ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

http://www.econometricsociety.org/

Econometrica, Vol. 87, No. 5 (September, 2019), 1475–1505

NONPARAMETRIC INFERENCE ON STATE DEPENDENCE IN UNEMPLOYMENT

ALEXANDER TORGOVITSKY
Department of Economics, University of Chicago

The copyright to this Article is held by the Econometric Society. It may be downloaded, printed and reproduced only for educational or research purposes, including use in course packs. No downloading or copying may be done for any commercial purpose without the explicit permission of the Econometric Society. For such commercial purposes contact the Office of the Econometric Society (contact information may be found at the website http://www.econometricsociety.org or in the back cover of *Econometrica*). This statement must be included on all copies of this Article that are made available electronically or in any other format.

Econometrica, Vol. 87, No. 5 (September, 2019), 1475–1505

NONPARAMETRIC INFERENCE ON STATE DEPENDENCE IN UNEMPLOYMENT

ALEXANDER TORGOVITSKY
Department of Economics, University of Chicago

This paper is about measuring state dependence in dynamic discrete outcomes. I develop a nonparametric dynamic potential outcomes (DPO) model and propose an array of parameters and identifying assumptions that can be considered in this model. I show how to construct sharp identified sets under combinations of identifying assumptions by using a flexible linear programming procedure. I apply the analysis to study state dependence in unemployment for working age high school educated men using an extract from the 2008 Survey of Income and Program Participation (SIPP). Using only nonparametric assumptions, I estimate that state dependence accounts for at least 30–40% of the four-month persistence in unemployment among high school educated men.

KEYWORDS: State dependence, unemployment, nonparametric, partial identification, linear programming, dynamic discrete choice, moment inequalities.

1. INTRODUCTION

SUPPOSE THAT A RESEARCHER OBSERVES A BALANCED PANEL CONSISTING of a binary outcome $Y_{it} \in \{0, 1\}$ at time periods $t = 0, 1, \dots, T$ for a cross-section of agents indexed by i. The researcher's goal is to determine to what extent the outcome in the previous period, $Y_{i(t-1)}$, has a causal effect on the current period outcome, Y_{it}. For example, Heckman (1981a) studied whether past employment has a causal effect on future employment for married women. A negative causal effect could arise from search costs, human capital depreciation during non-employment, or quality signaling in hiring processes ("stigma" or "scarring" effects), among other explanations. Such an effect is commonly described as state dependence, or "true" state dependence for emphasis.

Positive serial correlation in employment outcomes $Y_i \equiv (Y_{i0}, Y_{i1}, \dots, Y_{iT})$ does not necessarily indicate state dependence. An alternative explanation is that individuals have persistent latent heterogeneity in their propensities for employment and, as a result, some individuals are always more likely to be employed than others (Heckman and Willis (1977), Heckman (1978, 1981a)). This would lead to positive serial correlation in observed employment outcomes even if there is no state dependence in employment.

Alexander Torgovitsky: atorgovitsky@gmail.com

The first draft of this paper was circulated under the title "Partial Identification of State Dependence" and dated February 12, 2015. The second draft was circulated under the title "Nonparametric Inference on State Dependence with Applications to Employment Dynamics" and dated January 29, 2016. This paper was presented at the University of Wisconsin at Madison, Northwestern University, the University of Chicago, Columbia University, the 2015 World Congress of the Econometric Society, the 2016 North American Winter Meeting of the Econometric Society, and the 2016 East Asian Meeting of the Econometric Society. My thanks to the audiences at those presentations. I thank Ivan Canay, Andres Santos, and Azeem Shaikh for several helpful discussions. Useful comments and feedback were provided by Stéphane Bonhomme, Xiaohong Chen, Iván Fernández-Val, Joachim Freyberger, Jim Heckman, Joel Horowitz, Hide Ichimura, Yuichi Kitamura, Pat Kline, Thibaut Lamadon, Chuck Manski, Whitney Newey, Matt Notowidigdo, Jack Porter, Nancy Stokey, and Tiemen Woutersen. Four anonymous referees provided constructive comments that led to substantial improvements in the paper. This research was supported in part by National Science Foundation Grant SES-1530538. I am grateful to the Department of Economics at the University of Chicago for supporting me hospitably through a visiting position while I completed part of this research in 2015–2016.

© 2019 The Econometric Society https://doi.org/10.3982/ECTA14138

The difference between these two explanations has important implications for the design and long-run efficacy of active labor market programs (Heckman (1978, 1981a, 1981b)). It is therefore important to have convincing econometric methods to quantify the degree to which persistence in employment is due to state dependence. In order to be convincing, these econometric methods must first address the difficult identification problem of distinguishing state dependence from persistent unobserved heterogeneity.

The main contribution of this paper is the development of a new nonparametric framework for tackling this problem. The framework is a dynamic potential outcomes (DPO) model, the premise of which is simple to state. Given a binary outcome $Y_{it} \in \{0, 1\}$ for agent i at time t, let $U_{it}(0)$ and $U_{it}(1)$ be two latent binary variables that represent the potential outcomes that would have been realized had the prior period outcome, $Y_{i(t-1)}$, counterfactually been 0 or 1, respectively. The observed outcome is therefore related to the potential outcomes as

$$Y_{it} = Y_{i(t-1)} U_{it}(1) + (1 - Y_{i(t-1)}) U_{it}(0).$$

The model primitive is the joint distribution of $U_{it}(0)$, $U_{it}(1)$ across all time periods.[1]

This model can be used to construct a number of measures of state dependence, including common measures such as the average treatment effect. I discuss several parameters that provide different measures of state dependence, and I show that they are usually not point identified. For three of them, I derive sharp worst-case bounds that use only the empirical evidence. The bounds are quite wide. In particular, the bounds imply that empirical evidence alone is never informative enough to reject the hypothesis of no state dependence. At the same time, the empirical evidence alone is also never informative enough to reject the hypothesis that *all* of the observed persistence in the data is due to state dependence.

I propose several additional nonparametric assumptions that can be maintained for more informative inference. The assumptions concern the temporal dependence and stationarity of the potential outcomes, as well as their relationships with other observed covariates. These assumptions are fully nonparametric and have intuitive interpretations. For additional interpretation, I consider the DPO model that is implied by a dynamic model of a forward-looking agent. I develop nonparametric conditions on the dynamic choice model that are sufficient to imply each of the conditions on the DPO model. I also consider specializations of these conditions to a widely used dynamic binary response model.

Since the DPO model is recursive, analytically deriving sharp bounds under additional assumptions is quite difficult.[2] Instead, I develop a general procedure for computing sharp bounds that is valid for broad classes of parameters and combinations of assumptions. In many cases, the procedure amounts to solving two linear programming problems and is therefore straightforward to implement. An attractive feature of this approach is its flexibility: The researcher is afforded greater freedom to choose parameters and combine assumptions, without needing to derive new analytic results for each new specification.

[1]After reading an early draft of this paper, Chuck Manski shared with me his slides for an invited talk in 2006 in which he proposed using the same type of model to study state dependence (Manski (2006)). This paper was developed independently and without knowledge of that talk. The analysis of the DPO model in this paper is significantly different than that in Manski's talk.

[2]For example, see Chernozhukov, Fernández-Val, Hahn, and Newey (2013), who showed how to construct bounds on state dependence under the "time is an instrument" assumption discussed in Section 4.4. Their analysis already requires subtle constructions, even with only a single assumption.

The econometric methodology proposed in this paper can be applied to any of the large variety of empirical settings in which identifying state dependence is important. These include the dynamics of welfare recipiency (Chay, Hoynes, and Hyslop (2004), Card and Hyslop (2005)), product choices among consumers (Keane (1997), Dubé, Hitsch, and Rossi (2010), Handel (2013)), self-reported health status (Contoyannis, Jones, and Rice (2004)), firm investment (Drakos and Konstantinou (2013)) and exporting (Bernard and Jensen (2004)) decisions, household investment behavior (Alessie, Hochguertel, and Soest (2004)), illicit drug usage (Deza (2015)), and eating disorders (Ham, Iorio, and Sovinsky (2013)). Irace (2018) used the methodology developed in this paper to study the dynamics of hospital choice.

I apply the methodology to study the employment dynamics of working age, high school educated men, using an extract from the 2008 Survey of Income and Program Participation (SIPP) that covers January 2011 to April 2013. I find little evidence of state dependence among employed workers. However, by maintaining a nonparametric stationarity assumption, I find evidence of substantial state dependence among unemployed workers. The main estimates indicate that at least 23% of unemployed workers would be employed if they had been employed in the previous period. Overall, state dependence accounts for at least 30–40% of the observed four-month persistence in unemployment. The results imply that short-term state dependence is an important phenomenon in the U.S. labor market.

The organization of this paper is as follows. In the next section, I develop the DPO model and connect it to a dynamic choice model. In Section 3, I discuss parameters of interest in the DPO model, derive worst-case bounds, and develop a general procedure for computing sharp identified sets. In Section 4, I propose an array of identifying assumptions that can be imposed in the DPO model. I analyze the economic content of these assumptions through the lens of the dynamic choice model. In Section 5, I apply the DPO model to study state dependence in unemployment. Section 6 contains a brief conclusion.

2. THE DYNAMIC POTENTIAL OUTCOMES MODEL

2.1. *Model*

The canonical *static* potential outcomes model is based on two unobserved outcomes, $U_i(0)$ and $U_i(1)$, that would have been obtained had a binary treatment, $D_i \in \{0, 1\}$, been exogenously manipulated to be 0 or 1. The observed outcome, Y_i, is related to the potential outcomes and the observed treatment state through $Y_i = D_i U_i(1) + (1 - D_i)U_i(0)$. The researcher is interested in inferring features of the unobservable distribution of $(U_i(0), U_i(1))$ from the observable distribution of (Y_i, D_i).

State dependence is the causal effect of a past outcome on a current outcome. At time t, the outcome is the current outcome, Y_{it}, and the "treatment" is the immediately preceding outcome, $Y_{i(t-1)}$.[3] I assume throughout the main text that $Y_{it} \in \{0, 1\}$ is binary for each t and discuss the extension to multi-valued outcomes in Appendix S1 of the Supplemental Material (Torgovitsky (2019b)). Thus, in analogy to the static potential outcomes model, suppose that, for each time period $t = 1, \dots, T$, there exist unobservable random variables $U_{it}(0)$ and $U_{it}(1)$ taking values in $\{0, 1\}$. These binary unobservables represent the outcome that would have been realized at time t had the past period outcome $Y_{i(t-1)}$ been exogenously manipulated to be 0 or 1, respectively.

[3]Note in particular the distinction with the dynamic treatment effects literature (e.g., Abbring and Heckman (2007), Angrist and Kuersteiner (2011)), in which the treatment and outcome variables are distinct.

The observed outcomes $Y_i \equiv (Y_{i0}, Y_{i1}, \dots, Y_{iT})$ together form a random vector with values in $\mathcal{Y} \equiv \{0,1\}^{T+1}$, the $(T+1)$-fold Cartesian product of $\{0,1\}$. The observed outcomes are related to potential outcomes $U_i(0) \equiv (U_{i1}(0), \dots, U_{iT}(0))$ and $U_i(1) \equiv (U_{i1}(1), \dots, U_{iT}(1))$ through the recursive relationship

$$Y_{it} = Y_{i(t-1)}U_{it}(1) + (1 - Y_{i(t-1)})U_{it}(0) = U_{it}(Y_{i(t-1)}) \quad \text{for all } t \geq 1. \tag{1}$$

In this formulation, the outcome in the initial period, Y_{i0}, is observed but not modeled. This avoids the initial conditions problem discussed by Heckman (1981c) by simply reducing the number of observed variables that are explicitly modeled, similar in spirit to the approach of Honoré and Tamer (2006) or Chen, Tamer, and Torgovitsky (2011).[4]

This specification presumes that the researcher is only interested in the causal effect of the outcome in the immediately preceding period on the outcome in the current period. In some settings, it may be interesting to analyze the causal effects of longer sequences of prior outcomes on the current period outcome. This can be accommodated by redefining the potential outcomes to include a separate potential outcome for every sequence up to a certain length. For clarity, I focus on the one-period causal effect in the main text and discuss this extension to longer sequences in Appendix S2 of the Supplemental Material. However, note that focusing on single period sequences in (1) does not place any restrictions on the temporal dependence of the potential outcomes. In particular, even though only first-order causal effects are being modeled, (1) *does not* imply that the potential outcomes follow a first-order Markov chain.

In addition to Y_i, the researcher also observes a vector $X_i = (X_{i0}, X_{i1}, \dots, X_{iT})$ of covariates with support $\mathcal{X}$. The components of X_{it} may be time-varying or time-invariant. I assume for simplicity that $\mathcal{X}$ is a finite set, so that X_i is discretely distributed.[5] Some of the components of X_{it} may be thought of as conditioning variables that describe observed heterogeneity, while others might be viewed as instruments that satisfy certain exclusion or monotonicity conditions. These types of assumptions are discussed in Section 4.5.

The DPO model captures state dependence through the possibility that $U_{it}(0) \neq U_{it}(1)$. That is, the outcome $Y_{it} = U_{it}(Y_{i(t-1)})$ that actually occurred for agent i in period t may have been different had $Y_{i(t-1)}$ been different. The DPO model allows for "occurrence," "duration," and "lagged duration" dependence, as defined by Heckman and Borjas (1980). It also allows for general forms of both observed and unobserved heterogeneity. Observed heterogeneity is captured through differences in the distributions of $(U_i(0), U_i(1))|X_i = x$ for different values of x. Unobserved heterogeneity is captured through variation in $(U_i(0), U_i(1))$, conditional on $X_i = x$. For example, the model allows for the possibility that, conditional on $X_i = x$, $U_{it}(1) - U_{it}(0)$ is a random variable taking values in $\{-1, 0, 1\}$ for agents that differ along unobservable characteristics such as preferences or private information. The basic DPO model does not separate this unobserved heterogeneity into persistent and transitory components, and so does not impose any restrictions on the serial dependence of the potential outcomes. In Section 4, I discuss several assumptions that can be imposed to create a permanent-transitory distinction.

[4]In particular, note that the DPO model *does not* impose independence between Y_{i0} and any of the subsequent potential outcomes. It is straightforward to add such a condition as an additional identifying assumption, but this is often difficult to justify (Heckman (1981c)), so I do not consider it in this paper.

[5]Continuous covariates do not present any conceptual difficulty for the identification analysis; see the discussion in, for example, Torgovitsky (2019a). However, as is usually the case in nonparametric analyses, they do complicate estimation and statistical inference, so for simplicity I focus on the discrete case.

2.2. *DPO Models Implied by Dynamic Choice Models*

In this section, I connect the DPO model to a discrete time dynamic choice (DC) model of a rational, forward-looking economic agent. This serves two purposes. First, the DC model will be used to motivate and interpret the additional identifying assumptions for the DPO model that are proposed in Section 4. Second, since the DC model nests standard "structural" and "reduced form" models as special cases, it provides a vehicle for comparing these models to the DPO model.

The DC model is as follows. Time runs from some initial period $\underline{T}$ that occurs at or before $t = 0$ to some terminal period $\bar{T}$ that occurs at or after T, where $\bar{T}$ may be either finite or infinite. In each period t, agent i chooses $C_{it} \equiv (Y_{it}, D_{it})$. One of these choice variables is the binary outcome, Y_{it}, that the researcher observes in periods $t = 0, 1, \dots, T$. The other choice variables, D_{it}, could take any number of values, and could be either observed or unobserved by the researcher.

Agent i receives flow utility in period t of $\mu(C_{it}, S_{it})$, where $S_{it} \equiv (C_{i(t-1)}, Z_{it})$ are state variables that may affect this utility. I assume throughout that the flow utility is bounded. The state variables include the previous period choices, $C_{i(t-1)}$, and an additional vector of exogenous state variables, Z_{it}, which could contain both observable and unobservable components.

Each agent maximizes their expected present-discounted utility using discount factor $\delta \in (0, 1)$.[6] Under mild regularity conditions (see, e.g., Stokey, Lucas, and Prescott (1989) or Rust (1994)), agent i's problem can be written recursively in terms of the Bellman equation

$$\nu(S_{it}) = \max_{c' \in \mathcal{C}} \left\{ \mu(c', S_{it}) + \delta \int \nu(c', z') \, d\Lambda(z' | S_{it}) \right\} \equiv \max_{c' \in \mathcal{C}} \mathring{\nu}(c', S_{it}), \tag{2}$$

where ν is the value function, $\mathcal{C} \equiv \{0, 1\} \times \mathcal{D}$ is the feasible set of choices, $\Lambda(\cdot|s)$ is a distribution function for $Z_{i(t+1)}$, conditional on $S_{it} = s$, and $\mathring{\nu}$ is shorthand notation that combines the flow utility and continuation value. The distribution function, Λ, captures the agent's beliefs about the evolution of the exogenous state variables, including deterministic laws of motion. Neither μ nor Λ has an i or a t subscript because all observable and unobservable differences across agents and time are viewed notationally as part of Z_{it}. So, for example, one component of Z_{it} could be the time period itself, t, which would allow for these functions to vary over time.

To compare this model to the DPO model, I will isolate the Y_{it} choice. Profile (2) as

$$\nu(S_{it}) = \max_{y' \in \{0,1\}} \max_{d' \in \mathcal{D}} \mathring{\nu}(y', d', S_{it}). \tag{3}$$

Suppose that there is a unique solution to the inner problem in (3) given the (possibly suboptimal) choice of y', and denote it as

$$d^\star(S_{it} \parallel y') \equiv \arg\max_{d' \in \mathcal{D}} \mathring{\nu}(y', d', S_{it}). \tag{4}$$

Then (3) can be written as

$$\nu(S_{it}) = \max_{y' \in \{0,1\}} \mathring{\nu}(y', d^\star(S_{it} \parallel y'), S_{it}) \equiv \max_{y' \in \{0,1\}} \mathring{\nu}(S_{it} \parallel y'). \tag{5}$$

[6]In all of the following, δ could be replaced by δ_i and allowed to vary over the population as long as $\delta_i \in (0, 1)$ with probability 1. In this case, one could treat δ_i as a component of Z_{it}.

The observed binary choice, Y_{it}, is assumed to be the optimizer of (5):

$$Y_{it} = \underset{y' \in \{0,1\}}{\arg\max}\, \mathring{\nu}\big(S_{it} \parallel y'\big) = \mathbb{1}\big[\Delta\mathring{\nu}(S_{it}) \geq 0\big], \tag{6}$$

where $\Delta\mathring{\nu}(S_{it}) \equiv \mathring{\nu}(S_{it} \parallel 1) - \mathring{\nu}(S_{it} \parallel 0)$, and ties are broken in favor of $Y_{it} = 1$.

Corresponding to the agent's actual choice, Y_{it}, are two counterfactual choices that they would have made in period t, had they actually chosen $y \in \{0, 1\}$ in period $t-1$. The counterfactual entertained here is that the agent also re-optimizes their choice of d at time $t-1$, given their choice of $Y_{i(t-1)} = y$. Thus, instead of $D_{i(t-1)}$, they choose $d^\star(S_{i(t-1)} \parallel y)$. The other state variables, Z_{it}, are presumed to remain the same in both counterfactual states.[7] Let $S_{it}(y) \equiv (y, d^\star(S_{i(t-1)} \parallel y), Z_{it})$ denote the state variables that would have been realized in period t had the agent chosen y in period $t-1$. Then the agent's choice in period t if they had chosen y in period $(t-1)$ can be written as

$$U_{it}(y) = \mathbb{1}\big[\Delta\mathring{\nu}\big(S_{it}(y)\big) \geq 0\big]. \tag{7}$$

Equation (7) shows that a DC model generates a DPO model.[8]

2.3. *DPO Models Implied by Dynamic Binary Response Models*

Dynamic binary response (DBR) models are commonly used for measuring state dependence.[9] A textbook version of the model (e.g., Wooldridge (2010, Section 15.8.4)) has the threshold-crossing equation

$$Y_{it} = \mathbb{1}\big[\beta_0 Y_{i(t-1)} + X_{it}'\beta_1 + A_i + V_{it} \geq 0\big] \quad \text{for } t \geq 1, \tag{8}$$

where (β_0, β_1) are unknown parameters, and the exogenous state variables $Z_{it} = (X_{it}, A_i, V_{it})$ consist of an observable component, X_{it}, an unobservable time-invariant component, A_i, and an unobservable time-varying component, V_{it}.[10] This model can be viewed as a special case of (6) in which $\Delta\mathring{\nu}(S_{it}) = \beta_0 Y_{i(t-1)} + X_{it}'\beta_1 + A_i + V_{it}$. The associated potential outcomes are special cases of (7):

$$U_{it}(y) = \mathbb{1}\big[\beta_0 y + X_{it}'\beta_1 + A_i + V_{it} \geq 0\big] \quad \text{for } y \in \{0, 1\} \text{ and } t \geq 1. \tag{9}$$

The textbook implementation of (8) maintains the assumption that X_i is independent of $(A_i, V_{i1}, \ldots, V_{iT})$, that A_i is normally distributed, and that V_{it} are normally (or logistically) distributed. However, there are also several known results concerning semi- and nonparametric modifications of this and similar models. These are surveyed in Appendix S3. One criticism of (8)–(9) is that the relationship between the primitives of the choice problem (2) and the specification of $\Delta\mathring{\nu}$ in (9) can be obscure.[11]

[7]This by itself is not restrictive since any components that change can be treated as part of D_{it}.

[8]Note that $U_{it}(y)$ as defined in (7) satisfies the law of motion (1), since $U_{it}(Y_{i(t-1)}) = \mathbb{1}[\Delta\mathring{\nu}(S_{it}(Y_{i(t-1)})) \geq 0] = \mathbb{1}[\Delta\mathring{\nu}(S_{it}) \geq 0] = Y_{it}$.

[9]Most of the empirical papers listed in the Introduction use some variety of DBR model. An early example is Heckman (1981a). Linear probability models are also occasionally used to analyze state dependence in binary outcomes (e.g., pp. 1265–1266 of Hyslop (1999)); however, they have highly undesirable properties when viewed as causal models (Manski and Pepper (2009, p. S210)).

[10]Additionally, one must address the determination of Y_{i0} to account for the initial conditions problem observed by Heckman (1981c). One popular way to do this, proposed by Wooldridge (2005), is to include the initial period outcome Y_{i0} as one of the observed state variables, and interpret inference as conditional on Y_{i0}.

[11]Although, see Hyslop (1999), who developed a choice model that approximately implies (8)–(9).

2.4. *Discussion*

The attraction of the "structural" DC model (6) is that it is derived directly from a model of choice behavior. However, starting with Rust (1994), it has been recognized that additional parametric assumptions must be maintained in order to point identify primitive features of this model; see also Magnac and Thesmar (2002). These assumptions include finitely parameterized functional forms for the flow utility, μ, as well as parametric distributions for unobserved components of the exogenous state variables. It is also commonly assumed that the observed exogenous state variables are independent of the unobserved components, that δ is fixed at a value known to the researcher, that agents have knowledge of the distribution of the unobserved state variables, and, frequently, that agents have perfect foresight or rational expectations over all observed exogenous state variables. Many of these assumptions are often questionable in applications.

Of particular concern are parametric assumptions about the distributions of unobservables. As Flinn and Heckman (1982, p. 132) observed, "It [economic theory] is silent on the topic of the correct specification of functional forms for the distributions of unobservables." More bluntly, there is typically little economic rationale for assuming that an unobservable state variable follows a normal distribution as opposed to a logistic, Gumbel, or mixture of normals, among many other possibilities. Even advocates of structural modeling recognize parametric functional form restrictions as undesirable and "extra-theoretic" (e.g., Keane, Todd, and Wolpin (2011, p. 452)).[12]

The DPO model is fully nonparametric, so does not suffer from this drawback. However, like the DBR model (8), the DPO model has a more tenuous connection with DC models. The approach taken in this paper is to view the DPO model as being generated by an underlying DC model through (7). In Section 4, I show that nonparametric assumptions on the DC model imply nonparametric assumptions on its generated DPO model. In Section 5, I use a specific job search model to motivate the assumptions used in the application to unemployment dynamics. While the DPO model can be used by itself without taking this view, the link back to the familiar DC model can be helpful for interpretation.

3. IDENTIFICATION

3.1. *Definitions*

In this section, I develop a general procedure for constructing identified sets in the DPO model. The procedure is abstract with respect to the assumptions and parameter of interest; concrete examples are discussed ahead. I assume throughout the analysis that the panel is balanced. Periods are indexed by $t = 0, 1, \dots, T$ for T small and fixed, and probabilities are taken over agents i drawn from the population.

The DPO model is (1). The primitive of the DPO model is a probability mass function P with support contained in $\mathcal{U} \times \mathcal{X}$, where $\mathcal{U} \equiv \{0, 1\}^{2T+1}$ is the collection of all possible realizations of $U_i \equiv (Y_{i0}, U_i(0), U_i(1))$. A function P with domain $\mathcal{U} \times \mathcal{X}$ is a probability

[12] Norets and Tang (2014) have made progress on rigorously characterizing the effect of such assumptions on identification for DC models. However, a completely distribution-free characterization remains a difficult problem even in static discrete choice models (Torgovitsky (2019a)). See also Heckman and Navarro (2007), who considered point identification in nonparametric dynamic choice models using exogenous variables with large support. More semi- and nonparametric results are available for DBR models like (9); see Appendix S3.

mass function on $\mathcal{U} \times \mathcal{X}$ if and only if it takes values in $[0, 1]$ and

$$\sum_{u \in \mathcal{U}, x \in \mathcal{X}} P(u, x) = 1. \tag{10}$$

Let $\mathcal{P}$ denote the set of all functions $P : \mathcal{U} \times \mathcal{X} \to [0, 1]$ that satisfy (10).

The *parameter space*, $\mathcal{P}^\dagger$, is the subset of $\mathcal{P}$ that satisfies the researcher's prior assumptions. Notationally, it is convenient to describe $\mathcal{P}^\dagger$ as $\mathcal{P}^\dagger = \{P \in \mathcal{P} : \rho(P) \geq 0\}$, where $\rho : \mathcal{P} \to \mathbb{R}^{d_\rho}$ is a function representing restrictions on P, and the inequality is interpreted component-wise. Equality restrictions can be incorporated into $\mathcal{P}^\dagger$ by including pairs of inequalities in the function ρ. The restrictions may depend on features of the observable distribution of (Y_i, X_i), but this is suppressed in the notation.

The *identified set*, $\mathcal{P}^\star$, is defined as the *observationally equivalent* subset of $\mathcal{P}^\dagger$ that could have generated the observed data through relationship (1). Let $\mathbb{P}[Y_i = \cdot, X_i = \cdot]$ denote the observable probability mass function of (Y_i, X_i), where $Y_i \equiv (Y_{i0}, Y_{i1}, \dots, Y_{iT})$. Then $P \in \mathcal{P}^\star$ requires that, for every $y \equiv (y_0, y_1, \dots, y_T) \in \mathcal{Y}$ and $x \in \mathcal{X}$,

$$\begin{aligned}\mathbb{P}[Y_i = y, X_i = x] &= \mathbb{P}_P[Y_i = y, X_i = x] \\ &= \mathbb{P}_P\big[Y_{i0} = y_0, U_{it}(y_{t-1}) = y_t \text{ all } t \geq 1, X_i = x\big],\end{aligned}$$

where $\mathbb{P}_P[\cdot]$ denotes the probability of an event when (U_i, X_i) is distributed according to P and Y_i is determined recursively through (1). This expression can be rewritten as a linear function of P:

$$\mathbb{P}[Y_i = y, X_i = x] = \sum_{u \in \mathcal{U}_{\text{oeq}}(y)} P(u, x), \tag{11}$$

where $\mathcal{U}_{\text{oeq}}(y)$ is the set of all $u \equiv (u_0, u_1(0), \dots, u_T(0), u_1(1), \dots, u_T(1)) \in \mathcal{U}$ for which $u_0 = y_0$ and $u_t(y_{t-1}) = y_t$ for all $t \geq 1$. Figure 1 illustrates (11) for $T = 2$.

Usually, a researcher is interested in a low-dimensional *target parameter*, that is, a low-dimensional function $\theta : \mathcal{P} \to \mathbb{R}^{d_\theta}$ of P. The researcher's primary object of interest is then the identified set for θ, denoted by $\Theta^\star \equiv \{\theta(P) : P \in \mathcal{P}^\star\}$. In the next section, I discuss some target parameters that provide useful measures of state dependence.

3.2. *Target Parameters for Measuring State Dependence*

A natural measure of state dependence is the proportion of agents that would have experienced a different outcome in period t had their outcome in period $t-1$ been different, that is, the proportion of agents for which $U_{it}(0) \neq U_{it}(1)$. These agents are represented by the events $[U_{it}(0) = 0, U_{it}(1) = 1]$ and $[U_{it}(0) = 1, U_{it}(1) = 0]$. The proportion of the first group under P is denoted by

$$\text{SD}_t^+(P) \equiv \mathbb{P}_P\big[U_{it}(0) = 0, U_{it}(1) = 1\big].$$

Agents in this first group can be said to experience positive state dependence, since an exogenous manipulation of their period $t-1$ outcome from 0 to 1 would cause an increase in their period t outcome from 0 to 1. The measure of the second group under P is denoted by

$$\text{SD}_t^-(P) \equiv \mathbb{P}_P\big[U_{it}(0) = 1, U_{it}(1) = 0\big].$$

Potential Outcomes					Observed Outcomes		
Y_{i0}	$U_{i1}(0)$	$U_{i1}(1)$	$U_{i2}(0)$	$U_{i2}(1)$	Y_{i0}	Y_{i1}	Y_{i2}
$\boxed{0}$	$\boxed{0}$	0	$\boxed{0}$	0	0	0	0
$\boxed{0}$	$\boxed{0}$	0	$\boxed{0}$	1	0	0	0
$\boxed{0}$	$\boxed{0}$	1	$\boxed{0}$	0	0	0	0
$\boxed{0}$	$\boxed{0}$	1	$\boxed{0}$	1	0	0	0
	⋮		⋮			⋮	
$\boxed{1}$	0	$\boxed{0}$	$\boxed{1}$	0	1	0	1
$\boxed{1}$	0	$\boxed{0}$	$\boxed{1}$	1	1	0	1
$\boxed{1}$	1	$\boxed{0}$	$\boxed{1}$	0	1	0	1
$\boxed{1}$	1	$\boxed{0}$	$\boxed{1}$	1	1	0	1
	⋮		⋮			⋮	

FIGURE 1.—Observational equivalence, $T = 2$. *Notes*: The full diagram would have $2^{2T+1} = 2^5 = 32$ rows corresponding to all possible realizations of the unobservables $U_i \equiv (Y_{i0}, U_{i1}(0), U_{i2}(0), U_{i1}(1), U_{i2}(1))$. Here, the rows shown are those corresponding to the potential outcomes that could generate $Y_i = (0, 0, 0)$ or $Y_i = (1, 0, 1)$ through the recursive relationship (1), that is, the elements of $\mathcal{U}_{\text{oeq}}(0, 0, 0)$ and $\mathcal{U}_{\text{oeq}}(1, 0, 1)$ in (11). The observed outcome of Y_i is determined by the potential outcomes that are in boxes, but is unaffected by the other potential outcomes.

This is the proportion of agents who can be said to experience negative state dependence. The total proportion of agents experiencing state dependence under P is

$$\text{SD}_t(P) \equiv \mathbb{P}_P\big[U_{it}(0) \neq U_{it}(1)\big] = \text{SD}_t^+(P) + \text{SD}_t^-(P).$$

The average treatment effect of $Y_{i(t-1)}$ on Y_{it} is defined as

$$\text{ATE}_t(P) \equiv \mathbb{E}_P\big[U_{it}(1) - U_{it}(0)\big],$$

where $\mathbb{E}_P$ denotes expectation taken with respect to P. This parameter is widely used to study state dependence; however, it confounds two effects. To see this, notice that the relationship between ATE_t, SD_t^+, and SD_t^- is given by

$$\begin{aligned} \text{ATE}_t(P) &\equiv \mathbb{P}_P\big[U_{it}(1) = 1\big] - \mathbb{P}_P\big[U_{it}(0) = 1\big] \\ &= \big(\mathbb{P}_P\big[U_{it}(1) = 1, U_{it}(0) = 0\big] + \mathbb{P}_P\big[U_{it}(1) = 1, U_{it}(0) = 1\big]\big) \\ &\quad - \big(\mathbb{P}_P\big[U_{it}(1) = 0, U_{it}(0) = 1\big] + \mathbb{P}_P\big[U_{it}(1) = 1, U_{it}(0) = 1\big]\big) \\ &= \text{SD}_t^+(P) - \text{SD}_t^-(P). \end{aligned} \tag{12}$$

Thus, ATE_t is the proportion of the population that experiences positive state dependence at time t, less the proportion that experiences negative state dependence.

An implication of (12) is that it is possible for ATE_t to be small or zero even if there is substantial positive and negative state dependence. In such cases, ATE_t may be a misleading measure of state dependence. For example, suppose that Y_{it} denotes welfare status as in Chay, Hoynes, and Hyslop (2004) or Card and Hyslop (2005). Then SD_t^- represents the "at risk" proportion of the population that would receive welfare in period t as a direct result of having not received it in the previous period, while SD_t^+ represents the proportion of the population that is in the "welfare trap." In contrast, ATE_t represents the difference in the sizes of the two groups, which is less interpretable. For this reason, I do not analyze ATE_t in this paper.

In many settings, it is interesting to consider modifying SD_t^+ and SD_t^- to be conditional on realizations of Y_{it}. For example, if Y_{it} is welfare status, a researcher may be interested in measuring positive state dependence among just the individuals currently receiving welfare, that is, those with $Y_{it} = 1$. This parameter is given by

$$\text{SD}_t^+(P|1) \equiv \mathbb{P}_P\big[U_{it}(0) = 0, U_{it}(1) = 1 | Y_{it} = 1\big].$$

Alternatively, in the application to employment dynamics in Section 5, I consider positive state dependence among the unemployed, which is given by

$$\text{SD}_t^+(P|0) \equiv \mathbb{P}_P\big[U_{it}(0) = 0, U_{it}(1) = 1 | Y_{it} = 0\big].$$

These parameters are analogous to the treatment on the (un)treated parameters commonly considered in the analysis of the static potential outcome models (see, e.g., Heckman and Vytlacil (2007)).

This type of conditioning can be extended a period further to define

$$\begin{aligned}
&\text{SD}_t^+(P|00) \equiv \mathbb{P}_P\big[U_{it}(0) = 0, U_{it}(1) = 1 | Y_{it} = 0, Y_{i(t-1)} = 0\big] \quad \text{and} \\
&\text{SD}_t^+(P|11) \equiv \mathbb{P}_P\big[U_{it}(0) = 0, U_{it}(1) = 1 | Y_{it} = 1, Y_{i(t-1)} = 1\big],
\end{aligned}$$

which quantify positive state dependence among agents whose state in the previous period is the same as in the current period. Like $\text{SD}_t^+(\cdot|0)$ and $\text{SD}_t^+(\cdot|1)$, these parameters have a treatment on the (un)treated interpretation. However, they can also be interpreted as the *proportion* of the observed persistence in outcomes that is due to state dependence.

To see this, consider the quantity $\mathbb{P}[Y_{it} = 0|Y_{i(t-1)} = 0]$ as a measure of the observed persistence in state 0. For an observationally equivalent P, this quantity can be decomposed as

$$\begin{aligned}
&\mathbb{P}[Y_{it} = 0|Y_{i(t-1)} = 0] \\
&\quad = \mathbb{P}_P\big[Y_{it} = 0, U_{it}(1) = 0|Y_{i(t-1)} = 0\big] + \mathbb{P}_P\big[Y_{it} = 0, U_{it}(1) = 1|Y_{i(t-1)} = 0\big] \\
&\quad = \mathbb{P}_P\big[U_{it}(0) = 0, U_{it}(1) = 0|Y_{i(t-1)} = 0\big] + \mathbb{P}_P\big[U_{it}(0) = 0, U_{it}(1) = 1|Y_{i(t-1)} = 0\big].
\end{aligned}$$

The second term is the contribution to $\mathbb{P}[Y_{it} = 0|Y_{i(t-1)} = 0]$ that is due to positive state dependence. The size of this quantity as a proportion of the observed persistence is given by

$$\frac{\mathbb{P}_P\big[Y_{it} = 0, U_{it}(1) = 1|Y_{i(t-1)} = 0\big]}{\mathbb{P}[Y_{it} = 0|Y_{i(t-1)} = 0]} = \mathbb{P}_P\big[U_{it}(1) = 1|Y_{it} = 0, Y_{i(t-1)} = 0\big] = \text{SD}_t^+(P|00),$$

where the second equality follows because $[Y_{i(t-1)} = 0, Y_{it} = 0]$ implies $[U_{it}(0) = 0]$. A similar argument shows that $\text{SD}_t^+(\cdot|11)$ can be interpreted as the proportion of the observed persistence in state 1 that is due to positive state dependence. These parameters constitute a natural rubric for measuring the role of state dependence in the persistence of observed outcomes.

3.3. *Empirical-Evidence-Only Bounds*

The data alone do not provide enough information to point identify SD_t^+ or SD_t^-. The reasons are the same as in a static potential outcomes model. First, an analyst never observes both $U_{it}(0)$ and $U_{it}(1)$, since only $Y_{it} = U_{it}(Y_{i(t-1)})$ is observed. Thus, quantities like SD_t^+ which concern the joint distribution of $(U_{it}(0), U_{it}(1))$ are inherently not point identified (see, e.g., Heckman, Smith, and Clements (1997)). Second, even the marginal distributions of $U_{it}(0)$ and $U_{it}(1)$ will typically not be point identified due to the endogeneity of prior outcomes. That is, in general we expect that for observationally equivalent P,

$$\mathbb{P}[Y_{it} = 1 | Y_{i(t-1)} = 1, X_i] = \mathbb{P}_P\big[U_{it}(1) = 1 | Y_{i(t-1)} = 1, X_i\big] \neq \mathbb{P}_P\big[U_{it}(1) = 1 | X_i\big], \tag{13}$$

since $Y_{i(t-1)} = 1$ depends on $(U_{i(t-1)}(0), U_{i(t-1)}(1))$, and $U_{it}(1)$ is likely dependent with $(U_{i(t-1)}(0), U_{i(t-1)}(1))$, even conditional on X_i, due to persistent latent heterogeneity.

While SD_t^+ and SD_t^- are not point identified, they are not completely unconstrained by the data. The next proposition provides sharp bounds on SD_t^+, SD_t^-, and SD_t that use only the empirical evidence. All proofs are contained in Appendix S4.

PROPOSITION 1: *Suppose that $\mathcal{P}^\dagger = \mathcal{P}$. If $\theta = \mathrm{SD}_t^+$, then*

$$\Theta^\star = \big[0, \mathbb{P}[Y_{i(t-1)} = 0, Y_{it} = 0] + \mathbb{P}[Y_{i(t-1)} = 1, Y_{it} = 1]\big]. \tag{14}$$

If $\theta = \mathrm{SD}_t^-$, then

$$\Theta^\star = \big[0, \mathbb{P}[Y_{i(t-1)} = 0, Y_{it} = 1] + \mathbb{P}[Y_{i(t-1)} = 1, Y_{it} = 0]\big]. \tag{15}$$

If $\theta = \mathrm{SD}_t$, then $\Theta^\star = [0, 1]$.

The intuition behind the bounds in (14) is as follows. The target parameter is SD_t^+, which is the proportion of individuals with $U_{it}(0) = 0$ and $U_{it}(1) = 1$. Individuals with $Y_{i(t-1)} = 0$ and $Y_{it} = 1$ cannot have these potential outcomes, since they must have $U_{it}(0) = 1$. Similarly, individuals with $Y_{i(t-1)} = 1$ and $Y_{it} = 0$ have $U_{it}(1) = 0$, so they also do not contribute to SD_t^+. So, both of these observed groups can be removed from the calculation in (14).

The remaining observed groups are those with $Y_{i(t-1)} = Y_{it}$. Among individuals with $Y_{i(t-1)} = 0$ and $Y_{it} = 0$, there are those who have positive state dependence ($U_{it}(1) = 1$), and those who would have been in state 0 regardless ($U_{it}(1) = 0$). Similarly, the group with $Y_{i(t-1)} = 1$ and $Y_{it} = 1$ consists of individuals both with and without positive state dependence. The upper bound in (14) is obtained when all individuals in both observed groups exhibit positive state dependence, while the lower bound is obtained when none do. Analogous reasoning leads to the bounds in (15).

The lower bounds in (14)–(15) are always 0, regardless of the distribution of the data. Thus, the empirical evidence alone never enables a rejection of the hypothesis that there is no positive or negative state dependence. The upper bound in (14) will be large when the observed outcomes have strong positive serial dependence. Similarly, the upper bound in (15) will be large when the observed outcomes have strong negative serial dependence. Bounds on SD_t^+ will therefore tend to be wide when bounds on SD_t^- are narrow, and vice versa.

In fact, the third finding of Proposition 1 is that these upper bounds perfectly offset each other, so that the (sharp) identified set for state dependence of both sorts, SD_t, is always the entire logically possible interval [0, 1]. Thus, empirical evidence alone cannot discriminate between the hypothesis that there is no state dependence of any sort, with all persistence caused by latent heterogeneity, and the reverse, that there is no latent heterogeneity and all persistence in the data is due to state dependence. The existence or non-existence of state dependence can only be established by incorporating additional identifying assumptions, such as those discussed ahead.

3.4. *Computing Identified Sets*

Proposition 1 was proven using a standard two-step argument. First, one proposes bounds. Second, one argues that there are parameter values for which these bounds are obtained. This strategy provides analytic expressions, which can be useful both for intuition and for statistical inference. However, the argument becomes increasingly complicated as the parameter space becomes increasingly complex.[13] Yet, the takeaway from Proposition 1 was that we need to impose more assumptions in order to obtain interesting empirical conclusions. Since more assumptions typically make the parameter space more complex, this creates a problem.

One way to resolve this problem is to recognize that when θ is scalar-valued, its identified set can usually be determined by solving two optimization problems.[14]

PROPOSITION 2: *Suppose that $\mathcal{P}^\dagger$ is closed and convex, and that θ is a continuous, scalar-valued function of P. Then, as long as $\mathcal{P}^\star$ is nonempty, $\Theta^\star = [\theta^\star_{\text{lb}}, \theta^\star_{\text{ub}}]$, where*

$$\theta^\star_{\text{lb}} \equiv \min_{P\in\mathcal{P}^\star} \theta(P) = \min_{\{P(u,x)\in[0,1]:u\in\mathcal{U},x\in\mathcal{X}\}} \theta(P) \quad \text{s.t. } \rho(P) \geq 0, (10), \text{and } (11)\ \forall y, x \quad \text{and}$$

$$\theta^\star_{\text{ub}} \equiv \max_{P\in\mathcal{P}^\star} \theta(P) = \max_{\{P(u,x)\in[0,1]:u\in\mathcal{U},x\in\mathcal{X}\}} \theta(P) \quad \text{s.t. } \rho(P) \geq 0, (10), \text{and } (11)\ \forall y, x.$$

Proposition 2 provides a computational approach to characterizing the sharp identified set for θ. The feasibility of this approach depends on how difficult it is to solve the problems that define $\theta^\star_{\text{lb}}$ and $\theta^\star_{\text{ub}}$. Observe that (11) places linear restrictions on $P = \{P(u, x) : u \in \mathcal{U}, x \in \mathcal{X}\}$. The requirement that $P \in \mathcal{P}$ also places linear restrictions on P, namely, (10) and $1 \geq P(u, x) \geq 0$ for all $u \in \mathcal{U}$, $x \in \mathcal{X}$. Thus, if ρ and θ are linear, then the two optimization problems in Proposition 2 are linear programs. This means that Proposition 2 can be applied even in high dimensions as long as we limit attention to parameters and assumptions that can be expressed as linear functions of P.

Linearity turns out to be not very restrictive in the sense that it still permits a wide range of interesting parameters and assumptions. In Appendix S5, I show that each of the target parameters discussed in Section 3.2 is a linear function of P. When considering additional prior assumptions in Section 4, I will restrict attention to those that are linear due to computational considerations. This is not essential to Proposition 2, which applies more generally, but it is important for practical implementation.

[13]For example, compare the analysis in Okumura and Usui (2014) to that in Manski and Pepper (2000) and Manski (1994), or the analysis of Mourifié (2015) to that of Shaikh and Vytlacil (2011).

[14]This general point about partial identification analysis has been appreciated (sometimes implicitly) by many previous authors, including Honoré and Tamer (2006), Manski (2007), Molinari (2008), Chiburis (2010), Kitamura and Stoye (2013, 2018), Manski (2014), Freyberger and Horowitz (2015), and Lafférs (2013, 2018). In particular, Lafférs (2013, 2018) used a similar computational strategy as in this paper, but for a static potential outcomes model; see also the subsequent work by Demuynck (2015). The benefits in the static setting are smaller than in the dynamic case considered here, since a large number of analytic partial identification results already exist for static potential outcomes models. The representation of bounds in terms of linear programming dates back to at least Balke and Pearl (1994, 1997) for similar problems in causal inference, or to Hansen, Heaton, and Luttmer (1995) for different problems in finance.

4. IDENTIFYING ASSUMPTIONS FOR THE DPO MODEL

The empirical-evidence-only identified sets derived in Proposition 1 are wide. In this section, I propose identifying assumptions that can be added to the DPO model to narrow these bounds. These assumptions are implemented by including restrictions in the ρ function of Section 3 and then applying Proposition 2. Any subset of these assumptions can be combined by simply adding or removing the appropriate restrictions. All assumptions could be modified to be conditional on covariates (X_i). I keep this implicit for the sake of notation, but indicate situations in which conditioning may be important.[15] Along the way, I motivate and interpret the assumptions in the context of the DC and DBR models discussed in Section 2.

4.1. *Stationarity*

Stationarity assumptions are ubiquitous in panel data models. Indeed, combining empirical evidence from different time periods *requires* an assumption that the past shares at least some features in common with the future. In the DPO model, one stationarity assumption is that the joint distribution of $(U_{it}(0), U_{it}(1))$ is invariant across $t \geq 1$. A stronger form of stationarity uses multiple time periods, for example, that the distribution of $(U_{i(t-1)}(0), U_{it}(0), U_{i(t-1)}(1), U_{it}(1))$ does not vary across $t \geq 2$. The following is a general version.

ASSUMPTION ST: *Let m be a nonnegative integer chosen by the researcher and define $U_{i(t-m:t)}(0) \equiv (U_{i(t-m)}(0), \dots, U_{it}(0))$ and $U_{i(t-m:t)}(1) \equiv (U_{i(t-m)}(1), \dots, U_{it}(1))$ for $t \geq m+1$. For any $u \equiv (u_m(0), u_m(1)) \in \{0,1\}^{2(m+1)}$, define*

$$\Sigma^u_{t,m}(P) \equiv \mathbb{P}_P\big[U_{i(t-m:t)}(0) = u_m(0), U_{i(t-m:t)}(1) = u_m(1)\big].$$

Then, for any $P \in \mathcal{P}^\dagger$, every $u \in \{0,1\}^{2(m+1)}$, and every $t, t' \geq m+1$,

$$\Sigma^u_{t,m}(P) = \Sigma^u_{t',m}(P). \tag{16}$$

Distributions of potential outcomes that satisfy Assumption ST do not need to generate distributions of observed outcomes that are stationary. To see this, first observe that, for any P,

$$\begin{aligned}\mathbb{P}_P[Y_{it} = 0] &= \mathbb{P}_P\big[U_{it}(0) = U_{it}(1), Y_{it} = 0\big] + \mathbb{P}_P\big[U_{it}(0) \neq U_{it}(1), Y_{it} = 0\big] \\ &= \mathbb{P}_P\big[U_{it}(0) = 0, U_{it}(1) = 0\big] + \mathbb{P}_P\big[U_{it}(0) \neq U_{it}(1), Y_{it} = 0\big],\end{aligned} \tag{17}$$

where the second equality follows because $\mathbb{P}_P[U_{it}(0) = U_{it}(1) = 1, Y_{it} = 0] = 0$ by (1). Thus, if P satisfies Assumption ST, then from (17) there is marginal stationarity in the observed outcomes, that is, $\mathbb{P}_P[Y_{it} = 0] = \mathbb{P}_P[Y_{i(t-1)} = 0]$, if and only if

$$\mathbb{P}_P\big[U_{it}(0) \neq U_{it}(1), Y_{it} = 0\big] = \mathbb{P}_P\big[U_{i(t-1)}(0) \neq U_{i(t-1)}(1), Y_{i(t-1)} = 0\big]. \tag{18}$$

This restriction is not generally implied by Assumption ST. For (18) to hold would require a condition about the serial dependence of the potential outcomes at all previous lags, whereas Assumption ST does not restrict this dependence.

[15]Note that including rich covariate specifications quickly increases the dimension of the problems in Proposition 2. Three dimension reduction strategies are discussed in Appendix S6.

However, one case in which (18) does need to be true is when there is no state dependence, so that $\mathrm{SD}_{t'}(P) \equiv \mathbb{P}_P[U_{t'}(0) \neq U_{t'}(1)] = 0$ for $t' = t-1, t$. An important implication is that if Assumption ST is maintained, and if the distribution of observed outcomes is *not* stationary, then it is possible to bound the identified set for SD_t away from 0. Intuitively, in order for a DPO model that satisfies Assumption ST to generate a stationary distribution of observed outcomes, it must be the case that there is also no state dependence. To the extent that the observed outcome sequence is in fact nonstationary, one can therefore rule out the hypothesis that there is no state dependence.[16]

The potential outcomes implied by the DC model through (7) depend on $\Delta\mathring{\nu}(S_{it}(y))$. Whether $\Delta\mathring{\nu}(S_{it}(y))$ is stationary depends on the form of $\Delta\mathring{\nu}$ and the composition of the counterfactual state variables $S_{it}(y) \equiv (y, d^\star(S_{i(t-1)} \parallel y), Z_{it})$. Certainly, any time-invariant state variable is trivially stationary. It is also standard in empirical implementations of DC models to assume that the unobservable state variables are stationary. However, it may be undesirable to assume that time-varying components of the observed state variables are stationary, since this could be directly rejected by the data. In these cases, one can impose a version of Assumption ST that conditions on these variables.

Another conceptual issue highlighted by the DC model is the distinction between finite and infinite horizons. A finite horizon assumption can be viewed in terms of an infinite horizon problem by defining the flow utility to be 0 after the finite horizon has elapsed. This can be captured by an "age" state variable in Z_{it} that records the agent's location in their finite horizon. The value function—and therefore the potential outcomes $U_{it}(y)$—cannot be unconditionally stationary with a finite horizon, since it becomes identically 0 after a given time period. If a finite horizon is empirically important, then Assumption ST can be modified to be conditional on age.

4.2. *Diminishing Serial Correlation*

Persistent heterogeneity may cause potential outcomes to be positively serially correlated. If transitory heterogeneity is also present, then it may be natural to assume that this serial correlation is strongest between potential outcomes in adjacent periods and diminishes (or does not increase) as the distance between any two periods increases. This is the content of the following assumption.

ASSUMPTION DSC: *For every* $P \in \mathcal{P}^\dagger$, *and each* $y \in \{0, 1\}$, $\mathrm{Corr}_P(U_{it}(y), U_{i(t+t')}(y))$ *is decreasing in* $|t'|$ *for* $t' \in \{1-t, \dots, T-t\}$.

Assumption DSC places a nonlinear restriction on P. However, if Assumption ST holds (with any $m \geq 0$), then Assumption DSC becomes a linear restriction, equivalent to the statement that $\mathbb{P}_P[U_{it}(y) = 1, U_{i(t+t')}(y) = 1]$ is decreasing in $|t'|$ for $t' \in \{1-t, \dots, T-t\}$.[17] Due to this computational consideration, I will only consider imposing Assumption DSC when it is combined with Assumption ST.

The following proposition provides sufficient conditions for Assumption DSC when potential outcomes are generated by a DC model through (7).

[16] A similar observation was used by Heckman (1981b, p. 159) to establish point identification of β_0 in the parametric DBR model (8).

[17] This statement is justified in Appendix S5. If Assumption ST does not hold, then the statement that $\mathbb{P}_P[U_{it}(y) = 1, U_{i(t+t')}(y) = 1]$ is decreasing in $|t'|$ is equivalent to the statement that $(U_{it}(y), U_{i(t+t')}(y))$ is decreasing in the upper orthant order with respect to $|t'|$; see, for example, Shaked and Shanthikumar (2007, Section 6.G). The upper orthant order does not necessarily have a clear interpretation as a positive dependence concept, so imposing this condition directly does not seem attractive.

PROPOSITION 3: *Suppose that Assumption* ST *holds with any* $m \geq 0$. *Then Assumption* DSC *is satisfied if every* $P \in \mathcal{P}^{\dagger}$ *is consistent with the following conditions for each fixed* $y \in \{0, 1\}$: (i) $U_{it}(y)$ *is determined by* (7); (ii) *there are time-invariant random variables*, $\bar{S}_i$, *scalar time-varying random variables*, $\tilde{S}_{it}$, *and a function*, φ, *such that* $\Delta\mathring{\nu}(S_{it}(y)) = \varphi(\bar{S}_i, \tilde{S}_{it})$, *where* $\varphi(\bar{s}, \cdot)$ *is weakly increasing and right-continuous for each* $\bar{s}$; *and* (iii) $\{\tilde{S}_{it}\}_{t=1}^{T}|\bar{S}_i$ *is a first-order Markov chain with* $\mathbb{P}[\tilde{S}_{it} \leq \tilde{s}_t|\tilde{S}_{i(t-1)} = \tilde{s}_{t-1}, \bar{S}_i = \bar{s}]$ *weakly decreasing in* $\tilde{s}_{t-1}$ *for all* $\tilde{s}_t$ *and* $\bar{s}$.

Proposition 3 applies to the DBR model, (9), by taking $\bar{S}_i = A_i$, $\tilde{S}_{it} = X_{it}'\beta_1 + V_{it}$, and $\varphi(\bar{s}, \tilde{s}) \equiv \beta_0 y + \bar{s} + \tilde{s}$. The sufficient condition for Assumption DSC is that $X_{it}'\beta_1 + V_{it}$ is a first-order Markov chain with a stochastically increasing transition distribution. If there are no covariates ($\beta_1 = 0$), and $(V_{it}, V_{i(t+1)})$ are jointly normal (conditional on A_i), then the stochastic increasing assumption is equivalent to the correlation between V_{it} and $V_{i(t+1)}$ (given A_i) being nonnegative.

4.3. *Monotone Treatment Selection*

For a static model, Manski and Pepper (2000) considered the identifying content of assuming that potential outcomes are greater for agents who select into treatment than for those who do not. This monotone treatment selection (MTS) condition captures the idea that a researcher may be willing to make a prior assumption on the direction of bias that would arise from a simple treatment–control contrast. The following is a similar assumption for the DPO model.

ASSUMPTION MTS: *Every* $P \in \mathcal{P}^{\dagger}$ *satisfies*

$$\mathbb{P}_P\big[U_{it}(y) = 1|Y_{i(t-1)} = 1, Y_{i(t-2)} = \tilde{y}\big] \geq \mathbb{P}_P\big[U_{it}(y) = 1|Y_{i(t-1)} = 0, Y_{i(t-2)} = \tilde{y}\big] \tag{19}$$

for $y = 0, 1$, $\tilde{y} = 0, 1$, *and all* $t \geq 2$ *such that* $\mathbb{P}[Y_{i(t-1)} = 1|Y_{i(t-2)} = \tilde{y}] \in (0, 1)$.

Assumption MTS says that those with $Y_{i(t-1)} = 1$ would be more likely to have $Y_{it} = 1$ than those with $Y_{i(t-1)} = 0$, even if their outcomes in period $t-1$ were exogenously manipulated from 0 to 1 or vice versa. Stated differently, the assumption is that agents with $Y_{i(t-1)} = 1$ have a higher latent propensity to be in state 1 in period t than agents with $Y_{i(t-1)} = 0$. The additional conditioning on $Y_{i(t-2)} = y_{t-2}$ in these statements ensures that the outcome in year $t-1$ is comparable. That is, since the event $[Y_{i(t-1)} = y_{t-1}, Y_{i(t-2)} = y_{t-2}]$ is equivalent to the event $[U_{i(t-1)}(y_{t-2}) = y_{t-1}, Y_{i(t-2)} = y_{t-2}]$, conditioning on $Y_{i(t-2)} = y_{t-2}$ ensures that the conditioning events on the left and right sides of (19) are expressed in terms of the same potential outcome $U_{i(t-1)}(y_{t-2})$. A stronger form of Assumption MTS extends this conditioning back to period $t-q$.

ASSUMPTION MTS—Generalization: *Let* $q \geq 2$ *be an integer chosen by the analyst. For each* t, *if* $q < t$, *then let* $Y_{i(t-q):(t-2)} \equiv (Y_{i(t-q)}, \dots, Y_{i(t-2)})$; *otherwise, let* $Y_{i(t-q):(t-2)} \equiv (Y_0, \dots, Y_{t-2})$. *Every* $P \in \mathcal{P}^{\dagger}$ *satisfies*

$$\begin{aligned}&\mathbb{P}_P\big[U_{it}(y) = 1|Y_{i(t-1)} = 1, Y_{i(t-q):(t-2)} = y_{\text{past}}\big]\\ &\quad \geq \mathbb{P}_P\big[U_{it}(y) = 1|Y_{i(t-1)} = 0, Y_{i(t-q):(t-2)} = y_{\text{past}}\big]\end{aligned} \tag{20}$$

for $y = 0, 1$, $y_{\text{past}} \in \{0, 1\}^{\min\{q,t\}-1}$, *and all* $t \geq 2$ *such that* $\mathbb{P}[Y_{i(t-1)} = 1|Y_{i(t-q):(t-2)} = y_{\text{past}}] \in (0, 1)$.

When potential outcomes are generated by the DC model (7), a sufficient condition for Assumption MTS is that $\Delta\mathring{\nu}(S_{it}(y))$ and $\Delta\mathring{\nu}(S_{it}(\tilde{y}))$ exhibit local positive quadrant dependence for each $(y, \tilde{y})$.

PROPOSITION 4: *Assumption* MTS *is satisfied with* $q = 2$ *if every* $P \in \mathcal{P}^{\dagger}$ *is consistent with the following conditions for each* $(y, \tilde{y}) \in \{0, 1\}^2$: (i) $U_{it}(y)$ *is determined by* (7); (ii) $\Delta\mathring{\nu}(S_{it}(y))$ *and* $\Delta\mathring{\nu}(S_{it}(\tilde{y}))$ *are positive quadrant dependent at* $(0, 0)$, *conditional on* $Y_{i(t-2)} = \tilde{y}$, *that is*,

$$\begin{aligned}&\mathbb{P}\big[\Delta\mathring{\nu}\big(S_{it}(y)\big) \geq 0, \Delta\mathring{\nu}\big(S_{i(t-1)}(\tilde{y})\big) \geq 0 \mid Y_{i(t-2)} = \tilde{y}\big] \\ &\quad \geq \mathbb{P}\big[\Delta\mathring{\nu}\big(S_{it}(y)\big) \geq 0 \mid Y_{i(t-2)} = \tilde{y}\big]\mathbb{P}\big[\Delta\mathring{\nu}\big(S_{i(t-1)}(\tilde{y})\big) \geq 0 \mid Y_{i(t-2)} = \tilde{y}\big]. \qquad (21)\end{aligned}$$

Condition (ii) depends on the structure of $\Delta\mathring{\nu}$, as well as the composition and relationships among the state variables, $S_{it}(y)$. In Section 5, I discuss a job search model in which this condition can be made more primitive. The effective requirement is that there is positive dependence among the determinants of being in state 1.

In the DBR model (8), Assumption MTS is satisfied if $(X_{it}'\beta_1 + A_i + V_{it})$ and $(X_{i(t-1)}'\beta_1 + A_i + V_{i(t-1)})$ are locally positive quadrant dependent, conditional on $Y_{i(t-2)}$. If there are no covariates ($\beta_1 = 0$), and if $(V_{it}, V_{i(t-1)})$ and A_i are independent and normally distributed conditional on $Y_{i(t-2)}$, then this will be the case if V_{it} and $V_{i(t-1)}$ are weakly positively correlated, or even negatively correlated, so long as the magnitude of the covariance between V_{it} and $V_{i(t-1)}$ is smaller than the variance of A_i.

4.4. *Fixed Effects*

Another way to introduce a permanent-transitory distinction in unobserved heterogeneity is the following assumption, introduced by Chernozhukov et al. (2013). Those authors described it as "time is randomly assigned" or "time is an instrument" (TIV).

ASSUMPTION TIV: *Let* $U_{it} \equiv (U_{it}(0), U_{it}(1))$. *For every* $P \in \mathcal{P}^{\dagger}$, *there exists a random variable* A_i *such that if* U_i *is distributed according to* P *and* Y_i *is generated by* (1), *then*

$$\mathbb{P}[U_{it} = u | Y_{i(t-1)}, \ldots, Y_{i1}, Y_{i0}, A_i] = \mathbb{P}[U_{i1} = u | Y_{i0}, A_i] \quad (\textit{almost surely})$$

for all $u \in \{0, 1\}^2$ *and all* $t \geq 2$.

Assumption TIV implies that all persistent unobservable heterogeneity is captured by the time-invariant latent random variable A_i, which can be interpreted as a generalization of a fixed effect. After accounting for A_i and the initial state Y_{i0}, current potential outcomes are required to be independent of past realized outcomes.

Although A_i has no meaning itself within the DPO model, Assumption TIV implies many restrictions on the distribution of potential outcomes.

PROPOSITION 5: *Let* $Y_{i(0:t)} \equiv (Y_{i0}, Y_{i1}, \ldots, Y_{it})$. *If Assumption* TIV *holds, then, for every* $P \in \mathcal{P}^{\dagger}$, *and every* $t' > t \geq 1$,

$$\mathbb{P}_P[U_{it'} = u, Y_{i(0:t-1)} = y] = \mathbb{P}_P[U_{it} = u, Y_{i(0:t-1)} = y] \qquad (22)$$

for all $u \in \{0, 1\}^2$ *and* $y \in \{0, 1\}^t$. *As a consequence, Assumption* TIV *implies Assumption* ST *with* $m = 0$.

I have been unable to determine whether the converse of Proposition 5 is also true.[18] If it is not, then imposing (22) might yield a non-sharp identified set. In the application in Section 5, I test the implications of TIV in Proposition 5 and overwhelmingly reject them.

When potential outcomes are generated by the DC model, the following conditions are sufficient for Assumption TIV and therefore (22).

PROPOSITION 6: *Assumption* TIV *is satisfied if every* $P \in \mathcal{P}^\dagger$ *is consistent with the following conditions*: (i) $U_{it}(y)$ *is determined by* (7); *and* (ii) *the stochastic components of* $(S_{it}(0), S_{it}(1))$ *can be split into time-invariant components*, $\bar{S}_i$, *and time-varying components*, $\tilde{S}_{it}$, *in such a way that the distribution of* $\tilde{S}_{it}|\tilde{S}_{i(1:t-1)}, \bar{S}_i, Y_{i0}$ *is the same as that of* $\tilde{S}_{i1}|\bar{S}_i, Y_{i0}$ *for every* $t \geq 2$.

4.5. *Monotone Instrumental Variables*

Instrumental variables (IV) can be used by assuming that an observed state variable (the instrument) is independent of potential outcomes. The monotone instrumental variable (MIV) assumption introduced by Manski and Pepper (2000, 2009) weakens this assumption to only restrict the sign of the relationship between the potential outcomes and instrument. The following is one way to adapt the MIV assumption to the DPO model.

ASSUMPTION MIV: *Let* X^0_{it} *and* X^1_{it} *be subvectors of* X_i, *where* X^1_{it} *takes values in a partially ordered set*. *Every* $P \in \mathcal{P}^\dagger$ *is such that* $\mathbb{P}_P[U_{it}(y) = 1|X^0_{it} = x^0, X^1_{it} = x^1]$ *is weakly increasing in* x^1 *for every* x^0, *each* $y = 0, 1$, *and every* $t \geq 1$.

Of course, weakly increasing can be changed to weakly decreasing as appropriate. Assumption MIV can be strengthened to a full IV assumption by imposing both directions of weak monotonicity together.

The next proposition provides sufficient conditions for Assumption MIV when potential outcomes are generated by the DC model.

PROPOSITION 7: *Assumption* MIV *is satisfied if every* $P \in \mathcal{P}^\dagger$ *is consistent with the following conditions for each fixed* $y \in \{0, 1\}$: (i) $U_{it}(y)$ *is determined by* (7); (ii) $S_{it}(y)$ *can be partitioned as* $S_{it}(y) = (X^0_{it}, X^1_{it}, V_{it})$, *where* X^0_{it} *and* X^1_{it} *are observed to the researcher*, V_{it} *is unobserved*, *and* X^1_{it} *takes values in a partially ordered set*; (iii) X^1_{it} *is independent of* V_{it}, *conditional on* X^0_{it}; *and* (iv) $\Delta\mathring{\nu}(S_{it}(y))$ *can be written as* $\Delta\mathring{\nu}(S_{it}(y)) = \varphi(X^0_{it}, X^1_{it}, V_{it})$ *for a function* φ *that is increasing in* X^1_{it} *for all fixed* X^0_{it} *and* V_{it}.

The textbook implementation of the DBR model (8) (e.g., Wooldridge (2010, Section 15.8.4)) assumes that A_i is independent of X_{it}, and V_{it} is independent of (X_{it}, A_i) with (known) distribution Φ. In this case,

$$\mathbb{P}\big[U_{it}(y) = 1|X_{it} = x\big] = \mathbb{E}\big[\Phi\big(\beta_0 y + x'\beta_1 + A_i\big)\big]. \tag{23}$$

Thus, the DBR model implies that $\mathbb{P}[U_{it}(y) = 1|X_{it} = x]$ is monotone increasing in a component of x if the sign of the corresponding component of β_1 is positive. Assumption MIV amounts to placing a sign restriction on a component of β_1. Imposing Assumption MIV in both directions—that is, assuming that a particular component of β_1 is 0—corresponds to an exclusion restriction. Exclusion restrictions like these are often imposed in applications of DBR models by not including various leads and lags of the time-varying observables.

[18]That is, if P satisfies the condition in Proposition 5, does there exist a random variable A_i such that P satisfies the condition in Assumption TIV?

4.6. *Monotone Treatment Response*

In some applications, it may make sense to assume that state dependence is either positive or negative. This can be viewed as a dynamic version of Manski's (1997) monotone treatment response (MTR) assumption.

ASSUMPTION MTR: *Every* $P \in \mathcal{P}^{\dagger}$ *satisfies* $\mathbb{P}_P[U_{it}(1) \geq U_{it}(0)] = 1$ *for all* t.

If potential outcomes are determined by the DC model through (7), a sufficient condition for Assumption MTR is that $\Delta\mathring{\nu}(S_{it}(1)) \geq \Delta\mathring{\nu}(S_{it}(0))$ almost surely. Either this condition or its opposite is always satisfied in the special case of the DBR model (8)–(9), since the parameter β_0 on lagged outcomes is deterministic.

A weaker version of Assumption MTR only imposes monotonicity "on average."

ASSUMPTION MATR: *Every* $P \in \mathcal{P}^{\dagger}$ *satisfies* $\mathbb{P}_P[U_{it}(1) = 1] \geq \mathbb{P}_P[U_{it}(0) = 1]$ *for all* t.

Assumption MATR is equivalent to the assumption that $\text{ATE}_t(P)$ is positive. In light of (12), this is equivalent to the assumption that there is more positive state dependence than there is negative state dependence. For the DPO model (7), a sufficient and necessary condition for Assumption MATR is that $\Delta\mathring{\nu}(S_{it}(1))$ is more likely than $\Delta\mathring{\nu}(S_{it}(0))$ to be greater than 0.

5. THE UNEMPLOYMENT DYNAMICS OF HIGH SCHOOL EDUCATED MEN

5.1. *Background and Motivation*

The increase of long-term unemployment in the wake of the Great Recession has rejuvenated interest in potential state dependence in employment outcomes. There is an extensive literature that estimates parametric DBR models with European employment data, for example, Narendranathan and Elias (1993), Mühleisen and Zimmermann (1994), Arulampalam, Booth, and Taylor (2000), and Tumino (2015). This research typically finds substantial evidence of state dependence. On the other hand, the comparably few studies that used similar methods with U.S. data, such as Ellwood (1982) and Corcoran and Hill (1985), find little or no evidence of state dependence.

A recent line of field experiments (Oberholzer-Gee (2008), Kroft, Lange, and Notowidigdo (2013), Ghayad (2013), Eriksson and Rooth (2014), Farber, Silverman, and von Wachter (2016, 2017), Nunley, Pugh, Romero, and Seals (2016), Farber, Herbst, Silverman, and von Wachter (2018)) have provided convincingly-identified nonparametric estimates of the causal effect of employment gaps in fictitious resumes on the callback rates of prospective employers. The evidence from this literature has been mixed, with some studies finding no evidence of short-term state dependence (Eriksson and Rooth (2014), Nunley et al. (2016), Farber, Silverman, and von Wachter (2017)) and others finding negative effects (Kroft, Lange, and Notowidigdo (2013), Ghayad (2013)). Farber et al. (2018) examined potential explanations for these differences, but concluded that the contrast in results remains a puzzle.

In this section, I use the DPO model to take a different look at this topic. The analysis uses observational data, and so avoids a key criticism of the experimental literature that callbacks may have a limited relationship to actual employment outcomes (Jarosch and Pilossoph (2019)). This benefit to external validity comes at the cost of imposing assumptions that are less credible than random assignment in a controlled experiment. On the other hand, these assumptions are nonparametric, so they may be an attractive alternative to researchers who are concerned about the impact of specific functional forms. The cost of remaining nonparametric is the loss of point identification, but the results ahead show that one can nevertheless still obtain informative estimates.

5.2. *Data*

The data is an extract of the 2008 Survey of Income and Program Participation (SIPP), which is a nationally-representative longitudinal survey covering the period of September 2008 to December 2013. Individuals in the SIPP are surveyed in four-month waves and questioned retrospectively on their employment status at different points over the previous four months. In order to mitigate seam bias, I follow a conservative strategy of using only observations corresponding to the last month of the retrospective four-month interview period (Grogger (2004), Ham and Shore-Sheppard (2005)). Also, I limit my sample to the period between January 2011 and April 2013, so as to avoid the more turbulent times surrounding the Great Recession. This leaves seven ($T = 6$) periods that are four months long each.[19]

I used the following sample selection rules when constructing the extract. I restricted attention to the subpopulation of working age men who were between 18 and 55 years of age at the time of the initial survey. I kept only men who reported either being employed or actively searching for work during all periods, so that $Y_{it} = 1$ denotes employment and $Y_{it} = 0$ denotes unemployment.[20] I also limit the focus to men who had a high school education or the equivalent, but no college degree, and who were not enrolled in school at any point during the sample. I dropped men who reported having a work-preventing disability or serving in the military at any point in the sample. Also, I removed observations that were heavily imputed (referred to as "type z" in the SIPP) or had other indications of irregularity.

After balancing the panel, the cross-section consists of 3435 men. Table I reports some summary statistics on the employment dynamics of these men. The overall unemployment rate ranges from roughly 5% to 8% over the course of the sample and employment is highly persistent with roughly 97% of men employed in one period remaining employed in the next. Unemployment is less persistent, with the probability of an unemployed man remaining unemployed ranging from approximately 50% early in the sample to around 60% in the later periods. Fewer than 1% of men remained unemployed in every period, but roughly 18% experience at least one spell of unemployment. A naive estimate of the average treatment effect would be between 0.455 and 0.595 depending on the period considered. This estimate probably overstates the role of positive state dependence if there is a permanent source of latent heterogeneity that positively affects the propensity to be employed.

5.3. *A Model of Job Search With Endogenous Effort*

In this section, I develop a DC model of job search to help motivate and interpret the assumptions used in the empirical analysis. The model features on-the-job search and endogenous search effort as in Christensen, Lentz, Mortensen, Neumann, and Werwatz

[19] I did not include the final two survey waves because they have unusually high attrition rates.

[20] Following Chetty (2008), I classify a worker as employed if they report having a job the entire interview month and not being on layoff.

TABLE I
DESCRIPTIVE STATISTICS ON UNEMPLOYMENT DYNAMICS IN THE SIPP[a]

	Time Period t						
	0	1	2	3	4	5	6
$\mathbb{P}[Y_{it}=1]$	0.921	0.936	0.945	0.931	0.945	0.949	0.942
	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\mathbb{P}[Y_{it}\neq Y_{i(t-1)}]$	–	0.067	0.056	0.055	0.050	0.043	0.048
		(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
$\mathbb{P}[Y_{it}=0 \mid Y_{i(t-1)}=0]$	–	0.483	0.493	0.632	0.534	0.574	0.594
		(0.030)	(0.034)	(0.035)	(0.032)	(0.036)	(0.037)
$\mathbb{P}[Y_{it}=1 \mid Y_{i(t-1)}=1]$	–	0.972	0.975	0.964	0.981	0.979	0.971
		(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Naive ATE	–	0.455	0.468	0.595	0.515	0.554	0.565
		(0.030)	(0.034)	(0.035)	(0.032)	(0.036)	(0.037)

	Percentage of Agents With ...							
	0	1	2	3	4	5	6	7
Periods of unemployment	82.30	7.83	3.29	2.36	1.75	1.11	0.47	0.90
Unemployment spells	82.30	13.36	3.61	0.73	0.00	–	–	–
Transitions	83.20	6.78	6.43	2.21	1.25	0.12	0.00	–

[a]Standard errors are given in parentheses. A transition is defined as the event $[Y_{it}\neq Y_{i(t-1)}]$. The "naive ATE" is defined as $\mathbb{P}[Y_{it}=1|Y_{i(t-1)}=1]-\mathbb{P}[Y_{it}=1|Y_{i(t-1)}=0]$. The sample size is 3435.

(2005) and Faberman, Mueller, Şahin, and Topa (2017). The model is nonparametric with respect to the functional form of search effort costs, unobserved heterogeneity, and the distribution of wage offers. This generality is possible because the only way the job search model will be used is to motivate nonparametric assumptions for the DPO model.

Worker i begins period t having either been employed or unemployed in the previous period ($Y_{i(t-1)}=1$ or 0, respectively), and having exerted $E_{i(t-1)}$ units of search effort in the previous period. The worker receives a wage offer, $\omega(Y_{i(t-1)}, E_{i(t-1)}, A_i, V_{it})$, which depends on their work and effort choices in the previous period, a permanent (time-invariant) source of heterogeneity, A_i, and a time-varying wage shock, V_{it}.[21] After observing the offer, the worker decides to either accept it and work in period t ($Y_{it}=1$) or to remain unemployed ($Y_{it}=0$).[22] In either case, they also decide on a level of search effort, E_{it}, to exert in period t as an investment to getting a better offer in period $t+1$. The worker solves this problem with an infinite horizon ($\bar{T}=+\infty$).

The worker's flow utility from work decision y' and effort choice e' is given by

$$\mu(y', e', Y_{i(t-1)}, E_{i(t-1)}, A_i, V_{it}) = y'\omega(Y_{i(t-1)}, E_{i(t-1)}, A_i, V_{it}) - \kappa(y', e', A_i),$$

where $\kappa(y', e', A_i)$ is the cost of exerting e' units of search effort when making employment choice y', and A_i allows for heterogeneity in these costs across workers. Allowing both κ and ω to depend on employment status means that both the costs and efficacy of searching can vary for employed and unemployed workers, as in Faberman et al. (2017). These are the two potential sources of state dependence in the model.

The distribution of the wage shock, V_{it}, is assumed to be first-order Markov, conditional on permanent heterogeneity.

[21]Not receiving an offer (or being laid off) corresponds to receiving an offer of $-\infty$.

[22]Past offers cannot be recalled.

ASSUMPTION MC: $\{V_{it}\}_{t=1}^{T}$ *is first-order Markov, conditional on* A_i.

The worker is assumed to know the conditional distribution of V_{it}, so that, under Assumption MC, the Bellman equation is given by

$$\begin{aligned}
&\nu(Y_{i(t-1)}, E_{i(t-1)}, V_{it}, A_i) \\
&\quad = \max_{(y',e')\in\{0,1\}\times\mathcal{E}} \{y'\omega(Y_{i(t-1)}, E_{i(t-1)}, A_i, V_{it}) - \kappa(y', e', A_i) \\
&\qquad + \delta\mathbb{E}[\nu(y', e', V_{i(t+1)}, A_i) \mid V_{it}, A_i]\}. \qquad (24)
\end{aligned}$$

Thus, the state variables at time t are $S_{it} \equiv (Y_{i(t-1)}, E_{i(t-1)}, V_{it}, A_i)$.

This dynamic job search model generates a DPO model through (7). Assumptions about the job search model imply assumptions on the generated DPO model. Consider the following assumptions on the wage shocks.

ASSUMPTION W: (a) *The distribution of* $(V_{i(t-m')}, \dots, V_{it})|A_i = a$ *does not depend on* t *for any* a, *where* m' *is some positive integer.*

(b) V_{it} *and* $V_{i(t+1)}$ *are independent, conditional on* A_i, *for all* t.

Assumption W(a) is that the wage offer distribution is stationary, which is a common assumption in empirical and theoretical analyses of labor search models, as well as empirical implementations of dynamic discrete choice models more generally.[23] The assumption still allows the distribution of wage shocks to vary for workers with different time-invariant characteristics. Assumption W(b) is that the wage shocks are serially independent, conditional on these characteristics. This is also a common assumption in empirical implementations of dynamic discrete choice models.[24]

The next proposition shows that Assumptions W(a) and W(b) imply that the generated DPO model satisfies Assumptions ST and DSC, respectively. If both assumptions hold, and the time-invariant characteristics include the initial employment outcome, Y_{i0}, then the generated DPO model also satisfies Assumption TIV.

PROPOSITION 8: *Suppose that every* $P \in \mathcal{P}^\dagger$ *is consistent with* $U_{it}(y)$ *being generated through* (7) *by the DC model described in this section. Suppose that Assumption* MC *is also satisfied.*

(i) *If Assumption* W(a) *is satisfied, then Assumption* ST *is satisfied with* $m = m' - 1$.

(ii) *If Assumption* W(b) *is satisfied, then Assumption* DSC *is satisfied.*

(iii) *If Assumptions* W(a) *and* W(b) *are satisfied, then Assumption* ST *is satisfied with* $m = m'$.

(iv) *Suppose that* $A_i \equiv (\bar{A}_i, Y_{i0})$ *contains the worker's employment choice in period* 0. *Then Assumption* TIV *is satisfied if Assumptions* W(a) *and* W(b) *are satisfied.*

Requiring Y_{i0} to be included as part of A_i in Proposition 8(iv) is difficult to motivate, since in the current application period 0 simply reflects the first period of observing data,

[23]As Keane, Todd, and Wolpin (2011, p. 371) wrote when discussing stationarity, "Most DCDP [discrete choice dynamic programming] models in the literature which solve the full dynamic programming problem implicitly make such an assumption as well, though it is not dictated by the method."

[24]However, see Norets (2009) and Connault (2016) for approaches to incorporating serially correlated unobservables.

and not some initial period for the workers. The empirical results in the next section strongly reject the hypothesis that the model is correctly specified under Assumption TIV.

The next proposition provides sufficient conditions for Assumption MTS.

PROPOSITION 9: *Suppose that every $P \in \mathcal{P}^\dagger$ is consistent with $U_{it}(y)$ being generated through* (7) *by the DC model described in this section. Suppose that Assumptions* MC *and* W(b) *are satisfied, and that the following additional conditions hold*: (i) *A_i takes values in a partially ordered set, and it includes the initial conditions $(Y_{i\underline{T}}, E_{i\underline{T}})$ at the beginning of the choice problem*; (ii) *A_i and V_{it} are independent*; (iii) *there is no search effort decision, so that $\kappa(y', e', a) = 0$ and $\omega(y, e, a, v)$ does not depend on e*; (*iv*) *V_{it} is scalar and $\omega(y, a, v) \equiv \bar{\omega}(y, a) + v$ for some function $\bar{\omega}$ that is increasing in both y and a, and supermodular in (y, a). Then Assumption* MTS *is satisfied with any value of q.*

The intuitive interpretation of Assumption MTS is that agents who are employed tend to be so because they have a permanently higher latent propensity to be employed. The conditions in Proposition 9 achieve this by imposing enough structure to make A_i this propensity. In the current model, this requires taking a stand on the relative importance of unobserved heterogeneity for wage offers and the cost of search effort. A simple way to do this is to simply shut down the search effort choice, as in (iii). This is not a terribly attractive assumption, since it removes the endogenous source of state dependence in the model, although an exogenous source of state dependence still exists via the wage offer function. For this reason, Assumption MTS will only be used in a secondary capacity in the next section.

Faberman et al. (2017) found survey evidence that the distribution of wage offers for employed workers dominates that of non-employed workers. The next assumptions capture this idea in differing strengths. To state them, let $e^\star(S_{i(t-1)} \parallel y)$ denote the optimal choice of effort in period $t-1$, given employment choice y, and let $W_{it}(y) \equiv \omega(y, e^\star(S_{i(t-1)} \parallel y), A_i, V_{it})$ denote the wage that would have been received in period t under these choices.

ASSUMPTION W: (d) *The distribution of $W_{it}(1)$ first-order stochastically dominates that of $W_{it}(0)$, conditional on A_i.*

(e) *$W_{it}(1)$ is greater than $W_{it}(0)$ with probability* 1.

Assumption W(d) says that conditional on permanent heterogeneity, employed workers tend to receive higher wage offers than unemployed workers. Assumption W(e) strengthens this to assume that employed workers always receive better offers.

Assumptions W(d) and W(e) imply that the generated DPO model satisfies Assumption MATR and the stronger Assumption MTR, respectively.

PROPOSITION 10: *Suppose that every $P \in \mathcal{P}^\dagger$ is consistent with $U_{it}(y)$ being generated through* (7) *by the DC model described in this section. Suppose that Assumptions* MC *and* W(b) *are satisfied. If Assumption* W(d) *is satisfied, then Assumption* MATR *is satisfied. If Assumption* W(e) *is satisfied, then Assumption* MTR *is satisfied.*

There is ample reason to be skeptical of both Assumptions W(d) and W(e). For example, unemployed workers might receive higher offers than employed workers because they exert more effort searching for offers. Perhaps for this reason, the results in the next section indicate that the model is misspecified under Assumption MTR. While there is no such evidence against Assumption MATR, this assumption turns out to have little identifying content, so will not be used in the main results.

5.4. *Empirical Results*

Columns (1)–(9) of Table II report estimated identified sets for a variety of parameters under several combinations of assumptions. The reported target parameters are time-averages of those discussed in Section 3.2, for example,

$$\mathrm{SD}^{+}_{\mathrm{avg}}(P) \equiv \frac{1}{T}\sum_{t=1}^{T}\mathrm{SD}^{+}_{t}(P),$$

and similarly for the other parameters. For columns (1)–(7), the linear programs in Proposition 2 are feasible, so the estimated bounds are equal to the sharp identified set under the sample distribution of the data.[25] For columns (8) and (9), the programs were infeasible, so the estimated bounds are constructed using the estimation procedure described in Appendix S7. The reported 95% confidence intervals were constructed using the procedure of Chernozhukov, Newey, and Santos (2015), which is also described in Appendix S7. The Monte Carlo results reported there suggest that these confidence intervals control size but are excessively wide when the model is correctly specified.[26]

Column (1) contains the empirical-evidence-only bounds. These are wide for all target parameters, and in some cases completely uninformative. As predicted by Proposition 1, the bounds for $\mathrm{SD}^{+}_{\mathrm{avg}}$ include 0 and the identified set for $\mathrm{SD}_{\mathrm{avg}}$ is $[0, 1]$. The upper bound on $\mathrm{SD}^{+}_{\mathrm{avg}}$ is large (0.947), which reflects the strong positive serial correlation present in the observed data. To draw informative conclusions, more assumptions must be imposed.

Columns (2)–(6) impose Assumption ST for increasingly long sequence lengths, m. As expected, the bounds narrow as m increases, with the most informative bounds being obtained for $m = 4$, which is the strongest form of Assumption ST possible given the horizon of $T = 6$. For $m \geq 1$, the bounds on all target parameters exclude 0. In particular, the lower bound on overall state dependence, $\mathrm{SD}_{\mathrm{avg}}$, ranges between 0.018 and 0.061. This provides simple, nonparametric evidence against the hypothesis that persistence in employment outcomes is caused solely by persistent unobserved heterogeneity. However, the small magnitude of these lower bounds is still consistent with state dependence being of limited importance for the overall population.

On the other hand, the results indicate that the role of state dependence is quite important for unemployed men. With $m = 2$ in column (4), the lower bound of 0.193 for $\mathrm{SD}^{+}_{\mathrm{avg}}(\cdot|0)$ means that at least 19.3% of unemployed men remain unemployed in the subsequent period due to state dependence, that is, due to the fact that they are unemployed. Using the interpretation discussed in Section 3.2, the lower bound of 0.335 on $\mathrm{SD}^{+}_{\mathrm{avg}}(\cdot|00)$ means that this causal effect accounts for at least 33.5% of the four-month persistence in unemployment. With $m = 4$ in column (6), these numbers rise to 23.8% and 41.4%, respectively.

In contrast, Assumption ST does not by itself provide informative bounds for *employed* men. The bounds for employed men become informative when Assumption MTS is added

[25]The linear programs were solved using AMPL (Fourer, Gay, and Kernighan (2002)) and CPLEX (IBM (2010)).

[26]Note that a tuning parameter, called τ_n in Appendix S7, is required both for estimating identified sets when they are empty in sample, and for constructing confidence regions. I set this parameter to $\tau_n = 0.25$ throughout. This value was chosen because, in the Monte Carlo simulation in Appendix S7.6, it yields confidence regions with the correct coverage probability at the boundaries of the identified set.

TABLE II

ESTIMATES OF STATE DEPENDENCE IN UNEMPLOYMENT[a]

	DPO									PDBR		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Assumptions												
ST(m)		0	1	2	3	4	4	2		4		
MTS(q)							2			6	6	
MTR								✓				
TIV									✓	✓	✓	✓
Misspecification												
$\Theta^\star = \emptyset$ in sample	No	No	No	No	No	No	No	Yes	Yes			
p-value for $H_0 : \Theta^\star \neq \emptyset$								0.012	0.000			
Bounds and 95% Confidence Intervals												
SD^+_{avg}	0.000	0.000	0.005	0.011	0.013	0.013	0.013	0.036	0.027	0.067	0.067	0.067
	0.000	0.000	0.016	0.025	0.030	0.034	0.034	0.037	0.028	0.092	0.092	0.092
	0.947	0.933	0.933	0.933	0.933	0.933	0.379	0.932	0.904			
	0.950	0.939	0.939	0.939	0.939	0.939	0.468	0.935	0.904	0.115	0.114	0.114
$\text{SD}^+_{\text{avg}}(\cdot\|0)$	0.000	0.000	0.030	0.065	0.072	0.088	0.088	0.217	0.164	0.149	0.148	0.148
	0.000	0.000	0.096	0.193	0.214	0.238	0.238	0.218	0.164	0.192	0.191	0.192
	0.581	0.581	0.581	0.576	0.574	0.569	0.554	0.514	0.581			
	0.641	0.641	0.641	0.640	0.643	0.649	0.651	0.515	0.582	0.234	0.228	0.228
$\text{SD}^+_{\text{avg}}(\cdot\|00)$	0.000	0.000	0.054	0.114	0.128	0.153	0.153	0.379	0.286	0.264	0.266	0.266
	0.000	0.000	0.166	0.335	0.372	0.414	0.414	0.380	0.286	0.329	0.327	0.330
	1.00	1.00	1.00	0.992	0.990	0.980	0.956	0.891	1.00			
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.892	1.00	0.388	0.385	0.385
$\text{SD}^+_{\text{avg}}(\cdot\|1)$	0.000	0.000	0.002	0.003	0.006	0.006	0.006	0.021	0.016	0.062	0.062	0.062
	0.000	0.000	0.006	0.010	0.014	0.016	0.016	0.021	0.016	0.085	0.085	0.086
	0.970	0.970	0.968	0.966	0.964	0.963	0.371	0.961	0.926			
	0.975	0.975	0.974	0.973	0.972	0.971	0.463	0.964	0.926	0.108	0.107	0.107
$\text{SD}^+_{\text{avg}}(\cdot\|11)$	0.000	0.000	0.002	0.004	0.006	0.006	0.006	0.021	0.016	0.064	0.064	0.064
	0.000	0.000	0.006	0.010	0.014	0.017	0.017	0.022	0.017	0.088	0.088	0.088
	1.00	1.00	0.998	0.996	0.994	0.993	0.382	0.991	0.955			
	1.00	1.00	1.00	1.00	1.00	1.00	0.477	0.994	0.955	0.111	0.110	0.110
SD_{avg}	0.000	0.013	0.023	0.035	0.035	0.034	0.034	0.036	0.029	0.067	0.067	0.067
	0.000	0.018	0.036	0.054	0.058	0.061	0.061	0.037	0.030	0.092	0.092	0.092
	1.00	0.976	0.976	0.976	0.976	0.975	0.422	0.932	0.940			
	1.00	0.983	0.983	0.984	0.985	0.986	0.511	0.935	0.940	0.115	0.114	0.114

[a]Estimated bounds and 95% confidence intervals are reported in large and small font, respectively. Confidence intervals and misspecification p-values for the DPO model are obtained using the CNS method discussed in Appendices S7.3–S7.5 with 250 bootstrap draws and $\tau_n = 0.25$. When the sample identified set is empty, estimated bounds are constructed using the method discussed in Appendix S7.2. When the identified set is nonempty in sample, the p-value of the CNS misspecification test is 1, and therefore not reported—see Appendix S7.4. Point estimates and confidence intervals for the parametric DBR model are obtained by maximum likelihood and nonparametric bootstrap. The model is the dynamic random effects probit described in Section 3.2 and the covariates are taken to be a constant (column (10)), a full set of time dummies (column (11)), and a full set of time dummies with age (column (12)). The assumptions listed for the parametric DBR model are those that would be satisfied for the implied DPO model.

to Assumption ST in column (7).[27] The estimates imply that no more than 37.1% of employed workers remain employed due to state dependence, and that state dependence accounts for no more than 38.2% of the four-month persistence in employment. Intuitively, these upper bounds arise because Assumption MTS reflects an assumption that there is some persistent heterogeneity, which limits the role of state dependence in explaining persistence in observed outcomes.

Together, columns (4)–(7) yield two conclusions. The first is that state dependence plays an important role in unemployment persistence, accounting for at least 30–40% of the four-month persistence in unemployment among high school educated men. These estimates only depend on stationarity, which was easy to motivate in the job search model. The second conclusion is that state dependence explains less of the persistence in employment than it does unemployment. This conclusion depends on both Assumption ST and Assumption MTS. The latter can be interpreted as limiting persistent heterogeneity, which was harder to motivate in the job search model.

Column (8) adds Assumption MTR to Assumption ST with $m = 2$. Doing so causes the sample sharp identified set to become empty. This suggests misspecification, and a formal misspecification test provides evidence against the null of correct specification with a p-value of 0.012.[28] Such a finding is perhaps unsurprising given the strong assumptions required to motivate Assumption MTR in the job search model. Even if we were to put aside these concerns, adding Assumption MTR does little to narrow the bounds relative to column (4), so there would be little benefit to considering it anyway.[29]

Column (9) reports estimates that use the restrictions implied by Assumption TIV in Proposition 5. The sample identified set is empty, and a formal misspecification test rejects the null of correct specification with a p-value near 0. Like Assumption MTR, Assumption TIV was also difficult to justify in the job search model, so this finding may not be surprising. The estimated identified sets in column (9) are similar to those obtained under Assumption ST with $m = 2$ in column (4), and suggest in particular that state dependence is important for explaining persistence in unemployment. However, given the resounding rejection of the specification, it seems prudent not to put much weight on these estimates.

As a point of comparison, columns (10)–(12) report maximum likelihood estimates and bootstrapped confidence intervals from a parametric DBR model. The model is the "textbook version" of the dynamic random effects probit model described in Section 2.3. The assumptions indicated in Table II are those that would be implied (among others not listed) for the generated DPO model. Column (10) is a stationary specification with only a constant in X_{it}, while column (11) includes a full set of time dummies, and column (12) also adds the worker's age as a covariate. All three specifications yield nearly identical estimates across all target parameters considered. For unemployed workers, the estimates are near or outside the lower bound of the stationary DPO model in columns (4)–(6). For

[27]The reported bounds use conditioning length $q = 2$ in Assumption MTS. Extending this to $q = 3$ had little impact on the estimated bounds.

[28]As suggested by a referee, the confidence regions may be too narrow under misspecification. Moreover, the estimated bounds will not be consistent if the model is, as suggested by these tests, actually misspecified.

[29]Note that since the bounds in column (8) are estimated using the procedure in Appendix S7.2, they need not be narrower than those in column (4), even though column (4) maintains fewer assumptions. For example, the lower bound on SD_{avg} is larger in column (4) than in column (8). If the population identified sets were nonempty for both specifications, then asymptotically the bounds in column (8) would become subsets of those in column (4).

other parameters, the estimates are within the tightest DPO bounds in (7), but still closer to the lower bound. This suggests that the parametric DBR model might underestimate state dependence.

Table II does not include estimates for the DPO model under Assumptions DSC, MATR, or MIV. Assumptions DSC and MATR turned out to have very little impact on the bounds in this application. Estimated identified sets for these assumptions are reported in Appendix S8. Assumption MIV requires a compelling instrumental variable. This is not easy to find for the current application, but may be easier to find in other settings.[30]

5.5. *Sensitivity to Stationarity*

The results in the previous section show that Assumption ST is sufficient to reach the conclusion that state dependence is important for unemployed workers. Stationarity is a widely used assumption, and was easy to motivate in the job search model in Section 5.3. However, it may fail to hold if the causal mechanisms underlying the labor market changed over the sample of April 2011 to January 2013, or if there were seasonal changes within this horizon. It could also fail to the extent that the infinite horizon assumption ($\bar{T} = \infty$) used in Section 5.3 fails as a model of dynamic choice behavior. In this section, I examine the sensitivity to violations of Assumption ST.

To do so, I replace Assumption ST with the following, weaker condition.

ASSUMPTION ST(σ): *Let $\sigma \geq 0$ be a known scalar. Then, for any $P \in \mathcal{P}^\dagger$, every $u \in \{0,1\}^{2(m+1)}$, and every $t, t' \geq m+1$,*

$$(1-\sigma)\Sigma^u_{t',m}(P) \leq \Sigma^u_{t,m}(P) \leq (1+\sigma)\Sigma^u_{t',m}(P),$$

where $\Sigma^u_{t,m}$ is defined as in Assumption ST.

The parameter σ can be interpreted as the amount of "slippage" in $\Sigma_{t,m}(u; P)$ that is allowed in any two periods. For $\sigma = 0$, Assumption ST(σ) is the same as Assumption ST. For $\sigma > 0$, Assumption ST(σ) is strictly weaker than Assumption ST, since it allows latent probabilities to change by up to $100 \times \sigma\%$ in any two periods.

Figure 2 plots estimated identified sets and 95% confidence regions for $\mathrm{SD}^+_{\mathrm{avg}}(\cdot|0)$ and $\mathrm{SD}^+_{\mathrm{avg}}(\cdot|00)$ across different values of σ. The bounds widen as σ increases, reflecting the fact that the restriction of Assumption ST(σ) becomes monotonically weaker. Setting $\sigma = 0.15$ allows the probability of each potential employment sequence between two periods to change by up to 15%. This in turn allows SD^+_t—that is, the probability that a worker would be employed in period t if and only if they were employed the previous period—to change by 30 probability points over the time period considered, which would reflect an enormous change in the structure of the labor market. Yet, even for values of σ larger than this, the lower bounds on $\mathrm{SD}^+_{\mathrm{avg}}(\cdot|0)$ and $\mathrm{SD}^+_{\mathrm{avg}}(\cdot|00)$ are substantial, indicating that state dependence has an important impact on the unemployed. Thus, the main conclusions regarding the unemployed appear quite robust to violations of Assumption ST.

[30]One possibility, suggested by a referee, is to use variation in unemployment insurance benefits as in Farber, Rothstein, and Valletta (2015) or Farber and Valletta (2015).

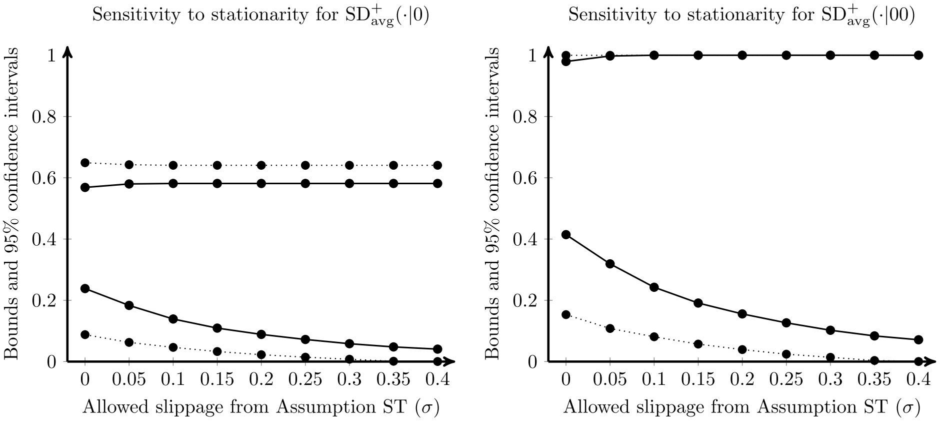

FIGURE 2.—Sensitivity of Table II, column (6) to Assumption ST.

As a further check, I estimate the same bounds for a sample of workers aged 40 or younger in the initial period. For these workers, the infinite horizon assumption is more reasonable, since they are farther from retirement. The estimated identified sets and confidence regions are plotted in Figure 3.[31] The lower bound estimates are substantially larger than those in Figure 2, and remain quite large even at values of σ that would represent massive changes in the underlying labor market. For younger workers, the conclusion that state dependence is important for the unemployed appears to be even stronger, even while this sample is also more likely to satisfy Assumption ST.

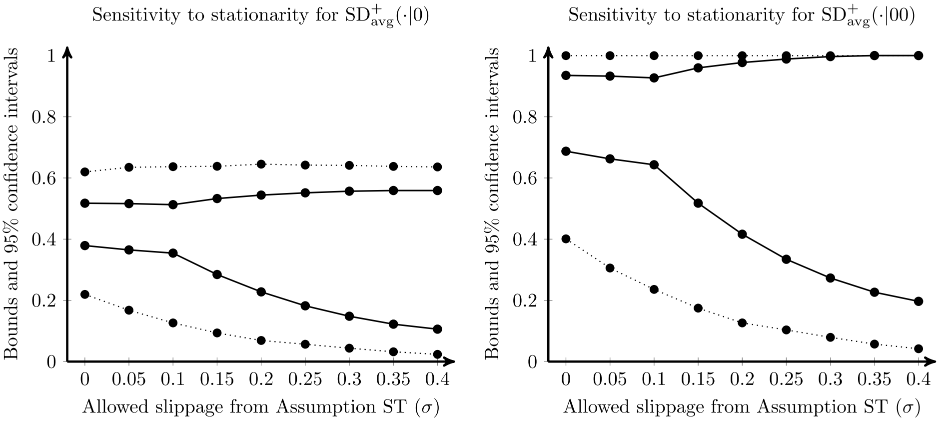

FIGURE 3.—Estimates for younger workers under Assumption ST(σ) with $m = 4$.

[31]Note that the analog identified sets are empty in the sample here, so the bounds are constructed using the estimation procedure discussed in Appendix S7.2. This explains the slight decline in the upper bound from $\sigma = 0.05$ to $\sigma = 0.1$. The p-values for testing the null of correct specification are larger than 0.55 for all values of σ.

6. CONCLUSION

I developed the dynamic potential outcomes (DPO) model as a tool for empirically measuring state dependence in dynamic discrete outcomes. The DPO model has the important advantage of being fully nonparametric. Its primary disadvantage is that causal parameters tend to only be partially identified. Nevertheless, in applying the method to study state dependence in unemployment using data from the SIPP, I demonstrated that the estimated identified sets can still be tight enough to be useful. In particular, the estimates provide nonparametric evidence using observational data that state dependence plays an important role in unemployment persistence, accounting for at least 30–40% of the four-month persistence. The estimates rest on a stationarity assumption, which says that the underlying economic environment remains stable, or at least does not change drastically.

REFERENCES

ABBRING, J. H., AND J. J. HECKMAN (2007): "Chapter 72 Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics*, Vol. 6, Part 2, ed. by J. J. Heckman and E. E. Leamer. Elsevier, 5145–5303. [1477]

ALESSIE, R., S. HOCHGUERTEL, AND A. V. SOEST (2004): "Ownership of Stocks and Mutual Funds: A Panel Data Analysis," *The Review of Economics and Statistics*, 86, 783–796. [1477]

ANGRIST, J. D., AND G. M. KUERSTEINER (2011): "Causal Effects of Monetary Shocks: Semiparametric Conditional Independence Tests With a Multinomial Propensity Score," *Review of Economics and Statistics*, 93, 725–747. [1477]

ARULAMPALAM, W., A. BOOTH, AND M. TAYLOR (2000): "Unemployment Persistence," *Oxford Economic Papers*, 52, 24–50. [1492]

BALKE, A., AND J. PEARL (1994): "Counterfactual Probabilities: Computational Methods, Bounds, and Applications," in *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI-94)*, ed. by R. Lopez de Mantras and D. Poole, 46–54. [1486]

——— (1997): "Bounds on Treatment Effects From Studies With Imperfect Compliance," *Journal of the American Statistical Association*, 92, 1171–1176. [1486]

BERNARD, A. B., AND J. B. JENSEN (2004): "Why Some Firms Export," *The Review of Economics and Statistics*, 86, 561–569. [1477]

CARD, D., AND R. HYSLOP (2005): "Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers," *Econometrica*, 73, 1723–1770. [1477,1483]

CHAY, K. Y., H. HOYNES, AND D. HYSLOP (2004): "True State Dependence in Monthly Welfare Participation: A Nonexperimental Analysis," Working Paper. [1477,1483]

CHEN, X., E. TAMER, AND A. TORGOVITSKY (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," Cowles Foundation Discussion Paper 1836. [1478]

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, J. HAHN, AND W. NEWEY (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81, 535–580. [1476,1490]

CHERNOZHUKOV, V., W. NEWEY, AND A. SANTOS (2015): "Constrained Conditional Moment Restriction Models," CEMMAP Working Paper CWP 59/15. [1497]

CHETTY, R. (2008): "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 116, 173–234. [1493]

CHIBURIS, R. C. (2010): "Semiparametric Bounds on Treatment Effects," *Journal of Econometrics*, 159, 267–275. [1486]

CHRISTENSEN, B. J., R. LENTZ, D. T. MORTENSEN, G. R. NEUMANN, AND A. WERWATZ (2005): "On-the-Job Search and the Wage Distribution," *Journal of Labor Economics*, 23, 31–58. [1493,1494]

CONNAULT, B. (2016): "Hidden Rust Models," Working Paper. [1495]

CONTOYANNIS, P., A. M. JONES, AND N. RICE (2004): "The Dynamics of Health in the British Household Panel Survey," *J. Appl. Econ.*, 19, 473–503. [1477]

CORCORAN, M., AND M. S. HILL (1985): "Reoccurrence of Unemployment Among Adult Men," *The Journal of Human Resources*, 20, 165–183. [1492]

DEMUYNCK, T. (2015): "Bounding Average Treatment Effects: A Linear Programming Approach," *Economics Letters*, 137, 75–77. [1486]

DEZA, M. (2015): "Is There a Stepping Stone Effect in Drug Use? Separating State Dependence From Unobserved Heterogeneity Within and Between Illicit Drugs," *Journal of Econometrics*, 184, 193–207. [1477]

DRAKOS, K., AND P. T. KONSTANTINOU (2013): "Investment Decisions in Manufacturing: Assessing the Effects of Real Oil Prices and Their Uncertainty," *J. Appl. Econ.*, 28, 151–165. [1477]

DUBÉ, J.-P., G. J. HITSCH, AND P. E. ROSSI (2010): "State Dependence and Alternative Explanations for Consumer Inertia," *The RAND Journal of Economics*, 41, 417–445. [1477]

ELLWOOD, D. T. (1982): "Teenage Unemployment: Permanent Scars or Temporary Blemishes?" in *The Youth Labor Market Problem: Its Nature, Causes, and Consequences*. University of Chicago Press, 349–390. [1492]

ERIKSSON, S., AND D.-O. ROOTH (2014): "Do Employers Use Unemployment as a Sorting Criterion When Hiring? Evidence From a Field Experiment," *American Economic Review*, 104, 1014–1039. [1492]

FABERMAN, R. J., A. MUELLER, A. ŞAHIN, AND G. TOPA (2017): "Job Search Behavior Among the Employed and Non-Employed," Tech. rep. [1494,1496]

FARBER, H., C. HERBST, D. SILVERMAN, AND T. VON WACHTER (2018): "Whom Do Employers Want? The Role of Recent Employment and Unemployment Status and Age," Tech. rep. [1492]

FARBER, H. S., AND R. G. VALLETTA (2015): "Do Extended Unemployment Benefits Lengthen Unemployment Spells?: Evidence From Recent Cycles in the U.S. Labor Market," *Journal of Human Resources*, 50, 873–909. [1500]

FARBER, H. S., J. ROTHSTEIN, AND R. G. VALLETTA (2015): "The Effect of Extended Unemployment Insurance Benefits: Evidence From the 2012–2013 Phase-out," *American Economic Review*, 105, 171–176. [1500]

FARBER, H. S., D. SILVERMAN, AND T. VON WACHTER (2016): "Determinants of Callbacks to Job Applications: An Audit Study," *American Economic Review*, 106, 314–318. [1492]

FARBER, H. S., D. SILVERMAN, AND T. M. VON WACHTER (2017): "Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study," *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 3, 168. [1492]

FLINN, C., AND J. HECKMAN (1982): "New Methods for Analyzing Structural Models of Labor Force Dynamics," *Journal of Econometrics*, 18, 115–168. [1481]

FOURER, R., D. M. GAY, AND B. W. KERNIGHAN (2002): *AMPL: A Modeling Language for Mathematical Programming*. Cengage Learning. [1497]

FREYBERGER, J., AND J. L. HOROWITZ (2015): "Identification and Shape Restrictions in Nonparametric Instrumental Variables Estimation," *Journal of Econometrics*, 189, 41–53. [1486]

GHAYAD, R. (2013): "The Jobless Trap," Working Paper. [1492]

GROGGER, J. (2004): "Welfare Transitions in the 1990s: The Economy, Welfare Policy, and the EITC," *J. Pol. Anal. Manage.*, 23, 671–695. [1493]

HAM, J. C., AND L. SHORE-SHEPPARD (2005): "The Effect of Medicaid Expansions for Low-Income Children on Medicaid Participation and Private Insurance Coverage: Evidence From the SIPP," *Journal of Public Economics*, 89, 57–83. [1493]

HAM, J. C., D. IORIO, AND M. SOVINSKY (2013): "Caught in the Bulimic Trap?: Persistence and State Dependence of Bulimia Among Young Women," *Journal of Human Resources*, 48, 736–767. [1477]

HANDEL, B. R. (2013): "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103, 2643–2682. [1477]

HANSEN, L. P., J. HEATON, AND E. G. J. LUTTMER (1995): "Econometric Evaluation of Asset Pricing Models," *The Review of Financial Studies*, 8, 237–274. [1486]

HECKMAN, J. J. (1978): "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence," *Annales de l'insée*, 227–269. [1475,1476]

——— (1981a): "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. by S. Rosen. University of Chicago Press. [1475,1476,1480]

——— (1981b): "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski and D. McFadden. Cambridge, MA: MIT Press. [1476,1488]

——— (1981c): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Data With Econometric Applications*, ed. by C. F. Manski and D. McFadden. Cambridge, MA: MIT Press. [1478,1480]

HECKMAN, J. J., AND G. J. BORJAS (1980): "Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers From a Continuous Time Model of Heterogeneity and State Dependence," *Economica*, 47, 247–283. [1478]

HECKMAN, J. J., AND S. NAVARRO (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136, 341–396. [1481]

HECKMAN, J. J., AND E. J. VYTLACIL (2007): "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in *Handbook of Econometrics*, Vol. 6, Part 2, ed. by J. J. Heckman and E. E. Leamer. Elsevier, 4779–4874. [1484]

HECKMAN, J. J., AND R. J. WILLIS (1977): "A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," *Journal of Political Economy*, 85, 27–58. [1475]
HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *The Review of Economic Studies*, 64, 487–535. [1485]
HONORÉ, B. E., AND E. TAMER (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629. [1478,1486]
HYSLOP, D. R. (1999): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67, 1255–1294. [1480]
IBM (2010): "IBM ILOG AMPL Version 12.2," International Business Machines Corporation. [1497]
IRACE, M. (2018): "Patient Loyalty in Hospital Choice: Evidence From New York," Working Paper. [1477]
JAROSCH, G., AND L. PILOSSOPH (2019): "Statistical Discrimination and Duration Dependence in the Job Finding Rate," *The Review of Economic Studies*, 86 (4), 1631–1665. [1492]
KEANE, M. P. (1997): "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," *Journal of Business & Economic Statistics*, 15, 310–327. [1477]
KEANE, M. P., P. E. TODD, AND K. I. WOLPIN (2011): "Chapter 4—the Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications," in *Handbook of Labor Economics*, Vol. 4, Part A. Elsevier, 331–461. [1481,1495]
KITAMURA, Y., AND J. STOYE (2013): "Nonparametric Analysis of Random Utility Models: Testing," CEMMAP Working Paper CWP36/13. [1486]
——— (2018): "Nonparametric Analysis of Random Utility Models," *Econometrica*, 86, 1883–1909. [1486]
KROFT, K., F. LANGE, AND M. J. NOTOWIDIGDO (2013): "Duration Dependence and Labor Market Conditions: Evidence From a Field Experiment," *The Quarterly Journal of Economics*, 128, 1123–1167. [1492]
LAFFÉRS, L. (2013): "Essays in Partial Identification," Ph.D. thesis, Norwegian School of Economics. [1486]
——— (2018): "Bounding Average Treatment Effects Using Linear Programming," *Empirical Economics*. [1486]
MAGNAC, T., AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816. [1481]
MANSKI, C. (1994): "The Selection Problem," in *Advances in Econometrics, Sixth World Congress*, Vol. 1, 143–170. [1486]
MANSKI, C. F. (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311–1334. [1492]
——— (2006): "Two Problems of Partial Identification With Panel Data," in *13th International Conference on Panel Data, July 7–9*. Cambridge. [1476]
——— (2007): "Partial Indentification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393–1410. [1486]
——— (2014): "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy," *Quantitative Economics*, 5, 145–174. [1486]
MANSKI, C. F., AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997–1010. [1486,1489,1491]
——— (2009): "More on Monotone Instrumental Variables," *Econometrics Journal*, 12, S200–S216. [1480, 1491]
MOLINARI, F. (2008): "Partial Identification of Probability Distributions With Misclassified Data," *Journal of Econometrics*, 144, 81–117. [1486]
MOURIFIÉ, I. (2015): "Sharp Bounds on Treatment Effects in a Binary Triangular System," *Journal of Econometrics*, 187, 74–81. [1486]
MÜHLEISEN, M., AND K. F. ZIMMERMANN (1994): "A Panel Analysis of Job Changes and Unemployment," *European Economic Review*, 38, 793–801. [1492]
NARENDRANATHAN, W., AND P. ELIAS (1993): "Influences of Past History on the Incidence of Youth Unemployment: Empirical Findings for the UK," *Oxford Bulletin of Economics and Statistics*, 55, 161–185. [1492]
NORETS, A. (2009): "Inference in Dynamic Discrete Choice Models With Serially Orrelated Unobserved State Variables," *Econometrica*, 77, 1665–1682. [1495]
NORETS, A., AND X. TANG (2014): "Semiparametric Inference in Dynamic Binary Choice Models," *The Review of Economic Studies*, 81, 1229–1262. [1481]
NUNLEY, J. M., A. PUGH, N. ROMERO, AND R. A. SEALS (2016): "The Effects of Unemployment and Underemployment on Employment Opportunities," *ILR Review*, 70, 642–669. [1492]
OBERHOLZER-GEE, F. (2008): "Nonemployment Stigma as Rational Herding: A Field Experiment," *Journal of Economic Behavior & Organization*, 65, 30–40. [1492]
OKUMURA, T., AND E. USUI (2014): "Concave-Monotone Treatment Response and Monotone Treatment Selection: With an Application to the Returns to Schooling," *Quantitative Economics*, 5, 175–194. [1486]

RUST, J. (1994): "Chapter 51 Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics*, Vol. 4, ed. by R. F. Engle and D. L. McFadden. Elsevier, 3081–3143. [1479,1481]

SHAIKH, A. M., AND E. J. VYTLACIL (2011): "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 79, 949–955. [1486]

SHAKED, M., AND J. G. SHANTHIKUMAR (2007): *Stochastic Orders*. Springer. [1488]

STOKEY, N. L., R. E. LUCAS, AND E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press. [1479]

TORGOVITSKY, A. (2019a): "Partial Identification by Extending Subdistributions," *Quantitative Economics*, 10, 105–144. [1478,1481]

——— (2019b): "Supplement to 'Nonparametric Inference on State Dependence in Unemployment'," *Econometrica Supplemental Material*, 87, https://doi.org/10.3982/ECTA14138. [1477]

TUMINO, A. (2015): "The Scarring Effect of Unemployment From the Early '90s to the Great Recession," Working Paper. [1492]

WOOLDRIDGE, J. M. (2005): "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models With Unobserved Heterogeneity," *J. Appl. Econ.*, 20, 39–54. [1480]

——— (2010): *Econometric Analysis of Cross Section and Panel Data*. MIT Press. [1480,1491]

Co-editor Liran Einav handled this manuscript.

Manuscript received 1 February, 2016; final version accepted 30 April, 2019; available online 6 May, 2019.