-320

[36] Hill, Perry and R. Willis (2004), Estimating Knightian Uncertainty from Survival Probability Questions on the HRS, mimeo

[37] Hotz, V. J. and R. Miller(1993), Conditional Choice Probabilities and the Estimation of Dynamic Models, Review of Economic Studies

[38] Hotz, V. J., Miller, R., Sanders, S. and J. Smith (1994), A Simulation Estimator for Dynamic Models of Discrete Choice, Review of Economic Studies

[39] Hu, Y. and M. Shum (2008) Identifying Dynamic Games with Serially-Correlated Unobservables, mimeo

[40] Hu, Y. and M. Shum (2009) Nonparametric Identification of Dynamic Models with Unobserved State Variables, mimeo

[41] Imai, S., N. Jain, and A. Ching (2009): "Bayesian Estimation of Dynamic Discrete Choice Models," Econometrica, 77(6), 1865-1899

[42] Jofre-Bonet and M. Pesendorfer (2003), Estimation of a Dynamic Auction Game, Econometrica

[43] Karni (2009) "A Mechanism for Eliciting Probabilities" Econometrica, Vol. 77, No. 2, March, 603–606

[44] Kasahara, H. and K. Shimotsu (2008), Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models, Journal of Econometrics, 146(1), 92-106

[45] Kasahara, H. and K. Shimotsu (2009), Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices, Econometrica, Vol. 77, No. 1 (January), 135–175

[46] Kasahara, H. and K. Shimotsu (2012), Sequential Estimation of Structural Models with a Fixed Point Constraint, Econometrica, Vol. 80, No. 5 (September), 2303–2319

[47] Keane, M. and K. I. Wolpin (1994) "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence" Review of Economics and Statistics

[48] Keane, M. and K. I. Wolpin (1997) "The Careers of Young Men" Journal of Political Economy

[49] Keane, M. and K. I. Wolpin (2009) "Empirical applications of discrete choice dynamic programming models" Review of Economic Dynamics

[50] Keane, M., Todd, P. and K. I. Wolpin (2011) "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications, Handbook of Labor Economics, Edited by O. Ashenfelter and D. Card, Volume 4, Part A, Chapter 4.

[51] Liu, Mroz, T. and W. van der Klaauw (2009), Maternal Employment, and Child Development, Journal of Econometrics, vol. 156(1), pages 212-228

[52] Magnac, T. and Thesmar (2002) "Identifying Dynamic Discrete Decision Processes", Econometrica, Vol. 70, No 2, March 801-816

[53] Mahajan, A. and Tarozzi, A. (2011), "Time Inconsistency, Expectations and Technology Adoption: The case of Insecticide Treated Nets", working paper

[54] McFadden, D. (1978) " Modelling the Choice of Residential Location", in A. Karlqvist et al., eds., Spatial Interaction: Theory and Planning Models. Amsterdam. North Holland Publishing.

[55] Mira, P. (2007) "Uncertain Infant Mortality, Learning and Life-Cycle Fertility". International Economic Review 48, 809-846

[56] Norets, A. (2009) “Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables”, Econometrica, 77(5), 1665-1682

[57] Pakes, A., Ostrovsky, M. and S. Berry (2008), Simple estimators for the parameters of discrete dynamic games (with entry/exit examples), RAND

[58] Pesendorfer, M. and P. Schmidt-Dengler (2008), Asymptotic Least Squares Estimators for Dynamic Games, Review of Economic Studies

[59] Rust, J. (1987), Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher, Econometrica

[60] Siebert, R., and C. Zulehner (2008): “The Impact of Market Demand and Innovation on Market Structure” Purdue University, mimeo.

[61] Su, C and Judd, K. (2012), "Constrained Optimization Approaches to Estimation of Structural Models", Econometrica, Vol. 80 (5), 2213-2230, September.

[62] Todd, P. and K. I. Wolpin (2009) "Structural Estimation and Policy Evaluation in Developing Countries" Annual Review of Economics, September 2010, Vol. 2

[63] van der Klaauw, W. (1996) Female Labor Supply and Marital Status Decisions: A Life Cycle Model, Review of Economics Studies, 1995, 63(2): 199-235.

[64] van der Klaauw, W. (2011) On the Use of Expectations Data in Estimating Structural Dynamic Choice Models, Journal of Labor Economics,

[65] van der Klaauw, W. and K. I. Wolpin (2008) Social Security and the Retirement and Savings Behavior of Low Income Households, Journal of Econometrics

[66] Walker, J. (2003), "Pregnancy and Fertility Expectations: Estimates of Bounded Rationality and Unintended Births", mimeo.

[67] Wiswall, M. and Zafar, B. "Determinants of College Major Choice: Identification Using an Information Experiment" Available at SSRN: http://ssrn.com/abstract=1919670.

[68] Wolpin, K. I. (1984), An Estimable Dynamic Stochastic Model of Fertility and Child Mortality, Journal of Political Economy,1984, Vol. 92: 852-875

[69] Wolpin, K. I. (1999), Commentary on "Analysis of Choice Expectations in Incomplete Scenarios", by C.F. Manski" Journal of Risk and Uncertainty

[70] Wolpin, K. I. and F. Gonul (1985) "On The Use of Expectations Data in Micro Surveys: The Case of Retirement", Report to the Employment and Training Administration, U. S. Department of Labor, Washington, DC, March.

[71] Zafar, B. (2011a) "Can Subjective Expectations Data be used in Choice Models? Evidence on Cognitive Biases", Journal of Applied Econometrics, Vol. 26, Issue 3, pp 520-544.

[72] Zafar, B. (2011b) "How do College Students Form Expectations? Journal of Labor Economics, Vol. 29, No. 2, pp. 301-348.