

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

Juan Pantano

Washington University
in St. Louis

Yu Zheng

Washington University
in St. Louis

This Version: September 20, 2010

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 just two 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. 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 participants at several conferences and workshops including Tulane University (2009), LACEA/LAMES Meetings in Buenos Aires (2009), Duke University (2009), UCLA (2010), University of Wisconsin-Madison (2010), the MOOD Meetings in Rome (2010), Universidad de San Andres (2010) and the World Congress of the Econometric Society in Shanghai (2010). 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 & 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 insight². 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

¹Within the full solution paradigm, Rust (1987) and Keane & Wolpin (1994,1997) provided substantial computational savings that stimulated most of the empirical research to date with these type of models. See Keane & Wolpin (2009), Todd & Wolpin (2009), and Keane, Todd & Wolpin (2010) for surveys of a substantial number of applications using full solution methods in development, labor, consumer behavior and other fields in applied microeconomics. More recently, Su & Judd (2007) proposed a novel, promising approach (MPEC) to further alleviate the computational burden associated with estimation via full solution methods by recasting the problem in a constrained optimization framework. See also Fox, Su & Dube (2009)

²See Hotz, Miller, Sanders & 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 & Mira (2002) for a recursive implementation of Hotz-Miller that improves small sample properties and for convergence to Full Information Maximum Likelihood. See Altug & Miller (1998) for a consistent account of aggregate shocks. See Jofre-Bonet & Pesendorfer (2003) for dynamic auctions. See Golan & Levy-Gayle (2008) for estimation of dynamic signaling models. See Bajari, Hong, Krainer & Nekipelov (2009) for similar ideas applied to estimation of static games. See Aguirregabiria & Mira (2007), Bajari, Benkard & Levin (2007), Pesendorfer & Schmidt-Dengler (2008) and Pakes, Ostrovsky & Berry (2008) for dynamic discrete games and Choo & Siow (2005) for the use of Hotz-Miller approaches in facilitating estimation of a dynamic two-sided matching game.

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

function.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 expectation data 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 structural 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.⁷ However our idea can be

⁴This is the so-called Heckman & Singer (1984) approach taken by Wolpin (1984), van der Klaauw (1996), Keane and Wolpin (1997), Eckstein and Wolpin (1999), Carro & Mira (2006), Mira (2007), Arcidiacono, Khwaja and Ouyang (2007), Blau & Gilleskie (2008), Liu, Mroz & van der Klaauw (2009), among many others. Alternative approaches to handle unobserved heterogeneity, which still require DP solutions have been advanced by Akerberg (1999,2009) and Bajari, Fox, Kim & 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.

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

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

⁷Casanova et al (2010) apply the methods in this paper to allow for unobserved heterogeneity in a two-stage, CCP-based estimation of a dynamic programming model of retirement.

applied to multiple agent contexts, in particular to dynamic discrete games. Indeed, much of the literature that built upon the Hotz-Miller strategy to estimate dynamic games is now being generalized to allow for game and/or player level unobserved heterogeneity.^{8,9}

We first characterize the power of expectation 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).¹⁰ Second, we show that when the use of more realistic, *focal*, subjective expectation 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, if we have more than one SR-CP available.¹¹

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 these 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 estimation strategy.

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 expectation data in more

⁸Aguirregabiria & Mira (2007) , Aguirregabiria, Mira & Roman (2007), Arcidiacono & Miller (2008), Siebert and Zulehner (2008), Hu & Shum (2008). Blevins (2009)

⁹Aguirregabiria & Mira (2009) 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

¹⁰By "focal measurement error" we mean the systematic tendency of respondents to report round numbers (focal points) when assessing their future choice probabilities.

¹¹Throughout this paper we allow for a specific form of (lack of) precision in the elicited subjective expectations. Allowing for more flexible forms of self-reporting error in is certainly important. In principle our framework could be generalized to allow for more flexible forms of measurement error in self reports but such generalization is beyond the scope of this paper.

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 expectation data are directly linked to the expectations used in the optimization problem. See Wolpin & Gonul (1985), van der Klaauw (2000), Wolpin (1999) and van der Klaauw & Wolpin (2008) 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 expectation 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 expectation data succeeds in identifying such heterogeneity. Section 4 provides Montecarlo experiments that describe the performance of our estimation strategy. Section 5 discusses some extensions for our framework. 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 expected current period profits from using a machine of age x_t be given by:

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

where $d_t = 1$ if the firm decides to replace the machine at t , R is the net cost of a new machine, and the ε_{ts} are time specific shocks to the utilities/profits from replacing and not

replacing. Let's assume that these ε_{ts} are i.i.d. across 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

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

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 first consider the case of finite discrete types. We then analyze the continuous case.

In the discrete case we index types by $k = 1, \dots, K$. An alternative set up considers the existence of unobserved state variables $x^u = \{x_1^u, x_2^u, \dots\}$ or, alternatively, a single unobserved discrete state variable $k_t \in \{1, \dots, K\}$ that captures every possible combination of unobserved states in x^u . We allow for the possibility that unobserved states may transition over time and allow for this transition to potentially depend on the choice d_t . Then, in general, we consider a state transition given by

$$f_{xk}(x', k' | d, x, k)$$

We can entertain several assumptions that restrict the generality of $f_{xk}(x', k'|d, x, k)$

- **Assumption F1((\mathbf{x}, \mathbf{k}) are conditionally independent):** conditional on (d, x, k) , x' and k' are independently distributed:

$$f_{xk}(x', k'|d, x, k) = f_k(k'|d, x, k) f_x(x'|d, x, k)$$

Similar to Arcidiacono & Miller (2008), we can also assume that

- **Assumption F2 (Exogenous Transitions for Unobserved States):** the transition of the unobserved state variables does not depend on the current choice nor the current observed state, but follows an exogenous and flexible markov stochastic process:

$$f_k(k'|d, x, k) = f_k(k'|k) = \pi_{kk'}$$

As in much of the literature using full-solution methods, in some situations we can further assume

- **Assumption F3 (Time Invariant Unobserved Heterogeneity):** the unobserved states are time invariant.

$$\pi_{kk} = 1 \text{ for all } k_t \in \{1, \dots, K\}$$

In some cases we can further assume that

- **Assumption F4 (Homogeneous Transitions for Observed State Variables):** the evolution of the observed states, x , does not depend on the unobserved heterogeneity, k .

$$f_x(x'|d, x, k) = f_x(x'|d, x) \text{ for all } k \in \{1, \dots, K\}$$

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.

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

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

$$p_i^{SR}(d_{t'+1} = 1 | x_{t'}, d_{t'}, k)$$

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 $d_{t'}$. 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. In particular, self-reports are not rounded off to the nearest "focal" probability.

A key question is then: under what conditions can we use these expectation data to reveal the underlying unobserved heterogeneity? 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,

$$\begin{aligned} p^{SR}(d_{A,t+1} = 1 | x_{At}, d_{At}, k_A) &\neq p^{SR}(d_{B,t+1} = 1 | x_{Bt}, d_{Bt}, k_B) \\ p^{SR}(d_{A,t+1} = 1 | x_t, d_t, k_A) &\neq p^{SR}(d_{B,t+1} = 1 | x_t, d_t, k_B) \end{aligned}$$

it must be the case that there is something unobserved by the econometrician but observed 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. In particular, note that if there are only two types, there can only be two

and only two different $p^{SR}(d_{t+1} = 1|x_t, d_t, k)$ reported by observations that have the same state-choice combination (x_t, d_t) .

It follows that the number of types, K can be readily identified by counting the number of different $p^{SR}(d_{t+1} = 1|x, d, k)$ elicited at each state-choice cell, (x, d) .¹² Can we use the self-reported probabilities to estimate the machine replacement problem a la Hotz-Miller but allowing for unobserved heterogeneity? The answer is yes. Below we provide details on how to do so.

3.1.1 Linking Technology and Type Revelation with Precise Self Reports

We now introduce our "linking technology". The basic idea is pretty simple and illustrates the power of eliciting self-reported choice probabilities to recover the underlying unobserved heterogeneity.

Let's assume there are only two types $k = 1, 2$. Then at any time t , the set of observations i with common observable state x_t and who made the same current choice d_t must be either type $k = 1$ or type $k = 2$. If they are of the same type, they face the same prospects regarding their state variables next period.¹³ Moreover, they also face a common distribution of idiosyncratic error terms next period $f(\varepsilon_{t+1})$. 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.¹⁴ We should then see two, and only two, different values of SR-CPs for each observed state-choice combination.¹⁵ Essentially, self-reported probabilities allow us to "reveal" type membership. Then, after

¹²Note 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 made the same current choice, we avoid this problem.

¹³Note that under Assumption F4, they would face the same prospects even if they were of different types.

¹⁴Different 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.

¹⁵Note that this holds regardless of whether Assumption F4 is true or, instead, the transitions for the observed states depend on the unobserved type.

uncovering the unobserved type, estimation methods such as those proposed by Hotz & Miller (1993) or Hotz, Miller, Sanders & Smith (1994) apply directly, treating type as an additional 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 & Thesmar (2002) apply directly.¹⁶

The linking technology, which we introduce more formally below, is a technique to "link" observations in the data. The linking is done via self-reports, which act as the chain's interconnecting links for each unobserved type. When we have two self-reports available for each observation, the only, rather weak, requirement for the linking technology to work is the absence of isolated islands in the space of feasible state-choice combinations. These "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 bridges to other regions of the state-choice space. Below we set up some notation and more formally define the linking technology along with the No-Islands assumption in which we rule out the existence of such islands.¹⁷

Definition 1 (*Revelation of types*) *A revelation of types is defined by an equivalence relation \sim on the set of observations $I = \{1, 2, \dots, N\}$. 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. The underlying parameter K is unknown. M does not necessarily recover K .*

Let the pair of self-reports be elicited at t' and t'' for all observations.

Definition 2 (*Linking Technology*) *Define a binary relation, R , in the following way:*

$$\forall i, j \in \{1, 2, \dots, N\},$$

$$i R j$$

¹⁶Magnac & 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.

¹⁷Alternatively, observations in the isolated islands can be discarded provided that suitable assumptions about their representativeness hold.

iff

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

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

$$i \sim j$$

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

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

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

Assumption SR-No Islands is defined after specifying a particular linking technology.

- **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. ■

We focus on the case in which we have permanent unobserved heterogeneity or "types" and where we have 1-period ahead SR-CPs. In the extensions section we discuss some variations. In section 5.1 we consider the elicitation of S-periods ahead SR-Cumulative CP. Later in sections 5.2 and 5.3 we address the case in which the unobserved heterogeneity is continuous as well as the case in which unobserved state variables evolve as a Markov process. In the remainder of this section we maintain Assumptions F1 and F3.

3.1.2 1-Period Ahead SR-CPs

For now, let's assume the available self-reports are about 1-period ahead CPs. In general, these self-reports can occur before or after the choice has been made this period. In what

follows, and unless noted otherwise, we assume that the 1-period ahead SR-CPs are elicited after the current choice, d_t has been made.

If the model in question were deterministic, 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, $p(d_{t+1} = 1|x_{t+1}, k)$ 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, $f_x(x'|d, x)$ as weights. In other words, the question elicits the "expected CCP". Formally,

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

In some case, it is also possible that the question actually elicits the one-period ahead **CCP at the Modal State**. In this case the respondent reports the CCP at the mode of the distribution of her own state variables next period. Given homogenous transitions, we can infer what that state is and we are then back to the simpler deterministic case. We can then "link" the self-reports at all those states. We call this the "Solve and Link" strategy: We solve out for the implied (modal) observed state at which the self-reported probability is being elicited. Then we link all the CCPs to trace out the unobserved types in the observed state space. It should be emphasized that rather than being something like "what's the probability that you'd choose $d_{t+1} = 1$?" here we are assuming the question eliciting the SR-CP is something more like the following: "Look one period ahead and consider what's your most likely situation at that time. In that situation, what would be the probability that

you'd choose $d = 1$?" Note that the introduction to the second question more explicitly instructs the respondent to situate himself in the most likely (modal) state one period ahead and then only report the choice probability assuming she will in fact be in that state. Formally, $p^{SR}(d_{t+1}|x_t, d_t, k) = \Pr(d_{t+1}|x_{t+1}^m, k)$ where $x_{t+1}^m = \text{Mode}(x_{t+1}|x_t, d_t)$ is the modal state at time $t + 1$ given x_t and d_t . That is $x_{t+1}^m = \arg \max_{x_{t+1}} \{f(x_{t+1}|x_t, d_t)\}$

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.

We focus on the case in which we have two self-reports available for each individual.¹⁸ In this case we can work within a very general class of models. We can exploit the self-reports to group observations into types, without trying to recover the implied CCPs. By having at least two self-reports we can connect observations at different points in the state space who belong to the same type. In particular, any two observations who share one common self-report at a given state-choice combination are of the same type and their other self-reports add to our signals to identify that type. The "linking technology" is extremely powerful. By having two self-reports we can trace out types in unrestricted models in which the profile of choice probabilities for different types may be allowed to cross in the state space. The linking technology allows us to overcome the ambiguity created by these crossings.

Recovering type-specific CCPs using 2 Self-Reports of Expected CCPs. While the identification of number of types (and type membership for each observation) doesn't actually require it, we can also explore the conditions under which we can recover the actual type-specific conditional choice probabilities. We will later make use of these results in more complex settings, but it is useful to introduce the issue now. When expected CCPs

¹⁸Again, well known surveys such as NLYS and HRS do include at least two self reports about subjective probability of future choices for the same individual. If only one Self-Report is available we need to restrict ourselves to cases where the CCPs are monotonic on type across the state space. For example, we could restrict ourselves to a class of models where one type always has higher choice probability. This is an important restriction. When profiles of SR-CPs for different types "cross" at some point in the state space, identification problems arise.

are reported, the respondent reports an average of CCPs, with the average taken using the transition probability. To recover the underlying CCPs we use the alternative "Link and Solve" strategy:

1. We first **link** SR-CPs from the same type and form a system of equations.
2. We then **solve** the system of equations and recover the type-specific CCPs.

To be specific, the first 1-period ahead SR-CP reported at t' gives us one equation for respondent i of type k_i .

$$p^{SR}(d_{i,t'+1} = 1|x_{it'}, d_{it'}, k_i) = \sum_{x_{i,t'+1}} \Pr(d_{i,t'+1} = 1|x_{i,t'+1}, k_i) f(x_{i,t'+1}|x_{it'}, d_{it'}, k_i)$$

where $p^{SR}()$ and $f()$ are known and $\Pr(d_{i,t'+1} = 1|x_{i,t'+1}, k_i)$ for all $x_{i,t'+1}$ are the unknowns. In general, we have $|X|$ unknowns so we need more equations. We then link this equation with a similar equation based on the respondent's second self-report and with the self-reports of other respondents j of the same type who have been linked to i to form a linear system of equations that has as many equations as unknowns.

$$\begin{aligned} p^{SR}(d_{i,t'+1} = 1|x_{it'}, d_{it'}, k_i) &= \sum_{x_{i,t'+1}} \Pr(d_{i,t'+1} = 1|x_{i,t'+1}, k_i) f(x_{i,t'+1}|x_{it'}, d_{it'}, k_i) \\ p^{SR}(d_{i,t'+1} = 1|x_{it''}, d_{it''}, k_i) &= \sum_{x_{i,t'+1}} \Pr(d_{i,t'+1} = 1|x_{i,t'+1}, k_i) f(x_{i,t'+1}|x_{it''}, d_{it''}, k_i) \\ &\vdots \\ p^{SR}(d_{j,t'+1} = 1|x_{jt'}, d_{jt'}, k_j) &= \sum_{x_{j,t'+1}} \Pr(d_{j,t'+1} = 1|x_{j,t'+1}, k_j) f(x_{j,t'+1}|x_{jt'}, d_{jt'}, k_j) \end{aligned}$$

where $k_i = k_j, \forall i, j$ is guaranteed by the "linking technology". We can then solve for the CCPs, $\{\Pr(d = 1|x, k)\}_{x \in X}$ by using standard techniques to solve systems of linear equations. There are $|X|$ unknowns and at most $|X| \times |D|$ different self-reports.

Note that once these type-specific CCPs have been recovered, they could be plugged-in directly instead of the non-parametric first stage probabilities in the typical Hotz-Miller two-step approach.

We have focused on discrete types, 1-period ahead self-reports and time-invariant unobserved heterogeneity. Our framework can be extend to relax each of these. We briefly discuss these extensions below in Section 5.

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.¹⁹ In particular, there is likely to be substantial "heaping" or "bunching" at common reference points like 0, 0.10, 0.50, 0.90 and 1. See Walker (2003), Hill, Perry & Willis (2004) and Blass, Lach & Manski (2010). Surprisingly, there is no bunching at 0.33 and 0.66 which a priori appear to be good focal points when the probability reflects 1 out of 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 order to make our method more empirically relevant, in this section we address the issue of "less than ideal" subjective choice probability assessments and 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 $B = 25$ "focal" or "reference" values, b , that have been consistently found in practice to account for most of the self-reported probabilities.²⁰ With a little abuse of notation, let B also denote the cardinality of the set B .

¹⁹See Karni (2009) for a formalization of truthful elicitation of probabilities.

²⁰Indeed, casual inspection of some of the responses to these type of questions in the National Longitudinal Survey of Youth 1997 NLSY97 reveals a pattern of clustering around the values in the particular set B we defined. However, our methods can be used with any set B . That is, the set of focal values can be modified on a case by case basis if the pattern of bunching in a specific survey is coarser or more detailed than this one.

Then

$$b \in B = \{0; 1; 2; 5; 10; 15; 20; 25; 30; 35; 40; 45; 50; 55; 60; 65; 70; 75; 80; 85; 90; 95; 98; 99; 100\}.$$

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 1-period ahead $E[CCP]$ self-reports. 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%.

In this section we will show that a variation of our linking technology, coupled with mild assumptions on the pattern of bunching across types and about the sampling of self-reports, can succeed in overcoming this problem. We maintain assumptions $F1$, $F3$, and $F4$ on the transition probability for state variables

A precise self-report of i at time t is defined to be a function of x_{it} , d_{it} , and k_i , which can be 1-period- or s -period-ahead expected CCP, modal CCP, etc. Following the notation in the previous sections, let $p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i)$ be the self-reported choice probability that satisfies SR-Precise. Now, consider two SRs at t' and t'' . 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 notation to emphasize it is now a focal self-report: $p_i^{SRF}(x_{it}, d_{it}, k_i)$. Formally,

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

where $p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i)$ may be a modal CCP or an Expected CCP. Actually, if $p_i^{SR}(d_{t+1} = 1|x_{it}, d_{it}, k_i) = E[CCP]$ we need to account for an additional layer of round-off

in the underlying CCPs, which we then denote FCCPs:²¹

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

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: $\tilde{p}_i^{SRF}(x_{it}, d_{it}) = p_i^{SRF}(x_{it}, d_{it}, k_i)$.

Definition 3 (Bunching) *Two SRs 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:

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

²¹This additional layer of rounding off corresponds to the idea that an additional source of discrepancy between the theoretical $E [CCP]$ 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 introduce some limited rationality at the self-report stage. Behavior continues to be fully rational.

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_{ij}^B, d_{ij}^B) .*

Assumption B1 guarantees that whenever there are two observations i and j of different types reporting at a bunching state-choice for them, the bunching of different types is immediately detected. Hence, the to-be-defined "linking technology under bunching" can use this notion of "bunching state-choice".

In particular, under Assumption B1, 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_{ij}^B, d_{ij}^B) .

However, Assumption B1 is not enough to identify the types. Consider the following example in Figure 2. There is no way of telling whether the observations are 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.

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_{ih}^B, d_{ih}^B) 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: $\forall i, j \in \{1, 2, \dots, N\}$,*

$$i R^B j$$

iff the following conditions are met:

Figure 1: Illustration of Immediate Detection of Bunching Observations

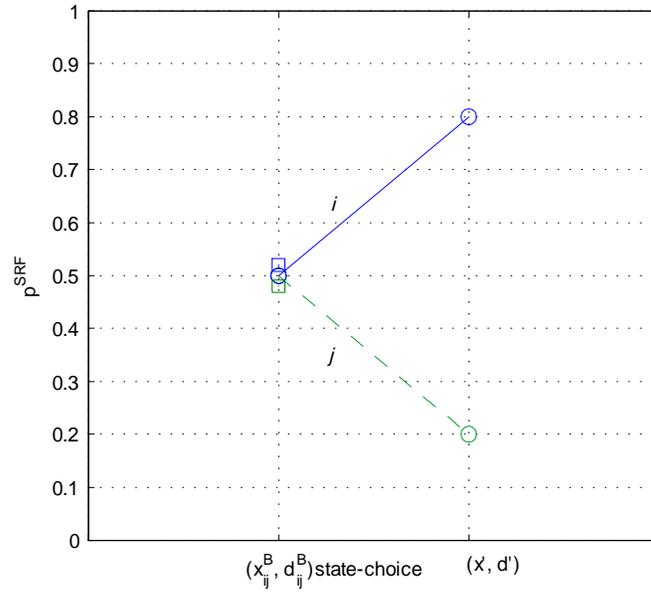


Figure 2: Problem Without Assumption B2

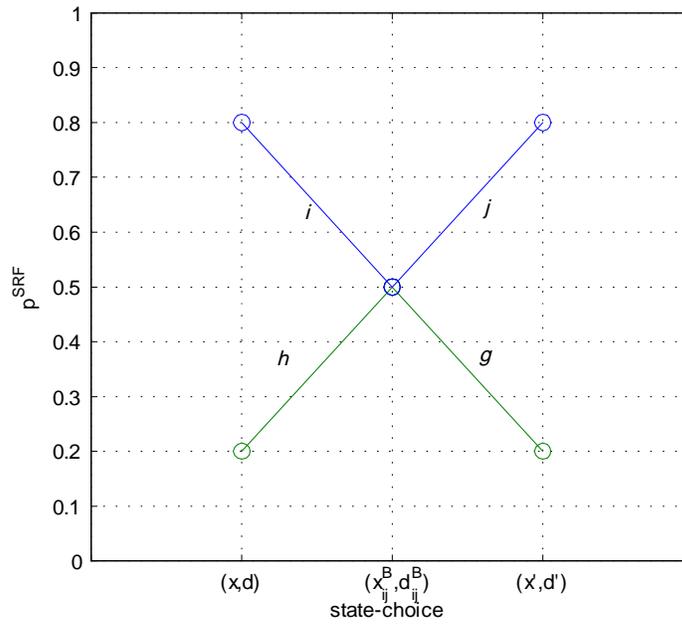
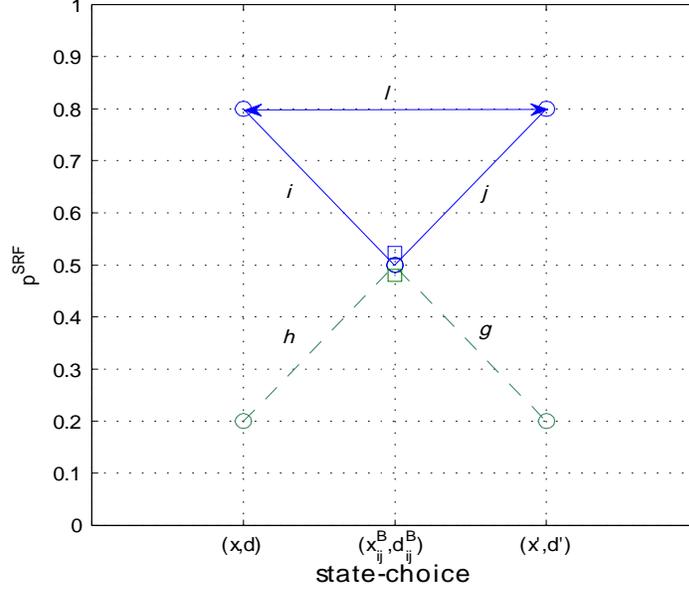


Figure 3: Illustration of Bridging the Bunching



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

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

2. if \exists observation h ,

$$\{\tilde{p}_i^{SRF}(x_{it'}, d_{it'}), \tilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \cap \{\tilde{p}_j^{SRF}(x_{jt'}, d_{jt'}), \tilde{p}_j^{SRF}(x_{jt''}, d_{jt''})\} = \{\tilde{p}_i^{SRF}(x_{ih}^B, d_{ih}^B)\}$$

then

$$\begin{aligned} & \exists l, \{\tilde{p}_i^{SRF}(x_{it'}, d_{it'}), \tilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \\ &= \{\tilde{p}_i^{SRF}(x_{it'}, d_{it'}), \tilde{p}_i^{SRF}(x_{it''}, d_{it''})\} \Delta \{\tilde{p}_j^{SRF}(x_{jt'}, d_{jt'}), \tilde{p}_j^{SRF}(x_{jt''}, d_{jt''})\}, \end{aligned}$$

where Δ denotes the set difference.

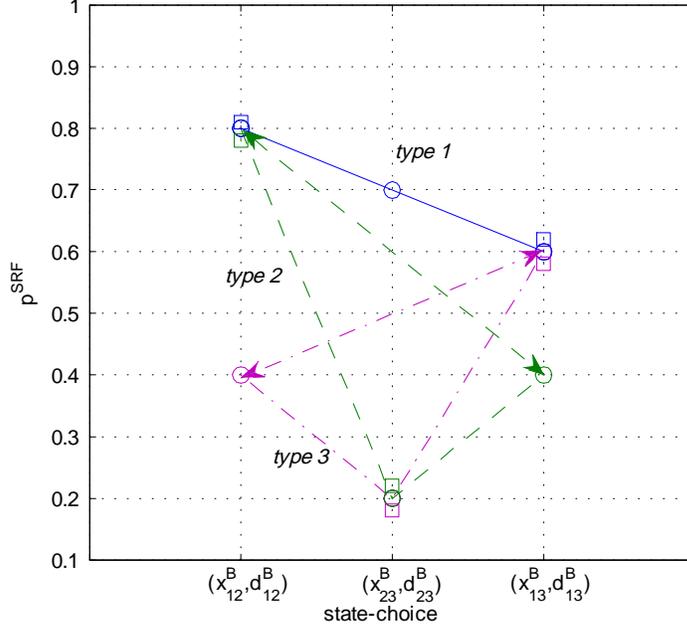
The linking technology under bunching is a relation \sim^B on $\{1, 2, \dots, N\}$: $\forall i, j \in \{1, 2, \dots, N\} = I$,

$$i \sim^B j$$

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

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

Figure 4: Identification of the Number of the Types



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

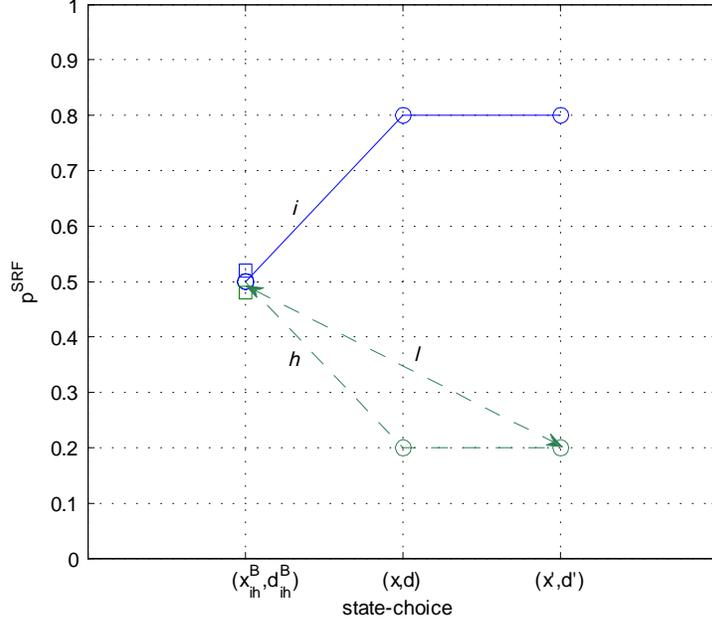
Lemma 2 *Under Assumptions B1, B2 and SR-No Islands, the linking technology under bunching recovers the true types exactly.*

Proof. See Appendix. ■

Note that the number of types is 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, $(x_{kk'}^B, d_{kk'}^B)$ 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

Figure 5: SRFs Allowed under Assumption B3 but not B1



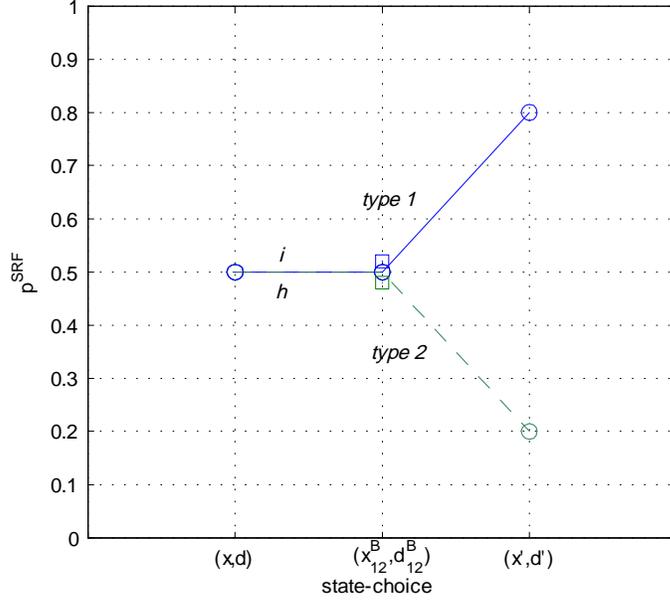
state-choice (x', d') s.t.

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

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 to the econometrician. 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 that links $p^{SRF}(x_{ih}^B, d_{ih}^B) = 0.5$ and $p^{SRF}(x', d') = 0.8$ for immediate detection of bunching types. Nevertheless, 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.

Figure 6: Non-identification of Types



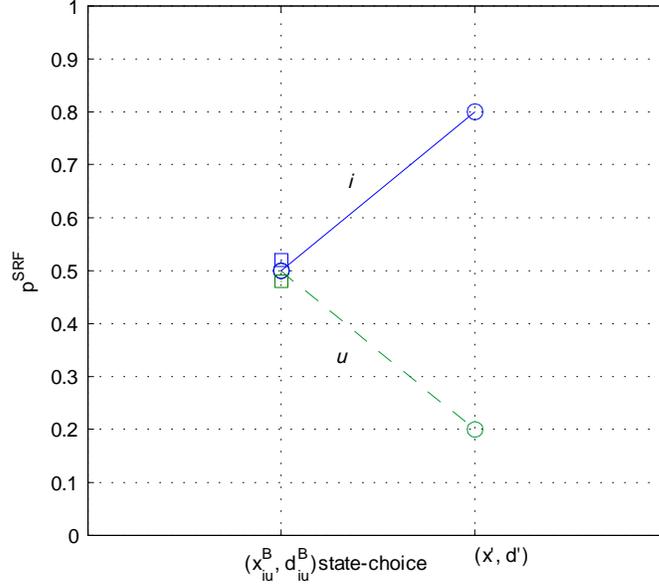
Assumption B4 (No observations with two "bunched" self-reports) Every observation i has at least one self-report elicited at a state-choice in which there is no bunching.

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 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

Figure 7: Immediate Detection of i-u Bunching



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 $\forall m R^B n$ 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.

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

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

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

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

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 B , the bias introduced by the approximation will be bounded. Since the focal CCPs differ from their precise values by at most ± 0.025 at each rounding-off, our procedure leads to a bounded bias of ± 0.05

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).

To that end note that for problematic observations we have

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

In the RHS, $\Pr(\{x_t, d_t\}_{t \neq t', t''} | k)$ 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

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

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.²³ 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 .²⁴

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, empirically relevant, 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.

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. True, 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

²³Alternatively, the denominator in the RHS, $\Pr(\{x_t, d_t\}_{t \neq t', t''})$, could be obtained by counting the proportion of observations who have this particular history of states and choices.

²⁴This 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.

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.²⁵

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:

$$\text{Type 1: } (\theta_{11}, R_1) = (-0.4, -3)$$

$$\text{Type 2: } (\theta_{12}, R_2) = (-1.2, -7)$$

We generate inputs to the simulated self-reports using our theory of self-report and further round off self-reported choice probabilities on the simulated elicitation according to the focal points described Section 3.2. 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 results of the Montecarlo simulations and illustrates that our linking technology allows quick and precise estimation of the unobserved heterogeneity in the structural model.²⁶The mean estimate over the R=500 repetitions is virtually the same as the truth. The standard deviation of the Montecarlo distribution is very small.²⁷

²⁵Only states visited with positive probability in the sample at hand (as oposed to all feasible states conceptually possible in the model) are used in this estimation strategy.

²⁶Convergence 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.

²⁷Standard Deviations for the montecarlo distribution of estimates are computed for each parameter as follows: $\sqrt{\frac{1}{R} \sum_{r=1}^R (\theta_r - \bar{\theta})^2}$ where $\bar{\theta} = \frac{1}{R} \sum_{r=1}^R \theta_r$

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

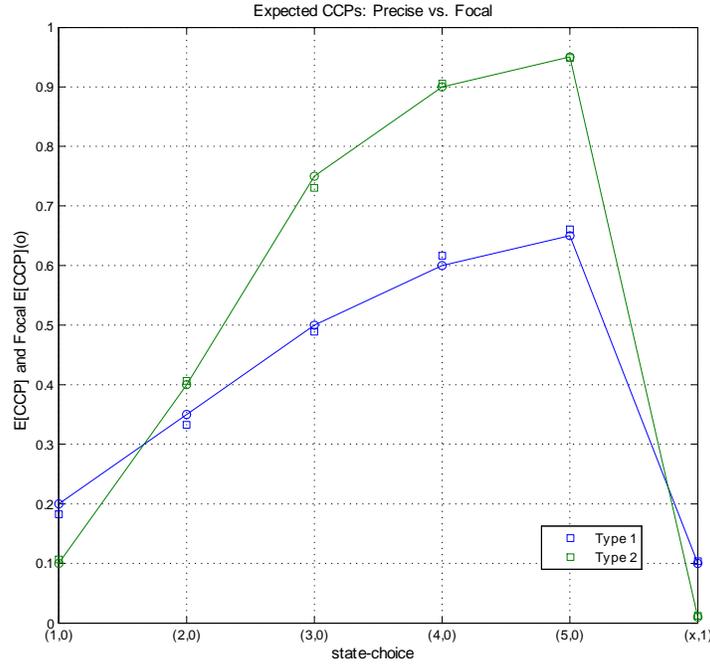


Table 1

	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	

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

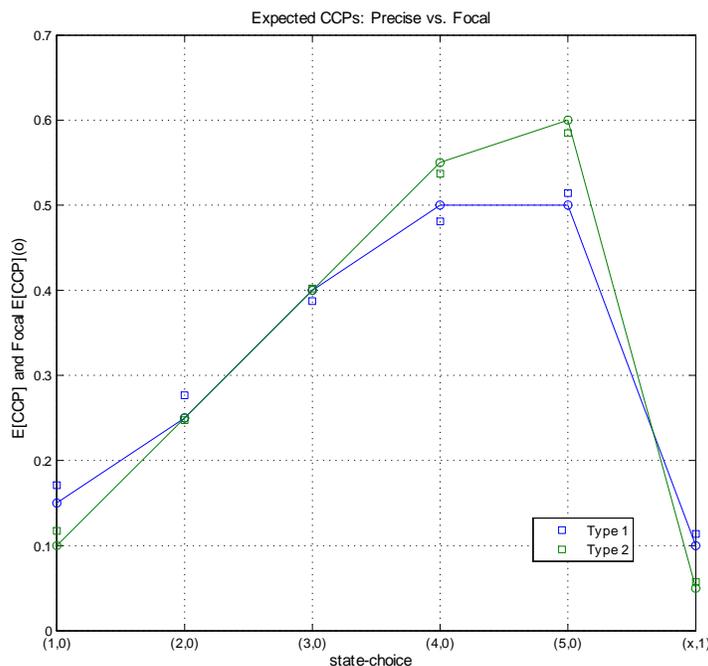
$$\text{Type 1: } (\theta_{11}, R_1) = (-0.27, -2.65)$$

$$\text{Type 2: } (\theta_{12}, R_2) = (-0.40, -3.75)$$

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 $E[\text{CCP}]$ s. The nearby circles show the corresponding "focal" $E[\text{CCP}]$ 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. **Discarded:** 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 costs, θ_1 are 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 $p(k | \{(x_t, d_t)\}_{t=t', t''}, \{x_t, d_t\}_{t \neq t', t''})$ using the Bayesian update described above and again assign types to "problematic" observations.²⁸ 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	"Discarded"		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	

²⁸In 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 Extensions

We first consider in some detail three important extensions in the ideal case in which we have self reports that are precisely elicited. We then briefly outline other directions for future research.

5.1 S-periods ahead Self-Reported Cumulative Choice Probability

In some cases the question explicitly specifies a longer planning horizon and the elicited subjective probability then refers to the probability of the action being taken at some point during the given planning horizon. In this case the linking technology still works and types can be revealed in the same fashion as in the 1-period ahead case. Then we proceed via standard Hotz-Miller using the revealed types as an extra observed state in the first stage.

Still, we can attempt to recover the underlying type-specific CCPs. Consider the 2-periods ahead SR-Cumulative CP. Then we have a nonlinear equation given by

$$\begin{aligned}
& p^{SR} (d = 1 \text{ at some point during the next two periods}) \\
&= \Pr (d_{t+1} = 1 \cup d_{t+2} = 1 | x_t, d_t, k) \\
&= 1 - \Pr (d_{t+1} = 0 \cap d_{t+2} = 0 | x_t, d_t, k) \\
&= 1 - \sum_{x_{t+1}} [\Pr (d_{t+1} = 0 \cap d_{t+2} = 0 | x_{t+1}, d_t, k)] f (x_{t+1} | x_t, d_t, k) \\
&= 1 - \sum_{x_{t+1}} [\Pr (d_{t+2} = 0 | x_{t+1}, d_{t+1} = 0, k) \Pr (d_{t+1} = 0 | x_{t+1}, k)] f (x_{t+1} | x_t, d_t, k) \\
&= 1 - \sum_{x_{t+1}} [\xi_{t+1}] f (d_{t+1} | x_t, d_t, k)
\end{aligned}$$

where

$$\xi_{t+1} = \left(\sum_{x_{t+2}} \Pr (d_{t+2} = 0 | x_{t+2}, k) f (x_{t+2} | x_{t+1}, d_{t+1} = 0, k) \right) \Pr (d_{t+1} = 0 | x_{t+1}, k)$$

Again if assumption F4 holds, in this equation the $f (x_{t+1} | x_t, d_t)$ and $p^{SR} ()$ are known, whereas the $\Pr (d = 0 | x, k) \forall x$ are unknown. Here we can also link the two self-reports of the same respondent and form additional equations with other self-reports from other

respondents of the same type until we have a system that can be solved.²⁹

5.2 Continuous Distribution of Unobserved Heterogeneity

Allowing for a continuous distribution of unobserved heterogeneity in dynamic programming models is very complicated. The only computationally feasible attempt we are aware of is the importance sampling strategy proposed by Akerberg (1999,2009). Let's assume we have the full set of precise CCPs or, alternatively, we have precise E[CCP]s at every possible state choice combination (x, d) . In our machine replacement example, this would amount to having SR-CPs requested at each and every one of the six state-choice combinations.³⁰ In this case we can estimate the model allowing for a nonparametric continuous distribution of unobserved heterogeneity as follows: With all the possible E[CCPs] in hand we can solve a system of equations and recover the individual specific CCPs at every possible state point. Then we can "plug-in" the individual-specific set of CCPs to compute individual specific W_i for each $i = 1, \dots, N$ exploiting Hotz & Miller's alternative representation. If we redefine structural parameter heterogeneity as deviations η_i from a common mean,

$$\begin{aligned}\theta_{1i} &= \mu_{\theta_1} + \eta_{1i} \\ R_i &= \mu_R + \eta_{Ri}\end{aligned}$$

the rest of the estimation protocol follows as in Hotz-Miller. The estimation routine searches for the mean of the structural parameters and, at each iteration, we solve out for the structural parameter deviations $\{\eta_i\}_{i=1}^N = \left\{ \left(\begin{array}{c} \eta_{1i} \\ \eta_{Ri} \end{array} \right) \right\}_{i=1}^N$ that are consistent with the

²⁹Unlike the 1-period ahead case, this system of equations is non-linear (even under precise elicitation) and the computational advantage of this strategy should be evaluated on a case by case basis for each specific model. This system of non-linear equations grows with the state-space so while our basic linking technology still works, recovering the underlying CCPs directly from the self-reports becomes more computationally demanding in realistic models. Still, it should be kept in mind that this whole step needs to be done only once so one can easily afford some computational cost.

³⁰Note that given the renewal structure of our model, we need only consider the following six state-choice combinations: $(1, 2)$, $(2, 0)$, $(3, 0)$, $(4, 0)$, $(5, 0)$ and $(x, 1)$

for any x . No matter what the state is, if a replacement decision is made, the state variable next period is "reset" to one with probability one.

data (given current trial parameter vector for the mean, $\begin{pmatrix} \mu_{\theta_1} \\ \mu_R \end{pmatrix}$). We do so by picking two CCPs implied by the self-reports ($\text{CCP}_i^{SR}(x_{it})$) and solving the linear system based on

$$Z\eta_i = E$$

where

$$\begin{aligned} Z &= \begin{pmatrix} \tilde{z}^{\hat{P}}(1, x_{it}) - \tilde{z}^{\hat{P}}(0, x_{it}) \\ \frac{1}{\text{CCP}_i^{SR}(x_{it})} - 1 \end{pmatrix} \\ E &= -\tilde{e}^{\hat{P}}(1, x_{it}) + \tilde{e}^{\hat{P}}(0, x_{it}) - \tilde{z}^{\hat{P}}(1, x_{it})\theta_u + \tilde{z}^{\hat{P}}(0, x_{it})\theta_u \end{aligned}$$

which can be derived by noting that³¹

$$\text{CCP}_i^{SR}(x_{it}) = \frac{\exp\left\{\tilde{z}^{\hat{P}}(1, x_{it})(\theta_u + \eta_i) + \tilde{e}^{\hat{P}}(1, x_{it})\right\}}{\exp\left\{\tilde{z}^{\hat{P}}(0, x_{it})(\theta_u + \eta_i) + \tilde{e}^{\hat{P}}(0, x_{it})\right\} + \exp\left\{\tilde{z}^{\hat{P}}(1, x_{it})(\theta_u + \eta_i) + \tilde{e}^{\hat{P}}(1, x_{it})\right\}}$$

After convergence we can "plot" the nonparametric joint distribution of the structural parameters and recovers its moments.

For example, consider the following DGP.

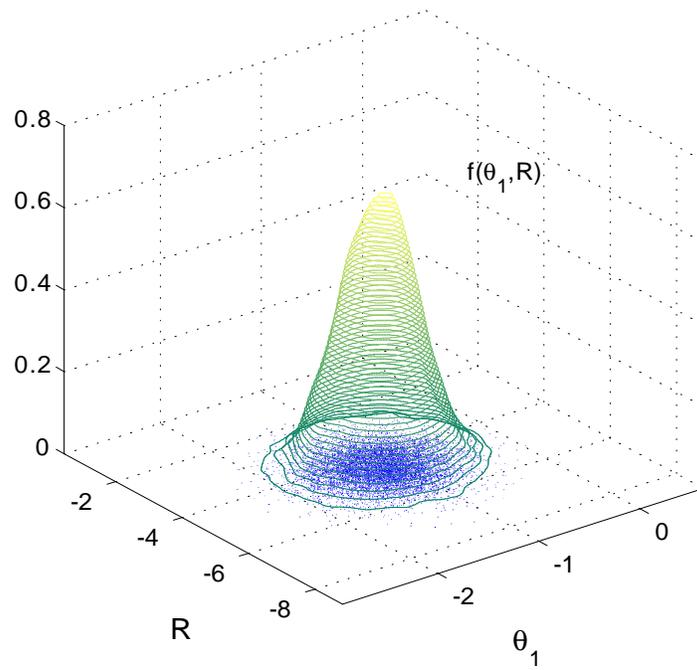
$$\begin{pmatrix} \theta_{1i} \\ R_i \end{pmatrix} \sim N \begin{pmatrix} \mu_{\theta_1} \\ \mu_R \end{pmatrix}, \Sigma$$

with true parameter values given by $\mu_{\theta_1} = -1.15$, $\mu_R = -4.45$, $\Sigma = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.5 \end{bmatrix}$.

Figure 11 depicts the true distribution of structural parameter values in the population.

³¹If current utility is given by $U(d, x_{it}, \varepsilon_{it}) = u(d, x_{it}) + \varepsilon_{it}(d)$ and we let $\theta = (\theta_u, \theta_f)$ be the vector of structural parameters. Consider a linear-in-parameters utility $u(d, x_{it}) = z(d, x_{it})'\theta_u$ where $z(d, x_{it})$ is a $\text{Dim}(\theta_u) \times 1$ vector. State variables evolve according to $f_x(x_{t+1}|d_t, x_t, \theta_f)$. The choice specific value functions can be re-written as $v(d, x, \theta) = \tilde{z}(d, x, \theta)\theta_u + \tilde{e}(d, x, \theta)$ where $\tilde{z}(d, x_t, \theta) = z(d, x) + \sum_{s=1}^{T-t} \beta^s E_{x_{t+s}|d_t=d, x_t} \left[\sum_{d'=0}^D P(d'|x_{t+s}, \theta) z(d', x_{t+s}) \right]$ and $\tilde{e}(d, x_t, \theta) = \sum_{s=1}^{T-t} \beta^s E_{x_{t+s}|d_t=d, x_t} \left[\sum_{d'=0}^D P(d'|x_{t+s}, \theta) e(d', x_{t+s}) \right]$. The policy function is $\alpha(x_t, \varepsilon_t) = \arg \max_d \{v(d, x_t, \theta) + \varepsilon_t(d)\}$ and the expected error conditional on optimal choice is $e(d, x_t) = E[\varepsilon_t(d) | x_t, \alpha(x_t, \varepsilon_t) = d]$. Hotz & Miller (1993) show that $e(d, x_t) = f(d, P(\cdot|x_t, \theta), G_\varepsilon)$. For example, if $\varepsilon_{it}(d)$ are iid Extreme Value, then $e(d, x_t) = \gamma - \log[P(d|x_{it}, \theta)]$

Figure 11: Distribution of Structural Parameter Population Heterogeneity



Our estimates are obtained in less than one minute. The estimated population means for both structural parameters are

$$\begin{aligned}\widehat{E}[\theta_1] &= -1.1496 \\ \widehat{E}[R] &= -4.4451\end{aligned}$$

which are almost exactly equal to the truth.

The estimated variance covariance matrix for the distribution of structural parameter heterogeneity in the population is

$$\widehat{Var}(\theta_1, R) = \begin{bmatrix} 0.1002 & 0.0028 \\ 0.0028 & 0.5043 \end{bmatrix}$$

which is consistent with the underlying bivariate normal DGP. Note however, that our approach is non-parametric in the sense that we could have recovered any kind of distribution since at no point did we use normality, other than to simulate the data.

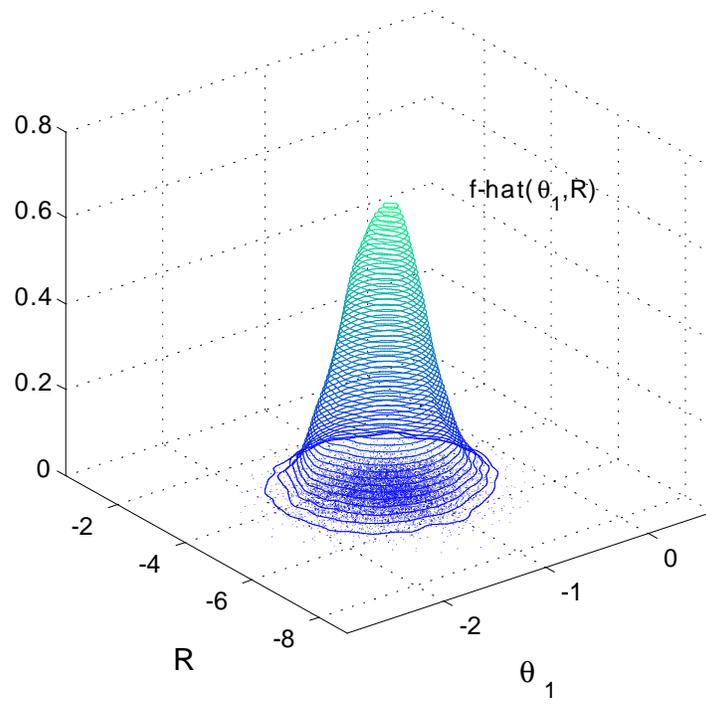
Figure 12 presents the plot for the non-parametric estimate of the distribution of structural parameters in the population.

Note that in more complex models, having the entire set of $E[\text{CCPs}]$ might be unrealistic. However, if we have three self-reports, we can use a variant of the linking technology that supplements our original linking strategy with an interpolation step. Indeed, if we assume Finite Dependence³², Modal CCP self-reports and Assumption F4, then we can recover the continuous distribution of unobserved heterogeneity in structural parameters by interpolating CCPs using nearest neighbors to complete each individual's full set of self-reports.³³

³²The concept of finite dependence was originally developed by Altug & Miller (1998) and further generalized by Arcidiacono & Miller (2008). It sharpens the insight of the Hotz-Miller original result by showing that for certain class of models only the 1-period (or, in general for finite $\rho > 0$, the ρ -period) ahead CCPs are needed in the alternative representation of the value functions.

³³Nearest neighbors are those observations that report similar SR-CPs in two of the same state-choice cells.

Figure 12: Estimated Distribution of Structural Parameter Heterogeneity



5.3 Time-Varying Unobserved Heterogeneity

As we discussed in the Introduction, recent efforts, to which our paper contributes, show that under some conditions, it is possible to accommodate permanent unobserved heterogeneity in two-step estimation strategies. The work of Arcidiacono & Miller (2008) pushes the frontier forward by not only allowing for unobserved heterogeneity but by letting it evolve in systematic ways over time.

In this subsection we briefly note that our linking approach can be modified to accommodate unobserved heterogeneity that evolves over time. To continue with our machine replacement example, we now think of firms as being in one of two possible unobserved states. These unobserved states are not permanent, but rather can change over time. We focus on cases in which we have two self-reports taken in consecutive periods.

In this scenario k is no longer fixed but becomes k_t , a random variable that exogenously evolves over time as in assumption F2. The key idea can be grasped with $K = 2$ (i.e. there are two possible unobserved states $k = 1, 2$). Suppose that at time t' the first self-report is collected. Among those with the same $(x_{t'}, d_{t'})$ we can identify those who give different SR-CPs and, following the reasoning of previous sections, those who have different $k_{t'}$. Without loss of generality we can assign one group to $k_{t'} = 1$ and the others to $k_{t'} = 2$. We can track this group into the next period. Suppose that next period, $t' + 1$, the second self-report is collected. We can see how the answers of each group split at the time of the second self-report. These splits give information on unobserved state transitions π_{jl} for $j, l = 1, 2$.

How do we know which % transitioned into which state? We can compare self-reports in the second period across different states and check against the self-reports collected at those same states in the first period as long as the model is stationary. We are then able to identify those that remained in their previous unobserved state and those who transitioned into a new one.

Finally, note that all of the above can be generalized to: a) any $K > 2$, b) self-reports collected in any two, not necessarily consecutive time periods (t', t'') , c) cases in which the first period collecting self-reports, t' , is not the first sample period and d) models with choice-dependent transitions for unobserved states.³⁴

³⁴Choice dependent transitions for unobserved states accommodate the following case: instead of firms

5.4 Other Extensions

1. In some special cases we can entertain the possibility of estimating the model off SR-CPs alone, rather than using actual revealed choice data.
2. Using Other Probability Questions:
 - As seen in Section 5.1, sometimes we have SR-Cumulative CPs like "What's the probability that you will choose $d = 1$, at least once, at some point during the next S years?". In some cases, we may have two of these questions at the same time t' , eliciting the cumulative probability that an action is taken or a state is reached within two or more time horizons. We may have, for example,

$$\left\{ \left(\begin{array}{l} p_i^{SR} (\cup_{s=1}^5 \{d_{t+s} = 1\} | x_t, d_t, k) \\ p_i^{SR} (\cup_{s=1}^{10} \{d_{t+s} = 1\} | x_t, d_t, k) \end{array} \right) \right\}_{i=1}^N$$

eliciting the 5- and 10-period ahead cumulative CP. This eliminates the need of a panel of self-reports.

- "What's the probability that you will have $x_{t+S} = x$ in S years?"
3. Using Other Types of Expectations: Some questions don't ask about the probability of making a given choice or reaching a given state but rather ask whether the respondent "expects" to make that choice or reach a given point in the state space.
 - "Do you expect to have $d_{t+S} = d$ in S years?" This case which severely limits the informational content of the self-report arguably asks whether $\Pr(d_{t+S} = d | x_t, d_t, k) > 50\%$
 - "What value of x do you expect to end up having over your planning horizon?" In these cases x may refer to the number of children or completed years of education

differing in k_t , machine differences are the underlying source of unobserved heterogeneity. Suppose when a firm replaces an old machine, the new machine may turn out to be an "easy maintenance, easy replacement" machine or a "problem" machine which is difficult to maintain and difficult to replace. Here the unobserved state may evolve over time but only if the renewal action is taken. To handle this case we relax Assumption F2.

that a person will have over their remaining lifetime, i.e. over the next $T - t$ years. Here, respondents could arguably be rounding-off the expected maximum, $E[x_T|x_t, d_t, k]$ to the nearest integer or providing the mode, $Mode[x_T|x_t, d_t, k] = \arg \max_{x_T} \{f(x_T|x_t, d_t, k)\}$

4. Application to multiple agent models.
5. Multinomial choice and use of GEV unobservables following McFadden (1978) , Breshnahan, Stern & Trajtenberg (1997), Arcidiacono (2005) and Arcidiacono & Miller (2008).

6 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 & 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. Indeed, the method can be implemented with only two unconditional self-reports about future choice probabilities per respondent. Our Montecarlo experiments show that 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. While our focus has been on single agent models of dynamic discrete choice, we believe that our approach can be generalized to the many other contexts discussed in the introduction as long as subjective expectations data is available to supplement traditional data on observed choices and states. We leave these and other extensions for future research. We believe this is a first step in a fruitful research program that leverages new forms of available data to be more flexible about the specification of unobserved heterogeneity in structural estimation.

Appendix: 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 $\{i_1, \dots, i_n\} \subseteq I$, such that $i R i_1 R \dots R i_n R j$. By transitivity of " \sim ", it is enough to show that $\forall m R n$, $k_m = k_n$. By definition, $m R n$ iff

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

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, $k_i = k_{i_1} = \dots = k_{i_n} = k_j$.

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 $i \not\sim j$. 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 $i \not\sim j$. So $\Sigma^{[i]} \cap \Sigma^{[j]} = \emptyset$. But it further contradicts Assumption SR-No Islands, because $\Sigma^{[i]} \cap \Sigma^{[j]} = \emptyset$, together with $i \not\sim j$, 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 $\{i_1, \dots, i_n\} \subseteq I$, such that $i R^B i_1 R^B \dots R^B i_n R^B j$. By transitivity of " \sim^B ", 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. $\{\tilde{p}_m^{SRF}(x_{mt'}, d_{mt'}), \tilde{p}_m^{SRF}(x_{mt''}, d_{mt''})\} \cap \{\tilde{p}_n^{SRF}(x_{nt'}, d_{nt'}), \tilde{p}_n^{SRF}(x_{nt''}, d_{nt''})\} \neq \emptyset$,

2. if \exists observation h ,

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

then

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

where Δ 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:

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

In this case, n qualifies as the observation h in the second condition, so

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

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 $i \not\sim^B j$. 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] Akerberg (1999) Importance Sampling and the Method of Simulated Moments, mimeo.
- [2] Akerberg (2009) A New Use of Importance Sampling to Reduce Computational Burden In Simulation Estimation, Quantitative Marketing and Economics, forthcoming
- [3] Arcidiacono (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., A. Khwaja and L. Ouyang (2007) Habit persistence and teen sex: Could increased access to contraception have unintended consequences for Teen pregnancies? mimeo. Duke University.
- [5] Arcidiacono & Miller (2008), CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity, mimeo.
- [6] Aguirregabiria & Mira (2009), Dynamic Discrete Choice Structural Models: A Survey. *Journal of Econometrics*.
- [7] Aguirregabiria & Mira (2002), Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models, *Econometrica*
- [8] Aguirregabiria & Mira (2007), Sequential Estimation of Dynamic Discrete Games, *Econometrica*
- [9] Aguirregabiria, Mira & Roman (2007), An Estimable Dynamic Model of Entry, Exit, and Growth in Oligopoly Retail Markets, *American Economic Review*
- [10] Altug & Miller (1998), The Effect of Work Experience on Female Wages and Labor Supply, *Review of Economic Studies*
- [11] Attanasio, O. P. (2009): Expectations and Perceptions in Developing Countries: Their Measurement and Their Use" *American Economic Review*, Papers and Proceedings.
- [12] Bajari, Hong, Krainer & Nekipelov (2009), Estimating Static Models of Strategic Interactions, mimeo.

- [13] Bajari, Benkard & Levin (2007), Estimating Dynamic Models of Imperfect Competition, *Econometrica*
- [14] Bajari, Fox, Kim & Ryan (2009), A Simple Nonparametric Estimator for the Distribution of Random Coefficients, mimeo.
- [15] Blass, Lach & Manski (2010), "Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability", *International Economic Review*, Vol. 51, No. 2, May
- [16] Blau, D. and D. Gilleskie (2008) The Role of Retiree Health Insurance in the Employment Behavior of Older Males. *International Economic Review*.
- [17] Blevins (2009) "Sequential MC Methods for Estimating Dynamic Microeconomic Models", Duke University, mimeo.
- [18] Bresnahan, Stern & Trajtenberg (1997), Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s, *RAND*
- [19] Buchinsky, Hahn & Hotz (2005), Cluster Analysis: A Tool for Preliminary Structural Analysis.
- [20] Carro, J. and P. Mira (2006), A Dynamic Model of Contraceptive Choice of Spanish Couples. *Journal of Applied Econometrics* 21, 955-980.
- [21] Casanova, M., He, Daifeng, Pantano, Juan & Zheng, Yu (2010)
- [22] Choo & Siow (2005): Dynamic Two-Sided Matching Games, mimeo
- [23] Dube, Fox and Su (2009), "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation" mimeo
- [24] Golan & Levy-Gayle (2008), Estimating a Dynamic Adverse-Selection Model: Labor-Force Experience and the Changing Gender Earnings Gap 1968-93, mimeo Carnegie Mellon University
- [25] Eckstein and Wolpin (1989) "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey" *Journal of Human Resources*, Fall, 24, 4

- [26] Eckstein and Wolpin (1999) "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities" *Econometrica*, 67(6): 1295-339
- [27] 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
- [28] Hill, Perry and Willis (2004), Estimating Knightian Uncertainty from Survival Probability Questions on the HRS, mimeo
- [29] Hotz & Miller(1993), Conditional Choice Probabilities and the Estimation of Dynamic Models, *Review of Economic Studies*
- [30] Hotz, Miller, Sanders & Smith (1994), A Simulation Estimator for Dynamic Models of Discrete Choice, *Review of Economic Studies*
- [31] Hu & Shum (2008) Identifying Dynamic Games with Serially-Correlated Unobservables, mimeo
- [32] Hu & Shum (2009) Nonparametric Identification of Dynamic Models with Unobserved State Variables, mimeo
- [33] Imai, S., N. Jain, and A. Ching (2009): "Bayesian Estimation of Dynamic Discrete Choice Models," *Econometrica*, 77(6), 1865-1899
- [34] Jofre-Bonet & Pesendorfer (2003), Estimation of a Dynamic Auction Game, *Econometrica*
- [35] Karni (2009) "A Mechanism for Eliciting Probabilities" *Econometrica*, Vol. 77, No. 2, March, 603–606
- [36] Kasahara & Shimotsu (2008), Pseudo-likelihood Estimation and Bootstrap Inference for Structural Discrete Markov Decision Models, *Journal of Econometrics*, 146(1), 92-106
- [37] Kasahara & Shimotsu (2009a), Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices, *Econometrica*

- [38] Kasahara & Shimotsu (2009b), Sequential Estimation of Structural Models with a Fixed Point Constraint, mimeo
- [39] Keane & Wolpin (1994) "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence" *Review of Economics and Statistics*
- [40] Keane & Wolpin (1997) "The Careers of Young Men" *Journal of Political Economy*
- [41] Keane & Wolpin (2009) "Empirical applications of discrete choice dynamic programming models" *Review of Economic Dynamics*
- [42] Todd & Wolpin (2009) "Structural Estimation and Policy Evaluation in Developing Countries" *Annual Review of Economics*, September 2010, Vol. 2
- [43] Keane, Todd & Wolpin (2010) "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications, *Handbook of Labor Economics*, forthcoming.
- [44] Liu, Mroz & van der Klaauw (2009), Mother's Work, Migration, and Child Achievement, *Journal of Econometrics*, forthcoming.
- [45] Magnac & Thesmar (2002) "Identifying Dynamic Discrete Decision Processes", *Econometrica*, Vol. 70, No 2, March 801-816
- [46] McFadden (1978) "Modelling the Choice of Residential Location", in A. Karlqvist et al., eds., *Spatial Interaction: Theory and Planning Models*. Amsterdam. North Holland Publishing.
- [47] Mira (2007) "Uncertain Infant Mortality, Learning and Life-Cycle Fertility". *International Economic Review* 48, 809-846
- [48] Norets, A. (2009) "Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables", *Econometrica*, 77(5), 1665-1682
- [49] Pakes, Ostrovsky & Berry (2008), Simple estimators for the parameters of discrete dynamic games (with entry/exit examples), *RAND*

- [50] Pesendorfer & Schmidt-Dengler (2008), Asymptotic Least Squares Estimators for Dynamic Games, *Review of Economic Studies*
- [51] Pesendorfer & Schmidt-Dengler (2008), Sequential Estimation of Dynamic Discrete Games: A Comment, mimeo
- [52] Rust (1987), Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher, *Econometrica*
- [53] Siebert, R., and C. Zulehner (2008): "The Impact of Market Demand and Innovation on Market Structure" Purdue University, mimeo.
- [54] Su, Che-Lin & Judd, Kenneth (2007), "Constrained Optimization Approaches to Estimation of Structural Models", mimeo
- [55] van der Klaauw (1996) Female Labor Supply and Marital Status Decisions: A Life Cycle Model, *Review of Economics Studies*, 1995, 63(2): 199-235.
- [56] van der Klaauw (2000) On the Use of Expectations Data in Estimating Structural Dynamic Models: An Analysis of Career Choices, mimeo, UNC-Chapel Hill
- [57] van der Klaauw & Wolpin (2008) Social Security and the Retirement and Savings Behavior of Low Income Households, *Journal of Econometrics*
- [58] Walker (2003), "Pregnancy and Fertility Expectations: Estimates of Bounded Rationality and Unintended Births", mimeo.
- [59] Wolpin (1984), An Estimable Dynamic Stochastic Model of Fertility and Child Mortality, *Journal of Political Economy*, 1984, Vol. 92: 852-875
- [60] Wolpin (1999), Commentary on "Analysis of Choice Expectations in Incomplete Scenarios", by C.F. Manski" *Journal of Risk and Uncertainty*
- [61] Wolpin & 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.