Nonparametric Identification of Dynamic Models with Unobserved State Variables*

Yingyao Hu
Johns Hopkins University

Matthew Shum
California Institute of Technology

First version: April 2008
This version: January 2009

Abstract

We consider the identification of a Markov process $\{W_t, X_t^*\}$ when only $\{W_t\}$, a subset of the variables, are observed. In structural dynamic models, W_t includes the choice variables and observed state variables of an optimizing agent, while X_t^* denotes the serially correlated unobserved state variables (or agent-specific unobserved heterogeneity). In the non-stationary case, we show that the Markov law of motion $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is identified from five periods of data $W_{t+1}, W_t, W_{t-1}, W_{t-2}, W_{t-3}$. In the stationary case, only four observations $W_{t+1}, W_t, W_{t-1}, W_{t-2}$ are required. Identification of $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is a crucial input in methodologies for estimating Markovian dynamic models based on the "conditional-choice-probability (CCP)" approach pioneered by Hotz and Miller.

1 Introduction

In this paper, we consider the identification of a Markov process $\{W_t, X_t^*\}$ when only $\{W_t\}$, a subset of the variables, is observed. In structural dynamic models, W_t typically consists of the choice variables and observed state variables of an optimizing agent. X_t^* denotes the serially correlated unobserved state variables (or agent-specific unobserved heterogeneity), which are observed by the agent, but not by the econometrician.

We demonstrate two main results. First, in the non-stationary case, where the Markov law of motion $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$, can vary across periods t, we show that, for any period

*The authors can be reached at yhu@jhu.edu and mshum@caltech.edu. We thank Xiaohong Chen, Jeremy Fox, Han Hong, Ariel Pakes, and Susanne Schennach for their suggestions. Seminar participants at BU, Harvard, LSE, MIT, NYU, Penn, Penn State, Toulouse, UCL, UCLA, USC, the Cowles 2008 Summer Conference at Yale, the 2008 ERID Conference at Duke, the 2008 Greater New York Econometrics Colloquium at Princeton, and the "Econometrics of Industrial Organization" workshop at Toulouse provided useful comments.

t, $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is identified from five periods of data $W_{t+1},\ldots,W_{t-3}$. Second, in the stationary case, where $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is the same across all t, only four observations $W_{t+1},\ldots,W_{t-2}$, for some t, are required for identification.

In most applications, W_t consists of two components $W_t = (Y_t, M_t)$, where Y_t denotes the agent's action in period t, and M_t denotes the period-t observed state variable(s). X_t^* are persistent unobserved state variables (USV for short), which are observed by agents and affect their choice of Y_t, but are unobserved by the econometrician. We begin by giving several motivating examples of well-known Markovian dynamic discrete-choice models which have been estimated in the existing literature.

[1] **Miller's (1984)** job matching model was one of the first empirical dynamic discrete choice models with unobserved state variables. Y_t is an indicator for the occupation chosen by a worker in period t, and the unobserved state variables X_t^* are the posterior means of workers' beliefs regarding their occupation-specific match values. The observed state variables M_t include job tenure and education level. ■

[2] In **Rust's (1987)** bus engine replacement model, Y_t is an indicator for whether Harold Zurcher (the bus depot manager) decides to replace the bus engine in week t. M_t is the accumulated mileage of the bus since the last engine replacement, in week t. Although Rust's original paper had no persistent unobserved state variable X_t^*, one could extend the model to allow for them. For example, X_t^* could be Harold Zurcher's health, or weather or road conditions during week t.■

[3] **Pakes (1986)** estimates an optimal stopping model of the year-by-year renewal decision on European patents. In his model, the decision variable Y_t is an indicator for whether a patent is renewed in year t, and the unobserved state variable X_t^* is the profitability from the patent in year t, which is not observed by the econometrician. The observed state variable M_t could be other time-varying factors, such as the stock price or total sales of the patent-holding firm, which affect the renewal decision. ■

The main result in this paper concerns the identification of the Markov law of motion $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$. Once this is known, it factors into conditional and marginal distributions of economic interest. For Markovian dynamic optimization models (such as the examples given above), $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ factors into

$$\begin{aligned} f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} &= f_{Y_t,M_t,X_t^*|Y_{t-1},M_{t-1},X_{t-1}^*} \\ &= \underbrace{f_{Y_t|M_t,X_t^*}}_{\text{CCP}} \cdot \underbrace{f_{M_t,X_t^*|Y_{t-1},M_{t-1},X_{t-1}^*}}_{\text{state law of motion}}. \end{aligned} \tag{1}$$

The first term denotes the conditional choice probability for the agent's optimal choice in period t. The second term is the Markovian law of motion for the state variables (M_t, X_t^*).

Once the CCP's and the law of motion for the state variables are recovered, it is straightforward to use them as inputs in a CCP-based approach for estimating dynamic discrete-choice models. This approach was pioneered in Hotz and Miller (1993) and Hotz, Miller, Sanders, and Smith (1994). Subsequent methodological developments in this vein include Aguirregabiria and Mira (2002), (2007), Pesendorfer and Schmidt-Dengler (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), and Hong and Shum (2007).[1] Alternatively, it is possible to use our identification results for the CCP's and laws of motions for the state variables as a "first-step" in an argument for identification of the per-period utility functions, in the spirit of Magnac and Thesmar (2002) and Bajari, Chernozhukov, Hong, and Nekipelov (2007), who considered the case of dynamic discrete-choice models without serially correlated unobserved state variables.

A general criticism of these CCP-based methods is that they cannot accommodate unobservables which are persistent over time. However, there are some recent papers focusing on the CCP-based estimation of dynamic discrete-choice models, in the presence of a latent state variable X_t^*. Aguirregabiria and Mira (2007), Buchinsky, Hahn, and Hotz (2004), and Houde and Imai (2006) develop estimation methodologies which accommodate agent-specific unobserved heterogeneity, corresponding to the case where the latent X_t^* is time-invariant. Arcidiacono and Miller (2006) develop a CCP-based approach to estimate dynamic discrete models where X_t^* can vary over time according to an exogenous and discrete first-order Markov process.[2]

While these papers have focused on estimation, our focus is on identification. Our identification approach is novel because it is based on recent econometric results in nonlinear measurement error models.[3] Specifically, we show that the identification results in Hu and Schennach (2008) and Carroll, Chen, and Hu (2008) for nonclassical measurement models (where the measurement error is not assumed to be independent of the latent "true"

[1] Applications applying the CCP insights to dynamic settings have grown quickly in recent years, and include Collard-Wexler (2006), Ryan (2006), and Dunne, Klimer, Roberts, and Xu (2006). See the discussion in Pakes (2008, section 3) and Ackerberg, Benkard, Berry, and Pakes (2007). All of these papers apply the CCP insight to dynamic games, which are more complex multi-agent generalizations of the single-agent dynamic setting consider in this paper.

[2] Several recent papers have focused on the estimation of parametric dynamic models with unobserved state variables, using non-CCP-based approaches. Imai, Jain, and Ching (2005) and Norets (2006) consider Bayesian MCMC estimation. Fernandez-Villaverde and Rubio-Ramirez (2007) develop an efficient simulation procedure (based on particle filtering) for estimating these models via simulation.

[3] See Li (2002) and Schennach (2004), (2007) for recent papers on nonlinear measurement error models, and Chen, Hong, and Nekipelov (2007) for a detailed survey.

variable) can be applied to Markovian dynamic models, and we use those results to establish nonparametric identification.

Kasahara and Shimotsu (2007, hereafter KS) consider the nonparametric identification of dynamic models in the presence of discrete unobserved heterogeneity, corresponding to the case where the latent variable X_t^* is time-invariant and discrete. KS prove the nonparametric identification of the Markov kernel $W_{t+1}|W_t, X^*$ in this setting, using six periods of data. KS's results rely on a matrix diagonalization argument, similar to the spectral decomposition arguments used in our identification proofs. However, our framework is more general than in KS, in that we allow X_t^* to vary over periods, and also to be drawn from a continuous distribution.

Henry, Kitamura, and Salanie (2008, hereafter HKS) exploit exclusion restrictions to identify Markov regime-switching models with a discrete and latent state variable. While our identification arguments, which rely on recent econometric results for nonclassical measurement error models, are quite distinct from those in HKS, our results share some of the intuition of HKS's findings, because we also exploit the feature of first-order Markovian models that, conditional on W_{t-1}, W_{t-2} is an "excluded variable" which affects W_t only via the unobserved state X_t^*.[4]

Cunha, Heckman, and Schennach (2006) apply the result of Hu and Schennach (2008) to show nonparametric identification of a nonlinear factor model consisting of $(W_t, W_t', W_t'', X_t^*)$, where the observed processes $\{W_t\}_{t=1}^T$, $\{W_t'\}_{t=1}^T$, and $\{W_t''\}_{t=1}^T$ constitute noisy measurements of the latent process $\{X_t^*\}_{t=1}^T$, contaminated with random disturbances. In contrast, we consider a setting where (W_t, X_t^*) jointly evolves as a dynamic Markov process. We use observations of W_t in different periods t to identify the conditional density of $\left(W_t, X_t^*|W_{t-1}, X_{t-1}^*\right)$. Thus, our model and identification strategy differ from theirs.

The paper is organized as follows. In Section 2, we introduce and discuss the main assumptions we make for identification. In Section 3, we present, in a sequence of lemmas, the proof of our main identification result. Subsequently, we also present several useful corollaries which follow from the main identification result. In Section 4, we discuss several examples, including a discrete case, to make our assumptions more transparent. We conclude in Section 5. While the proof of our main identification result is presented in the main text, the appendix contains the proofs for several lemmas and corollaries.

[4]In a similar vein, Bouissou, Laffont, and Vuong (1986) exploit the Markov restrictions on a stochastic process X to formulate tests for the noncausality of another process Y on X.

2 Overview of assumptions

Consider a dynamic process $\{(W_T, X_T^*), \dots, (W_t, X_t^*), ..., (W_1, X_1^*)\}_i$ for agent i. We assume that for each agent i, $\{(W_T, X_T^*), \dots, (W_t, X_t^*), ..., (W_1, X_1^*)\}_i$ is an independent random draw from a bounded distribution $f_{(W_T, X_T^*), \dots, (W_t, X_t^*), ..., (W_1, X_1^*)}$. The researcher observes a panel dataset consisting of an i.i.d. random sample of $\{W_T, W_{T-1}, \dots, W_1\}_i$, with $T \geq 5$, for many agents i.

We first consider identification in the nonstationary case, where the Markov law of motion $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ varies across periods. Note that this model subsumes the case of unobserved heterogeneity, in which X_t^* is fixed across all periods.

Below, we introduce our four assumptions. The first assumption below restricts attention to certain classes of models, while Assumptions 2-4 establish identification for the restricted class of models. Unless otherwise stated, all assumptions are taken to hold for all periods t.

2.1 Model restrictions

Assumption 1 *Model restrictions:*
(i) First-order Markov:

$$f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*, \Omega_{<t-1}} = f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*},$$

where $\Omega_{<t-1} \equiv \left\{W_{t-2}, ..., W_1, X_{t-2}^*, ..., X_1^*\right\}$*, the history of the process up to (but not including) period* $t-1$*.*
(ii) Limited feedback:

$$f_{W_t | W_{t-1}, X_t^*, X_{t-1}^*} = f_{W_t | W_{t-1}, X_t^*}.$$

Assumption 1(i) is just a first-order Markov assumption, which is satisfied for Markovian dynamic decision models (cf. Rust (1994)). Assumption 1(ii) is a "limited feedback" assumption, because it rules out direct feedback from the last period's USV, X_{t-1}^*, on the current value of the observed component W_t. When $W_t = (Y_t, M_t)$, where Y_t denotes the agent's action in period t, and M_t denotes the period-t observed state variable, Assumption 1 implies that:

$$\begin{aligned}
f_{W_t | W_{t-1}, X_t^*, X_{t-1}^*} &= f_{Y_t, M_t | Y_{t-1}, M_{t-1}, X_t^*, X_{t-1}^*} \\
&= f_{Y_t | M_t, Y_{t-1}, M_{t-1}, X_t^*, X_{t-1}^*} \cdot f_{M_t | Y_{t-1}, M_{t-1}, X_t^*, X_{t-1}^*} \\
&= f_{Y_t | M_t, X_t^*, Y_{t-1}, M_{t-1}} \cdot f_{M_t | Y_{t-1}, M_{t-1}, X_t^*}.
\end{aligned}$$

In the bottom line of the above display, the limited feedback assumption eliminates X^*_{t-1} as a conditioning variable in both terms. In Markovian dynamic optimization models, the first term (corresponding to the CCP) further simplifies to $f_{Y_t|M_t,X^*_t}$, because the Markovian laws of motion for the state variables (M_t, X^*_t) imply that the optimal policy function depends just on the current state variables, but not past values. Hence, Assumption 1 imposes weaker restrictions on the first term than Markovian dynamic optimization models.[5]

In the second term of the above display, the limited feedback condition rules out direct feedback from last period's unobserved state variable X^*_{t-1} to the current observed state variable M_t. However, it allows indirect effects via X^*_{t-1}'s influence on Y_{t-1} or M_{t-1}. Indeed, most empirical applications of dynamic optimization models with unobserved state variables satisfy the Markov and limited feedback conditions above. Examples of models in the industrial organization setting satisfying these conditions include Erdem, Imai, and Keane (2003), Crawford and Shum (2005), Das, Roberts, and Tybout (2007), Xu (2007), and Hendel and Nevo (2007). Finally, note that when X^*_t stands for unobserved heterogeneity and is time invariant, so that $X^*_t = X^*_{t-1}$, the limited feedback assumption is trivial.

Implicitly, the limited feedback assumption 1(ii) imposes a timing restriction, that X^*_t is realized before M_t, so that M_t depends on X^*_t. On the one hand, this is less restrictive than the assumption that M_t evolves independently of both X^*_{t-1} and X^*_t, which has been made in many applied settings, to enable the estimation of the M_t law of motion directly from the data. On the other hand, the limited feedback assumption does rule out models such as $M_t = h(M_{t-1}, X^*_{t-1}) + \eta_t$, which implies the alternative timing assumption that X^*_t is realized after M_t.

Moreover, our identification results can be extended to a general k-th order Markov process ($k < \infty$). For example, in a second-order Markov model, the analogous limited feedback condition allows W_t to depend on X^*_{t-1}, i.e., $f_{W_t|W_{t-1},W_{t-2},X^*_t,X^*_{t-1}}$. However, identifying a higher-order Markov law of motion would require additional data and stronger assumptions: it can be shown that the $3k+2$ observations $W_{t+k}, \ldots, W_{t-2k-1}$ can identify the k-th order Markov law of motion $f_{W_t,X^*_t|W_{t-1},\ldots,W_{t-k},X^*_{t-1},\ldots,X^*_{t-k}}$, under appropriate extensions of the assumptions presented in this section. We do not explore this here.

[5] Moreover, if we move outside the class of these models, the above display also shows that Assumption 1 does not rule out the dependence of Y_t on Y_{t-1} or M_{t-1}, which corresponds to some models of state dependence. These may include linear or nonlinear panel data models with lagged dependent variables, and serially correlated errors, cf. Arellano and Honore (2000). Arellano (2003, chs. 7–8) considers linear panel models with lagged dependent variables and persistent unobservables, which is also related to our framework.

2.2 Identification assumptions

For this paper, we assume that the unobserved state variable X_t^* is scalar-valued, and is drawn from a continuous distribution.[6] Hence, our identification results rely on linear operator-theoretic arguments, for which we must introduce some notation. Let R_1, R_2, R_3 denote three random variables, with support $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_3$, distributed according to the joint density $f_{R_1,R_2,R_3}(r_1, r_2, r_3)$.[7] Let $\mathcal{L}^p(\mathcal{X})$, $1 \leq p < \infty$ denote the space of functions $h(\cdot)$ with $\int_{\mathcal{X}} |h(x)|^p dx < \infty$. For any $1 \leq p \leq \infty$, we define an integral operator $L_{R_1|r_2,R_3} : \mathcal{L}^p(\mathcal{R}_3) \to \mathcal{L}^p(\mathcal{R}_1)$ for a given r_2 and any $h \in \mathcal{L}^p(\mathcal{R}_3)$, $\left(L_{R_1|r_2,R_3}h\right)(r_1) = \int f_{R_1|R_2,R_3}(r_1|r_2, r_3)h(r_3)dr_3$. Similarly, we define the diagonal operator $D_{r_1|r_2,R_3} : \mathcal{L}^p(\mathcal{R}_3) \to \mathcal{L}^p(\mathcal{R}_3)$ for a given (r_1, r_2) and any $h \in \mathcal{L}^p(\mathcal{R}_3)$, $\left(D_{r_1|r_2,R_3}h\right)(r_3) = f_{R_1|R_2,R_3}(r_1|r_2, r_3)h(r_3)$.

As we show in the next section, the crucial step in our identification argument is a spectral decomposition of a linear operator generated from $L_{W_{t+1},w_t|w_{t-1},W_{t-2}}$, which corresponds to the observed density $f_{W_{t+1},W_t|W_{t-1},W_{t-2}}$. (A spectral decomposition is the operator analog of the eigenvalue-eigenvector decomposition for matrices, in the finite-dimensional case.)[8] The next two assumptions guarantee the validity and uniqueness of this spectral decomposition.

Assumption 2 *Invertibility: There exists variable(s) $V \subseteq W$ such that for any $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$; equivalently, $E(h(V_{t+1})|w_t, w_{t-1}, v_{t-2}) = 0$ for any v_{t-2} implies $h = 0$.*
(i) $L_{V_{t-2},w_{t-1},w_t,V_{t+1}}$ is one-to-one.
(ii) $L_{V_{t+1}|w_t,X_t^*}$ is one-to-one.
(iii) L_{V_{t-2},w_{t-1},V_t} is one-to-one; equivalently, $E(h(V_t)|w_{t-1}, v_{t-2}) = 0$ for any v_{t-2} implies $h = 0$.

Assumption 2 restricts three operators to be one-to-one, which implies that they are invertible. The alternative statements of 2(i) and 2(iii) are the same as *completeness* conditions which have recently been employed in the nonparametric IV literature.[9] An operator $L_{R_1|r_2,R_3}$ is one-to-one if the corresponding density function $f_{R_1|r_2,R_3}$ satisfies a "complete-

[6] A discrete distribution for X_t^*, which is assumed in many applied settings (eg. Arcidiacono and Miller (2006)) is a special case, which we will consider as an example in Section 4 below.

[7] In this notation, capital letters denote random variables, while lower-case letters denote realizations of the random variables.

[8] Specifically, when W_t, X_t^* are both scalar and discrete with J $(< \infty)$ points of support, the operator $L_{W_{t+1},w_t|w_{t-1},W_{t-2}}$ is a $J \times J$ matrix, and spectral decomposition reduces to diagonalization of this matrix. This discrete case is discussed in detail in Section 4, example 1.

[9] See, for instance, Newey and Powell (2003), Blundell, Chen, and Kristensen (2007), d'Haultfoeuille (2008),Hu and Schennach (2008),Carrasco, Florens, and Renault (2005).

ness" condition; that is: for any r_2,

$$\left(L_{R_1|r_2,R_3}h\right)(r_1) = \int f(r_1|r_2,r_3)h(r_3)dr_2 = 0 \text{ for all } r_1 \text{ implies } h(r_3) = 0 \text{ for all } r_3. \quad (2)$$

We will elaborate on this condition in Section 4, when we discuss the one-to-one assumptions for several examples.

The variable(s) $V_t \subseteq W_t$ defined in Assumption 2 may be scalar, multidimensional, or W_t itself. Intuitively, by Assumption 2(ii), the variable(s) V_{t+1} are components of W_{t+1} which "transmit" information on the latent X_t^* conditional on W_t, the observables in the previous period. We consider suitable choices of V for specific examples in Section 4.[10]

A necessary condition for Assumption 2(i) is that the conditional density $f_{V_{t+1}|W_t,W_{t-1},V_{t-2}}$ depends on V_{t-2}, which is testable from the data. This rules out, for instance, the case where $V_{t+1} = h(W_t, W_{t-1}) + \eta_{t+1}$ with η_{t+1} being white noise. Similarly, Assumption 2(iii) is also testable from the data.

Assumption 2(ii) rules out models where X_t^* has a continuous support, but W_{t+1} contains only discrete components. In this case, there is no subset $V_{t+1} \subseteq W_{t+1}$ for which $L_{V_{t+1}|w_t,X_t^*}$ can be one-to-one. Hence, dynamic discrete-choice models with a continuous unobserved state variable X_t^*, but only discrete observed state variables M_t, fail this assumption, and may be nonparametrically underidentified without further assumptions. Moreover, models where the W_t and X_t^* processes evolve independently will also fail this assumption.

Assumption 3 *Uniqueness of spectral decomposition:*
(i) For any $(w_t, w_{t-1}, x_t^*) \in \mathcal{W}_t \times \mathcal{W}_{t-1} \times \mathcal{X}_t^*$, the density $f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*)$ is bounded away from zero and infinity.
(ii) For any $w_t \in \mathcal{W}_t$ and any $\overline{x}_t^* \neq \widetilde{x}_t^* \in \mathcal{X}_t^*$, there exists $w_{t-1} \in \mathcal{W}_{t-1}$ such that the density $f_{W_t|W_{t-1},X_t^*}$ satisfies

$$\frac{\partial^2}{\partial z_t \partial z_{t-1}} \ln f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, \overline{x}_t^*) \neq \frac{\partial^2}{\partial z_t \partial z_{t-1}} \ln f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, \widetilde{x}_t^*), \quad (3)$$

where z_t (resp. z_{t-1}) denotes a continuous-valued component of w_t (resp. w_{t-1}).

Assumption 3 ensures the uniqueness of the spectral decomposition of a linear operator generated from $L_{V_{t+1},w_t|w_{t-1},V_{t-2}}$. As we will discuss below, the eigenvalues in this decomposition involve the density $f_{W_t|W_{t-1},X_t^*}$, and conditions (i) and (ii) ensure that these eigen-

[10]There may be multiple choices of V which satisfy Assumption 2. In this case, the model may be overidentified, and it may be possible to do specification testing. We do not explore this possibility here.

values are, respectively, bounded and vary in x_t^*.[11] For 3(ii), a sufficient condition is that $\frac{\partial^3}{\partial z_t \partial z_{t-1} \partial x_t^*} \ln f_{W_t|W_{t-1},X_t^*}$ is continuous and nonzero, which implies that $\frac{\partial^2}{\partial z_t \partial z_{t-1}} \ln f_{W_t|W_{t-1},X_t^*}$ is monotonic in x_t^* for any (w_t, w_{t-1}).

Since our arguments are nonparametric, and X_t^* is unobserved, we need a normalization to pin down the the values of X_t^* relative to the values of the observables. For this purpose, we make a monotonicity assumption (similar to Matzkin (2003) and Hu and Schennach (2008)):

Assumption 4 *Normalization: For any given $w_t \in \mathcal{W}_t$, there exists a known functional G such that $G\left[f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)\right]$ is monotonic in x_t^*. Without loss of generality, we normalize x_t^* as $x_t^* = G\left[f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)\right]$.*

The functional G, which may depend on the value of w_t, could be the mean, mode, median, or another quantile of $f_{V_{t+1}|W_t,X_t^*}$. Assumptions 1-4 are the four main assumptions underlying our identification arguments. Of these four assumptions, all except Assumption 2(i) and 2(iii) involve densities not directly observed in the data, and are not directly testable.

3 Main nonparametric identification results

We present our argument for the nonparametric identification of the Markov law of motion $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ by way of several intermediate lemmas. The first two lemmas present convenient representations of the operators corresponding to the observed density $f_{V_{t+1},w_t|w_{t-1},V_{t-2}}$ and the Markov law of motion $f_{w_t,X_t^*|w_{t-1},X_{t-1}^*}$, for given values of $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$:

Lemma 1 (Representation of the observed density $f_{V_{t+1},w_t|w_{t-1},V_{t-2}}$):
For any $t \in \{3, \ldots, T-1\}$, Assumption 1 implies that for any given $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$,

$$L_{V_{t+1},w_t|w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}}. \tag{4}$$

Lemma 2 (Representation of Markov law of motion):
For any period $t \in \{3, \ldots, T-1\}$, Assumptions 1 and 2 imply that, for any given $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$,

$$L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} = L^{-1}_{V_{t+1}|w_t,X_t^*} L_{V_{t+1},w_t|w_{t-1},V_{t-2}} L^{-1}_{V_t|w_{t-1},V_{t-2}} L_{V_t|w_{t-1},X_{t-1}^*}. \tag{5}$$

[11] It turns out that Assumptions 2 and 3, as stated here, are stronger than necessary. An earlier version of the paper (Hu and Shum (2008)) contained less restrictive, but also less intuitive, versions of these assumptions.

Proofs: in Appendix.

Since $L_{V_{t+1},w_t|w_{t-1},V_{t-2}}$ and $L_{V_t|w_{t-1},V_{t-2}}$ are observed, Lemma 2 implies that the identification of the operators $L_{V_{t+1}|w_t,X_t^*}$ and $L_{V_t|w_{t-1},X_{t-1}^*}$ implies the identification of $L_{w_t,X_t^*|w_{t-1},X_{t-1}^*}$, the operator corresponding to the Markov law of motion. The next lemma postulates that $L_{V_{t+1}|w_t,X_t^*}$ is identified just from observed data.

Lemma 3 (Identification of $f_{V_{t+1}|W_t,X_t^*}$):
For any period $t \in \{3, \dots, T-1\}$, Assumptions 1, 2, 3, 4 imply that the density $f_{V_{t+1},W_t|W_{t-1},V_{t-2}}$ uniquely identifies the density $f_{V_{t+1}|W_t,X_t^*}$.

This lemma encapsulates the heart of the identification argument, which is the identification of $f_{V_{t+1}|W_t,X_t^*}$ via a spectral decomposition of an operator generated from the observed density $f_{V_{t+1},W_t|W_{t-1},V_{t-2}}$. Once this is established, re-applying Lemma 3 to the operator corresponding to the observed density $f_{V_t,W_{t-1}|W_{t-2},V_{t-3}}$ yields the identification of $f_{V_t|W_{t-1},X_{t-1}^*}$. Once $f_{V_{t+1}|W_t,X_t^*}$ and $f_{V_t|W_{t-1},X_{t-1}^*}$ are identified, then so is the Markov law of motion $f_{w_t,X_t^*|w_{t-1},X_{t-1}^*}$, from Lemma 2.

Proof. (Lemma 3)
By Lemma 1,

$$L_{V_{t+1},w_t|w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}}.$$

The first term on the RHS, $L_{V_{t+1}|w_t,X_t^*}$, does not depend on w_{t-1}, and the last term $L_{X_t^*|w_{t-1},V_{t-2}}$ does not depend on w_t. This feature suggests that, by evaluating Eq. (4) at the four pairs of points (w_t, w_{t-1}), $(\bar{w}_t, w_{t-1})$, $(w_t, \bar{w}_{t-1})$, $(\bar{w}_t, \bar{w}_{t-1})$, such that $w_t \neq \bar{w}_t$ and $w_{t-1} \neq \bar{w}_{t-1}$, each pair of equations will share one operator in common. Specifically:

$$\text{for } (w_t, w_{t-1}) \quad : \quad L_{V_{t+1},w_t|w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}}, \tag{6}$$

$$\text{for } (\overline{w}_t, w_{t-1}) \quad : \quad L_{V_{t+1},\overline{w}_t|w_{t-1},V_{t-2}} = L_{V_{t+1}|\overline{w}_t,X_t^*} D_{\overline{w}_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}}, \tag{7}$$

$$\text{for } (w_t, \overline{w}_{t-1}) \quad : \quad L_{V_{t+1},w_t|\overline{w}_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} L_{X_t^*|\overline{w}_{t-1},V_{t-2}}, \tag{8}$$

$$\text{for } (\overline{w}_t, \overline{w}_{t-1}) \quad : \quad L_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},V_{t-2}} = L_{V_{t+1}|\overline{w}_t,X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L_{X_t^*|\overline{w}_{t-1},V_{t-2}}. \tag{9}$$

Assumption 2(i) guarantees that the left-hand side operators can be inverted. Postmultiplying Eq. (6) by the inverse of Eq. (7) leads to

$$\begin{aligned}
\mathbf{A} &\equiv L_{V_{t+1},w_t|w_{t-1},V_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|w_{t-1},V_{t-2}} \\
&= L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}} \left(L_{V_{t+1}|\overline{w}_t,X_t^*} D_{\overline{w}_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}} \right)^{-1} \\
&= L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*}, \qquad (10)
\end{aligned}$$

eliminating $L_{X_t^*|w_{t-1},V_{t-2}}$. Similarly, postmultiplying Eq. (8) by the inverse of Eq. (9): eliminates $L_{X_t^*|\overline{w}_{t-1},Z_{t-2}}$:

$$\begin{aligned}
\mathbf{B} &\equiv L_{V_{t+1},w_t|\overline{w}_{t-1},V_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},V_{t-2}} \\
&= L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} D^{-1}_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*}. \qquad (11)
\end{aligned}$$

Finally, we postmultiply Eq. (10) by the inverse of Eq. (11) to obtain

$$\begin{aligned}
\mathbf{AB}^{-1} &= L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*} \times \\
&\quad \times \left(L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} D^{-1}_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*} \right)^{-1} \\
&= L_{V_{t+1}|w_t,X_t^*} \left(D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*} \right) L^{-1}_{V_{t+1}|w_t,X_t^*} \\
&\equiv L_{V_{t+1}|w_t,X_t^*} D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|w_t,X_t^*}, \qquad \text{where} \qquad (12)
\end{aligned}$$

$$\begin{aligned}
\left(D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} h \right)(x_t^*) &= \left(D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*} h \right)(x_t^*) \\
&= \frac{f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1},x_t^*) f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|\overline{w}_{t-1},x_t^*)}{f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|w_{t-1},x_t^*) f_{W_t|W_{t-1},X_t^*}(w_t|\overline{w}_{t-1},x_t^*)} h(x_t^*) \qquad (13) \\
&\equiv k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},x_t^*\right) h\left(x_t^*\right).
\end{aligned}$$

This equation implies that the observed operator $\mathbf{AB}^{-1}$ on the left hand side of Eq. (12) has an inherent eigenvalue-eigenfunction decomposition, with the eigenvalues corresponding to the function $k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},x_t^*\right)$ and the eigenfunctions corresponding to the density $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t,x_t^*)$. The decomposition in Eq. (12) is similar to the decomposition in Hu and Schennach (2008) or Carroll, Chen, and Hu (2008).

Assumption 3 ensures that this decomposition is unique. Specifically, Eq. (12) implies that the operator $\mathbf{AB}^{-1}$ on the LHS has the same spectrum as the diagonal operator $D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}$. Assumption 3(i) guarantees that the spectrum of the diagonal operator

$D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}$ is bounded. Since an operator is bounded by the largest element of its spectrum, Assumption 3(i) also implies that the operator $\mathbf{AB}^{-1}$ is bounded, whence we can apply Theorem XV.4.3.5 from Dunford and Schwartz (1971) to show the uniqueness of the spectral decomposition of bounded linear operators.

Several ambiguities remain in the spectral decomposition. First, Eq. (12) itself does not imply that the eigenvalues $k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*)$ are distinctive for different values x_t^*. When the eigenvalues are the same for multiple values of x_t^*, the corresponding eigenfunctions are only determined up to an arbitrary linear combination, implying that they are not identified. Assumption 3(ii) rules out this possibility. In the case that w_t (resp. w_{t-1}) is close to $\bar{w}_t$ (resp. $\bar{w}_{t-1}$), Eq. (13) implies that the logarithm of the eigenvalues in this decomposition can be represented as a second-order derivative of the log-density $f_{W_t|W_{t-1},X_t^*}$ as in Assumption 3(ii). Therefore, Assumption 3(ii) implies that for each w_t, we can find values $\bar{w}_t$, w_{t-1}, and $\bar{w}_{t-1}$ such that the eigenvalues are distinctive for all x_t^*.[12]

Second, the eigenfunctions $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ in the spectral decomposition (12) are unique up to multiplication by a scalar constant. However, these are density functions, so their scale is pinned down because they must integrate to one. Finally, both the eigenvalues and eigenfunctions are indexed by X_t^*. Since X_t^* is unobserved, the eigenfunctions are only identified up to an arbitrary one-to-one transformation of X_t^*. To resolve this issue, we need additional structure deriving from the economics or stochastic assumptions of the model, which "pin down" the values of the unobserved X_t^* relative to the observed variables. In Assumption 4, this additional structure comes in the form of the functional G which, when applied to the family of densities $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, \cdot)$ is monotonic in X_t^*, given w_t.[13] Given this monotonicity, we can normalize X_t^* by setting, $x_t^* = G\left[f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)\right]$ without loss of generality.

Therefore, altogether the density $f_{V_{t+1}|W_t,X_t^*}$ or $L_{V_{t+1}|w_t,X_t^*}$ is nonparametrically identified for any given $w_t \in \mathcal{W}_t$ via the spectral decomposition in Eq. (12). ■

By re-applying Lemma 3 to the observed density $f_{V_t,W_{t-1}|W_{t-2},V_{t-3}}$, it follows that the density $f_{V_t|W_{t-1},X_{t-1}^*}$ is identified.[14] Hence, by Lemma 2, we have shown the following result:

---

[12] Specifically, the operators $\mathbf{AB}^{-1}$ correpsonding to different values of $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1})$ share the same eigenfunctions $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$. Assumption 3(ii) implies that, for any two different eigenfunctions $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ and $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, \widetilde{x}_t^*)$, one can always find values of $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1})$ such that the two different eigenfunctions correspond to two different eigenvalues, i.e., $k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*) \neq k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, \widetilde{x}_t^*)$.

[13] See also Matzkin (2003) for a similar use of monotonicity in the nonparametric identification of nonseparable structural models.

[14] Recall that Assumptions 1-4 are assumed to hold for all periods t. Hence, applying Lemma 3 to the observed density $f_{V_t,W_{t-1}|W_{t-2},V_{t-3}}$ does not require any additional assumptions.

Theorem 1 (Identification of Markov law of motion, non-stationary case): *Under the Assumptions 1, 2, 3, and 4, the density* $f_{W_{t+1},W_t,W_{t-1},W_{t-2},W_{t-3}}$ *for any* $t \in \{4, \dots, T-1\}$ *uniquely determines the density* $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$.

3.1 Initial conditions

Some CCP-based estimation methodologies for dynamic optimization models (eg. Hotz, Miller, Sanders, and Smith (1994), Bajari, Benkard, and Levin (2007)) require simulation of the Markov process $(W_t, X_t^*, W_{t+1}, X_{t+1}^*, W_{t+2}, X_{t+2}^*, \dots)$ starting from some initial values W_{t-1}, X_{t-1}^*. When there are unobserved state variables, this raises difficulties because X_{t-1}^* is not observed.

However, it turns out that, as a by-product of the main identification results, we are also able to identify the marginal densities f_{W_{t-1},X_{t-1}^*}. For any given initial value of the observed variables w_{t-1}, knowledge of f_{W_{t-1},X_{t-1}^*} allows us to draw an initial value of X_{t-1}^* consistent with w_{t-1}.

Corollary 1 (Identification of initial conditions, non-stationary case): *Under the Assumptions 1, 2, 3, and 4, the density* $f_{W_{t+1},W_t,W_{t-1},W_{t-2},W_{t-3}}$ *for any* $t \in \{4, \dots, T-1\}$ *uniquely determines the density* f_{W_{t-1},X_{t-1}^*}.

Proof: in Appendix.

3.2 Stationarity

In the proof of Theorem 1 from the previous section, we only use the fifth period of data W_{t-3} for the identification of $L_{V_t|w_{t-1},X_{t-1}^*}$. Given that we identify $L_{V_{t+1}|w_t,X_t^*}$ using four periods of data, i.e., $\{W_{t+1}, W_t, W_{t-1}, W_{t-2}\}$, the fifth period W_{t-3} is not needed when $L_{V_t|w_{t-1},X_{t-1}^*} = L_{V_{t+1}|w_t,X_t^*}$. This is true when the Markov kernel density $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is time-invariant. Thus, in the stationary case, only four periods of data, $\{W_{t+1}, W_t, W_{t-1}, W_{t-2}\}$, are required to identify $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$. Formally, we make the additional assumption:

Assumption 5 *Stationarity: of the Markov law of motion of* (W_t, X_t^*) *is time-invariant:*

$$f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} = f_{W_2,X_2^*|W_1,X_1^*}, \ \forall\ 2 \leq t \leq T.$$

In dynamic optimization settings, this assumption is usually maintained in infinite-horizon models. Given the foregoing discussion, we present the next corollary without proof.

Corollary 2 (Identification of Markov law of motion, stationary case): *Under assumptions 1, 2, 3, 4, and 5, the observed density* $f_{W_{t+1},W_t,W_{t-1},W_{t-2}}$ *for any* $t \in \{3, \ldots, T-1\}$ *uniquely determines the density* $f_{W_2,X_2^*|W_1,X_1^*}$.

In the stationary case, initial conditions are still a concern. The following corollary, analogous to Corollary 1 for the non-stationary case, postulates the identification of the marginal density f_{W_t,X_t^*}, for periods $t \in \{1, \ldots, T-3\}$. For any of these periods, f_{W_t,X_t^*} can be used as a sampling density for the initial conditions.[15]

Corollary 3 (Identification of initial conditions, stationary case): *Under assumptions 1, 2, 3, 4, and 5, the observed density* $f_{W_{t+1},W_t,W_{t-1},W_{t-2}}$ *for any* $t \in \{3, \ldots, T-1\}$ *uniquely determines the density* f_{W_{t-2},X_{t-2}^*}.

Proof: in Appendix.

4 Comments on Assumptions in Specific Examples

Even though we focus on nonparametric identification, we believe that our results can be valuable for applied researchers working in a parametric setting, because they provide a guide for specifying models such that they are nonparametrically identified. As part of a pre-estimation check, our identification assumptions could be verified for a prospective model via either direct calculation, or Monte Carlo simulation using specific parameter values. If the prospective model satisfies the assumptions, then the researcher could proceed to estimation, with the confidence that underlying variation in the data, rather than the particular functional forms chosen, is identifying the model parameters, and not just the particular functional forms chosen. If some assumptions are violated, then our results suggest ways that the model could be adjusted in order to be nonparametrically identified.

To this end, in this section we present several examples of dynamic models. Because some of the assumptions that we made for our identification argument are quite abstract, we discuss these assumptions in the context of these examples.

4.1 Example 1: A discrete model

As a first example, let (W_t, X_t^*) denote a bivariate discrete first-order Markov process where W_t and X_t^* are both scalars, and binary: $\forall t$, $\text{supp}X_t^* = \text{supp}W_t \equiv \{0, 1\}$. This is the

[15] Note that even in the stationary case, where $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is invariant over time, the marginal density of f_{W_{t-1},X_{t-1}^*} may still vary over time (unless the Markov process (W_t, X_t^*) starts from the steady-state). For this reason, it is useful to identify f_{W_t,X_t^*} across a range of periods.

simplest example of the models considered in our framework. We assume that the laws of motion for both W_t and X_t^* exhibit state dependence:

$$Pr(W_t = 1|w_{t-1}, x_t^*) = p(w_{t-1}, x_t^*); \quad Pr(X_t^* = 1|x_{t-1}^*, w_{t-1}) = q(x_{t-1}^*, w_{t-1}) \tag{14}$$

These laws of motion satisfy Assumption 1.

This model is a binary version of Abbring, Chiappori, and Zavadil's (2008) "dynamic moral hazard" model of auto insurance. In that model, W_t is a binary indicator of claims occurrence, and X_t^* is a binary effort indicator, with $X_t^* = 1$ denoting higher effort. The mutual dependence of effort and claims in the laws of motion (14) arise from moral hazard, and experience rating in insurance pricing.

The main difference between this discrete case and the previous continuous case is that the linear integral operators are replaced by matrices. The L operators in the main proof correspond to 2×2 square matrices, and the D operators are 2×2 diagonal matrices.[16] Assumptions 2 and 3 are quite transparent to interpret in the matrix setting.

Assumption 2 implies the invertibility of certain matrices. From Lemma 1, the following matrix equality holds, for all values of (w_t, w_{t-1}):

$$L_{W_{t+1},w_t|w_{t-1},W_{t-2}} = L_{W_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},W_{t-2}}. \tag{15}$$

Given this equation, the invertibility of $L_{W_{t+1},w_t|w_{t-1},W_{t-2}}$ implies that $L_{W_{t+1}|w_t,X_t^*}$ and $L_{X_t^*|w_{t-1},W_{t-2}}$ are both invertible, and that all the elements in the diagonal matrix $D_{w_t|w_{t-1},X_t^*}$ are nonzero. Hence, in this discrete model, Assumptions 2(ii) is redundant, because it is implied by 2(i). That implies that Assumption 2 is fully testable from the observed data.

Assumption 3 puts restrictions on the eigenvalues in the spectral decomposition of the $\mathbf{AB^{-1}}$ operator. In the discrete case, $\mathbf{AB^{-1}}$ is an observed 2×2 matrix, and the spectral decomposition reduces to the usual matrix diagonalization. Assumption 3(i) implies that the eigenvalues are nonzero and finite, and 3(ii) implies that the eigenvalues are distinctive. For all values of (w_t, w_{t-1}), these assumptions can be verified, by directly diagonalizing the $\mathbf{AB^{-1}}$ matrix.

In this discrete case, Assumption 4 can be interpreted as an "ordering" assumption, which imposes an ordering on the columns of the $L_{W_{t+1}|w_t,X_t^*}$ matrix, corresponding to the eigenvectors of $\mathbf{AB^{-1}}$. If the goal is only to identify $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ for a single period t, then we could dispense with Assumption 4 altogether, and pick two arbitrary in recovering

[16]Specifically, for binary random variables R_1, R_2, R_3, the $(i+1, j+1)$-th element of the matrix L_{R_1,r_2,R_3} contains the joint probability that $(R_1 = i, r_2, R_3 = j)$, for $i, j \in \{0, 1\}$.

$L_{W_{t+1}|w_t,X_t^*}$ and $L_{W_t|w_{t-1},X_{t-1}^*}$. If we do this, we will not be able to pin down the exact value of X_t^* or X_{t-1}^*, but the recovered density of $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ will still be consistent with the two arbitrary orderings for X_t^* and X_{t-1}^* (in the sense that the implied transition matrix $X_t^*|X_{t-1}^*, w_{t-1}$ for every $w_{t-1} \in \mathcal{W}_{t-1}$ will be consistent with the true, but unknown ordering of X_t^* and X_{t-1}^*).[17]

But this will not suffice if we wish to recover the transition density $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ in two periods $t = t_1, t_2$, with $t_1 \neq t_2$. If we want to compare values of X_t^* across these two periods, then we must invoke Assumption 4 to pin down values of X_t^* which are consistent across the two periods. Hu (2008) suggests a number of ways to satisfy Assumption 4 in the discrete case. For example, a reasonable restriction, which satisfies Assumption 4, is that

$$\text{for } w_t = \{0,1\}: \quad f_{W_{t+1}|W_t,X_t^*}(0|w_t,1) > f_{W_{t+1}|W_t,X_t^*}(1|w_t,1),$$

which implies that claims occur less frequently with higher effort. This restriction can be ensured by appropriate restrictions on the values of the $p(\cdots)$ and $q(\cdots)$ functions in (14).

4.2 Example 2: Rust's (1987) bus engine replacement model

The second example is a version of Rust's (1987) bus-engine replacement model, augmented to allow for persistent unobserved state variables. In this model, $W_t = (Y_t, M_t)$, where Y_t is the indicator that the bus engine was replaced in week t, and M_t is the mileage since the last engine replacement.

We introduce two specifications of the model, which differ in how the unobserved state variable X_t^* enters. In both specifications, we will restrict X_t^* to have a bounded support: for $[0, U]$ such that $0 < U < +\infty$,

$$X_t^* = 0.5X_{t-1}^* + 0.3\psi\left(M_{t-1}\right) + 0.2\nu_t; \quad \psi\left(M_{t-1}\right) = U\frac{e^{M_{t-1}} - 1}{e^{M_{t-1}} + 1}. \tag{16}$$

ν_t is a truncated standard normal shock over the interval $[0, U]$, distributed independently over t. We also assume that the initial value $X_0^* \in [0, U]$, which guarantees that $X_t^* \in [0, U]$ for all t. Hence, $X_t^*|X_{t-1}^*, Y_{t-1}, M_{t-1}$ is distributed with density determined by $f_{\nu_t}(\cdot)$.

Let $S_t \equiv (M_t, X_t^*)$ denote the persistent state variables in this model. The period utility from each choice is additive in a function of the state variables S_t, and a choice-specific non-

[17] We thank Thierry Magnac for this insight.

persistent preference shock:

$$u_t = \begin{cases} u_0(S_t) + \epsilon_{0t} & \text{if } Y_t = 0 \\ u_1(S_t) + \epsilon_{1t} & \text{if } Y_t = 1 \end{cases}$$

where ϵ_{0t} and ϵ_{1t} are i.i.d. Type I Extreme Value shocks, which are independent over time, and also independent of the state variables S_t.

In **Specification A**, the choice-specific utility functions are:

$$u_0(S_t) = -c(M_t) + X_t^*; \quad u_1(S_t) = -RC. \tag{17}$$

In the above, $c(M_t)$ denotes the maintenance cost function, which is increasing in mileage M_t, and $0 < RC < +\infty$ denotes the cost of replacing the engine. We also assume that the maintenance cost function $c(\cdot)$ is bounded below and above: $c(0) = 0$; $\lim_{M\to+\infty} c(M) = \bar{c} < +\infty$. Mileage evolves as:

$$M_{t+1} - (1 - Y_t)M_t = \exp(\eta_{t+1}) \tag{18}$$

where $\eta_{t+1} > 0$ follows a standard normal random variable, truncated to $[0, 1]$, with density $\tilde{\phi}(\eta) \equiv \frac{\phi(\eta)}{\Phi(1)-\Phi(0)}$, where ϕ and Φ denote the standard normal density and CDF.[18] ■

In **Specification B**, the agent's per-period utility functions are given by:

$$u_0(S_t) = -c(M_t); \quad u_1(S_t) = -RC. \tag{19}$$

with the same assumptions on RC and $c(\cdot)$ as in Specification A. Mileage evolves as:

$$M_{t+1} - (1 - Y_t)M_t = \exp(\eta_{t+1} + X_{t+1}^*). \tag{20}$$

Hence, the main difference between the two specifications is that in Specification A, the unobserved state variable X_t^* affects utilities directly, but not the mileage process. In Specification B, X_t^* directly affects the evolution of mileage, but not the agent's utilities.[19] ■

Both specifications are stationary dynamic optimization models, in which the conditional choice probabilities take the multinomial logit form (for $Y_t = 0, 1$): $P(Y_t|S_t) = \exp\left(V_{Y_t}(S_t)\right) / \left[\sum_{y=0}^{1} \exp\left(V_y(S_t)\right)\right]$ where $V_y(S_t)$ is the choice-specific value function in pe-

[18]For this to be reasonable, assume that mileage is measured in units of 10,000 miles.

[19]Furthermore, following Rust's assumptions, in both specifications, previous mileage M_{t-1} has no direct effect on current mileage M_t when the engine was replaced in the previous period ($Y_{t-1} = 1$).

riod t, defined recursively by $V_y(S_t) = u_y(S_t) + \beta E\left[\log\left\{\sum_{y'=0}^{1}\exp\left(V_{y'}(S_{t+1})\right)\right\}|Y_t = y, S_t\right]$.

The first-order Markov and limited feedback assumptions are satisfied in both specifications. We discuss each remaining assumption in turn.

Assumption 2 contains three invertibility assumptions. For the V_t variables in Assumption 2, we use $V_t = M_t$, for all periods t. We begin by presenting a necessary condition for an operator to be one-to-one, which is useful to determine when one-to-one is *not* satisfied. Later, we will consider conditions which ensure that one-to-one *is* satisfied.

Lemma 4 (Necessary conditions for one-to-one): *If $L_{R_1|R_3}$ is one-to-one, then for any set $\mathcal{S}_3 \subseteq \mathcal{R}_3$ with $\Pr(\mathcal{S}_3) > 0$, there exists a set $\mathcal{S}_1 \subseteq \mathcal{R}_1$ such that $\Pr(\mathcal{S}_1) > 0$ and*

$$\frac{\partial}{\partial r_3} f_{R_1|R_3}(r_1|r_3) \neq 0 \textit{ almost surely for } \forall r_1 \in \mathcal{S}_1,\ \forall r_3 \in \mathcal{S}_3. \tag{21}$$

Proof: in Appendix.

Intuitively, the necessary condition ensures that there is enough variation in the distribution of R_1 for different values of R_3.

We first consider Assumption 2(i). For Specification A, pick an w_t such that $Y_t = 1$ (so that the engine is replaced in period t). In this case, $M_{t+1}|Y_t = 1$ follows a normal distribution truncated to $[0, 1]$, and does not depend stochastically on either w_{t-1} or M_{t-2}. Hence, Eq. (21) fails here, and hence so does Assumption 2(i).

Now consider Specification B, using the same w_t such that $Y_t = 1$. Because X_t^* directly enters the mileage process, the distribution of M_{t+1} depends on X_{t+1}^*. Similarly, the distribution of M_{t-2} depends on X_{t-2}^*. Since (X_{t+1}^*, X_{t-2}^*) are correlated, conditional on w_{t-1} (which does not include X_{t-1}^*), the distribution of M_{t+1} varies in M_{t-2}. The discussion of Assumption 2(iii) is very similar to that of 2(i), and we omit it for convenience here.

Assumption 2(ii) requires that, for all w_t, the mapping $L_{M_{t+1}|w_t, X_t^*}$ is one-to-one. As before, consider a value w_t such that $Y_t = 1$. In Specification A, $M_{t+1}|w_t, X_t^*$ is distributed according to a standard normal distribution truncated to $[0, 1]$, regardless of the value of X_t^*. Hence, Eq. (21) fails, and $L_{M_{t+1}|w_t, X_t^*}$ is not one-to-one. For Specification B, however, M_{t+1} is distributed according to a mixture distribution which depends on X_{t+1}^*. Since X_{t+1}^* and X_t^* are serial correlated, M_{t+1} will vary in X_t^*, for fixed w_t.

Hence, for Specification B, we cannot show that Assumption 2 is violated. Indeed, we can go further and directly confirm that Assumption 2 is satisfied, exploiting some sufficient conditions for an operator to be one-to-one, which have been provided in the existing literature. The details, which are specialized for this particular example, are given in the Appendix.

Assumption 3 contains two restrictions on the density $f_{W_t|W_{t-1},X_t^*}$, which factors as

$$f_{W_t|W_{t-1},X_t^*} = f_{Y_t|M_t,X_t^*} \cdot f_{M_t|Y_{t-1},M_{t-1},X_t^*}. \tag{22}$$

Assumption 3(i) requires that, for any (w_t, w_{t-1}), this density is bounded between 0 and $+\infty$. The first term is the CCP $f_{Y_t|M_t,X_t^*}$, which is a logit probability. Because the per-period utilities (under both specification A and B), net of the ϵ's, are bounded away from $-\infty$ and $+\infty$, the logit choice probabilities are also bounded away from zero. The second term is the mileage law of motion $f_{M_t|Y_{t-1},M_{t-1},X_t^*}$ which, by assumption, is a truncated normal distribution, so it is also bounded away from zero and $+\infty$. The bounded support assumption on M_t is crucial but, in practice, imply little loss in generality, because typically in estimating these models, one can take the upper and lower bounds on M_t from the observed data.

Assumption 3(ii) ensures that the eigenvalues in the decomposition (12) are distinctive. Because of the factorization (22), and the fact that the CCP's are bounded away from zero, a sufficient condition for Eq. (3) is that

$$\frac{\partial^2}{\partial m_t \partial m_{t-1}} \ln f_{M_t|Y_{t-1},M_{t-1},X_t^*}(m_t|y_{t-1}, m_{t-1}, x_t^*) \tag{23}$$

is strictly monotonic in x_t^*, for all m_t, x_t^*, and some $w_{t-1} = (y_{t-1}, m_{t-1})$.

For any value of m_t, pick any m_{t-1} such that $y_{t-1} = 0$ (ie., the bus engine was not replaced in period $t-1$). Under Specification B, the density of $M_t|Y_{t-1}, M_{t-1}, X_t^*$ for this pair of (m_t, m_{t-1}), is distributed with density $\widetilde{\phi}\left(\log\left(\frac{m_t - m_{t-1}}{\exp(x_t^*)}\right)\right) / [m_t - m_{t-1}]$ on the range $m_t \in [m_{t-1}, m_{t-1} + \exp(x_t^*)]$, where $\widetilde{\phi}(\cdot)$ denotes a truncated standard normal density. The second derivative of the log of this density is monotonic in x_t^*.

On the other hand, for Specification A, the sufficient condition (23) cannot be satisfied, because the conditional distribution $M_t|Y_{t-1}, M_{t-1}, X_t^*$ is never a function of x_t^*.

Assumption 4 presumes a known functional G such that $G\left[f_{M_{t+1}|Y_t,M_t,X_t^*}(\cdot|y_t, m_t, x_t^*)\right]$ is monotonic in x_t^*. For Specification B, Eqs. (16) and (20) imply that

$$M_{t+1} = (1 - Y_t)M_t + \exp(\eta_{t+1} + 0.2\nu_{t+1}) \cdot \exp(0.3\psi(M_t)) \cdot \exp(0.5X_t^*). \tag{24}$$

Let C_{med} denote the median of the random variable $\exp(\eta_{t+1}+0.2\nu_{t+1})$, which is a truncated log-normal random variable. Then

$$\text{med}\left[f_{M_{t+1}|Y_t,M_t,X_t^*}(\cdot|y_t,m_t,x_t^*)\right] = (1-y_t)m_t + C_{med}\cdot\exp(0.3\psi\left(m_t\right))\cdot\exp(0.5x_t^*)$$

which is monotonic in x_t^*. Hence, we can pin down $x_t^* = \text{med}\left[f_{M_{t+1}|Y_t,M_t,X_t^*}(\cdot|y_t,m_t,x_t^*)\right]$.

4.3 Example 3: generalized investment model

For the third example, we consider a dynamic model of firm R&D and product quality in the "generalized dynamic investment" framework described in Doraszelski and Pakes (2007). In this model, Y_t measures a firm's R&D, and X_t^* measures the firm's product quality, which evolves according to $X_{t+1}^* - X_t^* = h(Y_t)\cdot\exp(\nu_{t+1})$. The observed state variable M_t is installed base, which evolves as $M_{t+1} - M_t = k(X_{t+1}^*)\cdot\exp(\xi_{t+1})$ with $k'(\cdot) > 0$ (so that, *ceteris paribus*, product quality raises installed base).

Each period, a firm chooses its R&D to maximize the discounted future profits:

$$\begin{aligned} Y_t &= Y^*(M_t, X_t^*, \gamma_t) \\ &= \text{argmax}_y \big[\underbrace{\Pi(M_t, X_t^*)}_{\text{profits}} - \underbrace{\gamma_t}_{\text{shock}} \cdot \underbrace{c(Y_t, M_t, X_t^*)}_{\text{invt cost}} + \beta E \underbrace{V(M_{t+1}, X_{t+1}^*, \gamma_{t+1})}_{\text{value fxn}} \big] \end{aligned}$$

In the above, the errors (ξ_t, ν_t, γ_t) are assumed to be mutually independent, each distributed $N(0,1)$ i.i.d. across periods. These errors are introduced to induce randomness in (Y_t, M_t, X_t^*) conditional on $(Y_{t-1}, M_{t-1}, X_{t-1}^*)$. Since the mathematical structure of this example are very similar to those in the Rust example, we omit some details.

Obviously, Assumption 1 is satisfied with the above assumptions. As in Example 2, we use $V_t = M_t$, for all periods t here because, as noted in Levinsohn and Petrin (2000) and Ackerberg, Benkard, Berry, and Pakes (2007), Y_t may be equal to zero for many values of (M_t, X_t^*), and hence may not provide enough information on X_t^*. For Assumption 2, as in the Rust example, we focus on showing that the necessary condition (Lemma 4) is satisfied. For Assumption 2(i), note that M_{t+1} depends on X_{t+1}^*, which is correlated with M_{t-2}. Similarly, for Assumption 2(iii), note that M_t depends on X_t^*, which is correlated with M_{t-2}, so that the density $f_{M_t|w_{t-1},M_{t-2}}$ varies in M_{t-2}, for fixed w_{t-1}. For Assumption 2(ii), that $L_{M_{t+1}|w_t,X_t^*}$ is one-to-one, note that the conditional distribution of $M_{t+1}|w_t, X_t^*$ depends on X_t^*. Sufficient conditions for Assumption 2 to be satisfied can be constructed as in Appendix B for the Rust example, and we omit the details here.

The bounded eigenvalues restriction in Assumption 3(i) can be ensured by truncating supports of (γ_t, ξ_t, ν_t), and the range of the $k(\cdot)$, $h(\cdot)$, and $c(\cdot)$ functions, similarly to what was done for the Rust example above. For 3(ii), we derive that

$$\frac{\partial^2}{\partial m_t \partial m_{t-1}} \ln f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*) = \frac{\partial^2}{\partial m_t \partial m_{t-1}} \left[\ln f(y_t|m_t, x_t^*) + \ln f(m_t|w_{t-1}, x_t^*)\right].$$

The conditional density of $m_t|m_{t-1}, x_t^* \sim \phi\left(\log((m_t - m_{t-1})/k(x_t^*))\right) / \left[m_t - m_{t-1}\right]$, where ϕ denotes the standard normal density. This is decreasing in x_t^* for every (m_t, m_{t-1}), and implies the condition in Assumption 3(ii).

Finally, we note that

$$E[M_{t+1}|m_t, y_t, x_t^*] = m_t + \exp(0.5) \cdot E[k(X_{t+1}^*)|x_t^*, y_t].$$

Because the function $k(\cdot)$ is monotonic, the law of motion for product quality implies that $E[k(X_{t+1}^*)|x_t^*, y_t]$ is monotonic in x_t^*. Hence, we can set G as the mean functional, and pin down $x_t^* = E\left[f_{M_{t+1}|M_t,Y_t,X_t^*}(\cdot|m_t, y_t, x_t^*)\right]$.

5 Concluding remarks

We have considered the identification of a first-order Markov process $\{W_t, X_t^*\}_t$ when only $\{W_t\}_t$ is observed. Under non-stationarity, the Markov law of motion $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is identified from the distribution of the five observations $W_{t+1}, \ldots, W_{t-3}$ under reasonable assumptions. When stationarity is imposed, identification of $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ obtains with only four observations $W_{t+1}, \ldots, W_{t-2}$. Identification of $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is a crucial input in methodologies for estimating dynamic models based on the "conditional-choice-probability (CCP)" approach pioneered by Hotz and Miller. Once $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ is identified, nonparametric identification of the remaining parts of the models – particularly, the per-period utility functions – can proceed by straightforward application of the identification results in Magnac and Thesmar (2002) and Bajari, Chernozhukov, Hong, and Nekipelov (2007), which considered dynamic models without persistent latent variables X_t^*.

We have only considered the case where the unobserved state variable X_t^* is scalar-valued. An interesting extension is the case where X_t^* is a multivariate process, which may apply to dynamic game settings, where M_t and X_t^* may contain the set of, respectively, observed and unobserved state variables for all agents in the game.[20]

[20]However, when X_t^* is multi-dimensional, Assumption 2(ii), which requires that $L_{V_{t+1}|w_t,X_t^*}$ be one-to-one, can be quite restrictive. Ackerberg, Benkard, Berry, and Pakes (2007, Section 2.4.3) contains a discussion

Finally, this paper has focused on identification, but not estimation. In ongoing work, we are using our identification results to guide the specification and estimation of dynamic models with unobserved state variables.

A Proofs

Proof. (Lemma 1) By Assumption 1(i), the observed density $f_{W_{t+1},W_t,W_{t-1},W_{t-2}}$ equals

$$\begin{aligned}
& \int\int f_{W_{t+1},W_t,W_{t-1},W_{t-2},X_t^*,X_{t-1}^*} dx_t^* dx_{t-1}^* \\
= & \int\int f_{W_{t+1}|W_t,W_{t-1},W_{t-2},X_t^*,X_{t-1}^*} f_{W_t,X_t^*|W_{t-1},W_{t-2},X_{t-1}^*} f_{W_{t-1},W_{t-2},X_{t-1}^*} dx_t^* dx_{t-1}^* \\
= & \int\int f_{W_{t+1}|W_t,X_t^*} f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{W_{t-1},W_{t-2},X_{t-1}^*} dx_t^* dx_{t-1}^* \\
= & \int\int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} f_{X_t^*|W_{t-1},X_{t-1}^*} f_{W_{t-1},W_{t-2},X_{t-1}^*} dx_t^* dx_{t-1}^* \\
= & \int\int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} f_{X_t^*|W_{t-1},W_{t-2},X_{t-1}^*} f_{W_{t-1},W_{t-2},X_{t-1}^*} dx_t^* dx_{t-1}^* \\
= & \int\int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} f_{X_t^*,X_{t-1}^*,W_{t-1},W_{t-2}} dx_t^* dx_{t-1}^*.
\end{aligned}$$

(We omit all the arguments in the density functions.) Assumption 1(ii) then implies

$$\begin{aligned}
f_{W_{t+1},W_t,W_{t-1},W_{t-2}} &= \int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*} \left(\int f_{X_t^*,X_{t-1}^*,W_{t-1},W_{t-2}} dx_{t-1}^* \right) dx_t^* \\
&= \int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*} f_{X_t^*,W_{t-1},W_{t-2}} dx_t^*.
\end{aligned}$$

Hence, by combining the above two displays, we obtain

$$f_{W_{t+1},W_t|W_{t-1},W_{t-2}} = \int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*} f_{X_t^*|W_{t-1},W_{t-2}} dx_t^*. \tag{25}$$

In operator notation, given values of $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$, this is

$$L_{W_{t+1},w_t|w_{t-1},W_{t-2}} = L_{W_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},W_{t-2}}. \tag{26}$$

For the variable(s) $V_t \subseteq W_t$, for all periods t, introduced in Assumption 2, Eq. (26)

of the difficulties with multivariate unobserved state variables in the context of dynamic investment models.

implies that the joint density of $\{V_{t+1}, W_t, W_{t-1}, V_{t-2}\}$ is expressed in operator notation as

$$L_{V_{t+1},w_t|w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}}. \tag{27}$$

■

Proof. (Lemma 2) Assumption 1 implies the following two equalities:

$$\begin{aligned} f_{V_{t+1},W_t|W_{t-1},V_{t-2}} &= \int f_{V_{t+1}|W_t,X_t^*} f_{W_t,X_t^*|W_{t-1},V_{t-2}} dx_t^* \\ f_{W_t,X_t^*|W_{t-1},V_{t-2}} &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},V_{t-2}} dx_{t-1}^*. \end{aligned} \tag{28}$$

In operator notation, for fixed w_t, w_{t-1}, the above equations are expressed:

$$\begin{aligned} L_{V_{t+1},w_t|w_{t-1},V_{t-2}} &= L_{V_{t+1}|w_t,X_t^*} L_{w_t,X_t^*|w_{t-1},V_{t-2}} \\ L_{w_t,X_t^*|w_{t-1},V_{t-2}} &= L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*|w_{t-1},V_{t-2}}. \end{aligned}$$

Substituting the second line into the first, we get

$$\begin{aligned} L_{V_{t+1},w_t|w_{t-1},V_{t-2}} &= L_{V_{t+1}|w_t,X_t^*} L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*|w_{t-1},V_{t-2}} \\ \Leftrightarrow L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*|w_{t-1},V_{t-2}} &= L_{V_{t+1}|w_t,X_t^*}^{-1} L_{V_{t+1},w_t|w_{t-1},V_{t-2}}. \end{aligned} \tag{29}$$

Next, we eliminate $L_{X_{t-1}^*|w_{t-1},V_{t-2}}$ from the above. Again using Assumption 1, we have

$$f_{V_t|W_{t-1},V_{t-2}} = \int f_{V_t|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},V_{t-2}} dx_{t-1}^* \tag{30}$$

which, in operator notation (for fixed w_{t-1}), is

$$L_{V_t|w_{t-1},V_{t-2}} = L_{V_t|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*|w_{t-1},V_{t-2}} \quad \Rightarrow \quad L_{X_{t-1}^*|w_{t-1},V_{t-2}} = L_{V_t|w_{t-1},X_{t-1}^*}^{-1} L_{V_t|w_{t-1},V_{t-2}}$$

where the right-hand side applies Assumption 2(ii). Hence, substituting the above into Eq. (29), we obtain the desired representation

$$\begin{aligned} L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{V_t|w_{t-1},X_{t-1}^*}^{-1} L_{V_t|w_{t-1},V_{t-2}} &= L_{V_{t+1}|w_t,X_t^*}^{-1} L_{V_{t+1},w_t|w_{t-1},V_{t-2}} \\ \Rightarrow L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} &= L_{V_{t+1}|w_t,X_t^*}^{-1} L_{V_{t+1},w_t|w_{t-1},V_{t-2}} L_{V_t|w_{t-1},V_{t-2}}^{-1} L_{V_t|w_{t-1},X_{t-1}^*} \end{aligned} \tag{31}$$

where the second line applies Assumption 2(iii).[21] ■

Proof. (Lemma 4) Suppose Eq. (21) fails, so there exists an interval $\mathcal{S}_3 \equiv [\underline{r}, \bar{r}]$ such that, for $\forall r_3 \in \mathcal{S}_3$ and $\forall r_1 \in \mathcal{R}_1$, $\frac{\partial}{\partial r_3} f_{R_1|R_3}(r_1|r_3) = 0$. Define $h_0(r_3) = I_{\mathcal{S}_3}(r_3)\, g(r_3)$. Then

$$\begin{aligned}
\left(L_{R_1|R_3} h_0\right)(r_1) &= \int f_{R_1|R_3}(r_1|r_3) h_0(r_3) dr_3 \\
&= \int_{\mathcal{S}_3} f_{R_1|R_3}(r_1|r_3) g(r_3) dr_3 \\
&\equiv \int_{\mathcal{S}_3} f_{R_1|R_3}(r_1|r_3) dG(r_3) \\
&= f_{R_1|R_3}(r_1|r_3) G(r_3)\Big|_{\underline{r}}^{\overline{r}} - \int_{\mathcal{S}_3} G(r_3) \left(\frac{\partial}{\partial r_3} f_{R_1|R_3}(r_1|r_3) \right) dr_3 \\
&= f_{R_1|R_3}(r_1|\overline{r}) G(\overline{r}) - f_{R_1|R_3}(r_1|\underline{r}) G(\underline{r})
\end{aligned}$$

Notice that $f_{R_1|R_3}(r_1|\overline{r}) = f_{R_1|R_3}(r_1|\underline{r})$. Thus, for $\forall r_1 \in \mathcal{R}_1$

$$\left(L_{R_1|R_3} h_0\right)(r_1) = f_{R_1|R_3}(r_1|\overline{r}) \left[G(\overline{r}) - G(\underline{r})\right].$$

Then, pick any function g for which $G(\overline{r}) - G(\underline{r}) = \int_{\underline{r}}^{\bar{r}} g(r) dr = 0$, but $g(r) \neq 0$ for any r in a nontrivial subset of $[\underline{r}, \bar{r}]$. We have $L_{R_1|R_3} h_0 = 0$, but $h_0 \neq 0$. Therefore, Eq. (2) fails, and $L_{R_1|R_3}$ is not one-to-one. ■

Proof. (Corollary 1)

From Lemma 3, we know that $f_{V_t|W_{t-1},X_{t-1}^*}$ is identified from the observed density $f_{V_t,W_{t-1}|W_{t-2},V_{t-3}}$. The following equation

$$f_{V_t,W_{t-1}} = \int f_{V_t|W_{t-1},X_{t-1}^*} f_{W_{t-1},X_{t-1}^*} dx_{t-1}^*$$

implies that for any given $w_{t-1} \in \mathcal{W}_t$,

$$\begin{aligned}
f_{V_t,W_{t-1}=w_{t-1}} &= L_{V_t|w_{t-1},X_{t-1}^*} f_{W_{t-1}=w_{t-1},X_{t-1}^*} \\
\Leftrightarrow f_{W_{t-1}=w_{t-1},X_{t-1}^*} &= L_{V_t|w_{t-1},X_{t-1}^*}^{-1} f_{V_t,W_{t-1}=w_{t-1}}
\end{aligned}$$

where the second line applies Assumption 2(ii). Hence, f_{W_{t-1},X_{t-1}^*} is identified. ■

Proof. (Corollary 3)

[21] Eq. (29) also shows that Assumption 2(iii) could have been replaced by the invertibility of $L_{X_{t-1}^*|w_{t-1},V_{t-2}}$, since either assumption would have yielded the second row of Eq. (31). But we chose to assume the invertibility of $L_{V_t|w_{t-1},V_{t-2}}$ because it corresponds to an observed density function.

Under stationarity, the operator $L_{V_{t-1}|w_{t-2},X^*_{t-2}}$ is the same as $L_{V_{t+1}|w_t,X^*_t}$, which is identified from the observed density $f_{V_{t+1},W_t|W_{t-1},V_{t-2}}$ (by Lemma 3). Note that

$$f_{V_{t-1},W_{t-2}} = \int f_{V_{t-1}|W_{t-2},X^*_{t-2}} f_{W_{t-2},X^*_{t-2}} dx^*_{t-2}.$$

The same argument as in the proof of Corollary 1 then implies that we may identify $f_{W_{t-2},X^*_{t-2}}$ from the observed density $f_{V_{t-1},W_{t-2}}$. ■

B Additional Details on the Rust Example

In this section we present additional details for Specification B in Example 2 of Section 4, regarding Assumption 2. First, as shown in the proof of Lemma 2, Assumption 1 implies that

$$\begin{aligned} L_{M_{t+1},w_t|w_{t-1},M_{t-2}} &= L_{M_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} L_{X^*_t|w_{t-1},M_{t-2}} \\ &= L_{M_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} L_{X^*_t|w_{t-1},X^*_{t-1}} L_{X^*_{t-1}|w_{t-1},M_{t-2}} &\quad (32) \\ L_{M_t|w_{t-1},M_{t-2}} &= L_{M_t|w_{t-1},X^*_{t-1}} L_{X^*_{t-1}|w_{t-1},M_{t-2}}. &\quad (33) \end{aligned}$$

Furthermore, we have $L_{M_{t+1}|w_t,X^*_t} = L_{M_{t+1}|w_t,X^*_{t+1}} L_{X^*_{t+1}|w_t,X^*_t}$.

Hence, the invertibility of $L_{M_{t+1},w_t|w_{t-1},M_{t-2}}$, $L_{M_{t+1}|w_t,X^*_t}$, and $L_{M_t|w_{t-1},M_{t-2}}$, as required by Assumption 2, is implied by the invertibility of $L_{M_{t+1}|w_t,X^*_{t+1}}$, $D_{w_t|w_{t-1},X^*_t}$, $L_{X^*_t|w_{t-1},X^*_{t-1}}$ and $L_{X^*_{t-1}|w_{t-1},M_{t-2}}$.[22] It turns out that assumptions we have made already for this example ensure that three of these operators are invertible. We discuss each case in turn.

(i) For the diagonal operator $D_{w_t|w_{t-1},X^*_t}$, the inverse has a kernel function which is equal to $1/f_{w_t|w_{t-1},X^*_t}$. Hence, by Assumption 3(i), which guarantees that $f_{w_t|w_{t-1},X^*_t}$ is bounded away from 0 and ∞, the inverse exists.

(ii) For $L_{M_{t+1}|w_t,X^*_{t+1}}$, we use the fact, for every y_t, m_t, the corresponding density takes a convolution form, ie. $\log[M_{t+1} - (1 - Y_t)M_t] = X^*_{t+1} + \eta_{t+1}$. As is well-known, as long as the characteristic function of η_{t+1} has no real zeros, which is satisfied by the assumed truncated normal distribution, the corresponding operator is invertible.

(iii) Similarly, the density $f_{X^*_t|w_{t-1},X^*_{t-1}}$ is a convolution, ie. $X^*_t = 0.5X^*_{t-1} + 0.3\psi(M_{t-1}) + 0.2\nu_t$. Hence, the operator $L_{X^*_t|w_{t-1},X^*_{t-1}}$ is invertible if the characteristic function of ν_t has no real zeros, which is satisfied by the assumed normal distribution truncated to $[0, U]$.

(iv) For the last operator, corresponding to the density $f_{X^*_{t-1}|w_{t-1},M_{t-2}}$, the assumptions

[22]By stationarity, we do not need to consider $L_{X^*_{t+1}|w_t,X^*_t}$ and $L_{M_t|w_{t-1},X^*_{t-1}}$ separately.

made so far do not ensure that this operator is invertible. Given cases (ii) and (iii) immediately above, a sufficient condition for invertibility is to assume that X^*_{t-1} is a convolution of M_{t-2}: $X^*_{t-1} = h_1(w_{t-1}) + h_2(M_{t-2}) + \xi_{t-1}$, with h_2 an increasing function and ξ_{t-1} a random variable with a characteristic function without any real zeros.

References

Abbring, J., P. Chiappori, and T. Zavadil (2008): "Better Safe than Sorry? Ex Ante and Ex Post Moral Hazard in Dynamic Insurance Data," Tilburg Univeristy, working paper.

Ackerberg, D., L. Benkard, S. Berry, and A. Pakes (2007): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics, Vol. 6A*, ed. by J. Heckman, and E. Leamer. North-Holland.

Aguirregabiria, V., and P. Mira (2002): "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70, 1519–1543.

——— (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

Arcidiacono, P., and R. Miller (2006): "CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," Manuscript, Duke University.

Arellano, M. (2003): *Panel Data Econometrics*. Oxford University Press.

Arellano, M., and B. Honore (2000): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics, Vol. 5*. North-Holland.

Bajari, P., L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75, 1331–1370.

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2007): "Nonparametric and Semiparametric Analysis of a Dynamic Game Model," Manuscript, University of Minnesota.

Blundell, R., X. Chen, and D. Kristensen (2007): "Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613–1669.

Bouissou, M., J. J. Laffont, and Q. Vuong (1986): "Tests of Noncausality under Markov Assumptions for Qualitative Panel Data," *Econometrica*, 54, 395–414.

Buchinsky, M., J. Hahn, and J. Hotz (2004): "Estimating Dynamic Discrete Choice Models with Heterogeneous Agents: a Cluster Analysis Approach," Working Paper, UCLA.

Carrasco, M., J.-P. Florens, and E. Renault (2005): "Linear Inverse Problems and Structural Economics: Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics, Vol. 6*. North-Holland.

Carroll, R., X. Chen, and Y. Hu (2008): "Identification and estimation of nonlinear models using two samples with nonclassical measurement error," Manuscript, Johns Hopkins University.

Chen, X., H. Hong, and D. Nekipelov (2007): "Measurement Error Models," manuscript, Yale University.

Collard-Wexler, A. (2006): "Demand Fluctuations and Plant Turnover in the Ready-to-Mix Concrete Industry," manuscript, New York University.

Crawford, G., and M. Shum (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73, 1137–1174.

Cunha, F., J. Heckman, and S. Schennach (2006): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," manuscript,University of Chicago.

Das, S., M. Roberts, and J. Tybout (2007): "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 75, 837–874.

d'Haultfoeuille, X. (2008): "On the Completeness Condition in Nonparametric Instrumental Problems," CREST-INSEE working paper.

Doraszelski, U., and A. Pakes (2007): "A Framework for Dynamic Analysis in IO," in *Handbook of Industrial Organization, Vol. 3*, ed. by M. Armstrong, and R. Porter, chap. 30. North-Holland.

Dunford, N., and J. Schwartz (1971): *Linear Operators*, vol. 3. Wiley.

Dunne, T., S. Klimer, M. Roberts, and D. Xu (2006): "Entry and Exit in Geographic Markets," manuscript, Penn State University.

Erdem, T., S. Imai, and M. Keane (2003): "Brand and Quantity Choice Dynamics under Price Uncertainty," *Quantitative Marketing and Economics*, 1, 5–64.

Fernandez-Villaverde, J., and J. Rubio-Ramirez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," University of Pennsylvania, working paper.

Hendel, I., and A. Nevo (2007): "Measuring the Implications of Sales and Consumer Stockpiling Behavior," Forthcoming, *Econometrica*.

Henry, M., Y. Kitamura, and B. Salanie (2008): "Nonparametric identification of mixtures with exclusion restrictions," University of Montreal, working paper.

Hong, H., and M. Shum (2007): "Pairwise-Difference Estimation of a Dynamic Optimization Model," Revised manuscript, Stanford University.

Hotz, J., and R. Miller (1993): "Conditional Choice Probabilties and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

Hotz, J., R. Miller, S. Sanders, and J. Smith (1994): "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, 61, 265–289.

Houde, J.-F., and S. Imai (2006): "Identification and 2-step Estimation of DDC models with Unobserved Heterogeneity," Working Paper, Queen's University.

Hu, Y. (2008): "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: a General Solution," *Journal of Econometrics*, 144, 27–61.

Hu, Y., and S. Schennach (2008): "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 76, 195–216.

Hu, Y., and M. Shum (2008): "Nonparametric Identification of Dynamic Models with Unobserved State Variables," Jonhs Hopkins University, Dept. of Economics working paper #543.

Imai, S., N. Jain, and A. Ching (2005): "Bayesian Estimation of Dynamic Discrete Choice Models," Queen's University, mimeo.

Kasahara, H., and K. Shimotsu (2007): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choice," Manuscript, University of Western Ontario.

Levinsohn, J., and A. Petrin (2000): "Estimating Production Functions Using Intermediate Inputs to Control for Unobservables," Manuscript, University of Michigan.

Li, T. (2002): "Robust and consistent estimation of nonlinear errors-in-variables models," *Journal of Econometrics*, 110, 1–26.

Magnac, T., and D. Thesmar (2002): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70, 801–816.

Matzkin, R. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1376.

Miller, R. (1984): "Job Matching and Occupational Choice," *Journal of Political Economy*, 92, 1086–1120.

Newey, W., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Norets, A. (2006): "Inference in dynamic discrete choice models with serially correlated unobserved state variables," Manuscript, Princeton University.

Pakes, A. (1986): "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks," *Econometrica*, 54(4), 755–84.

——— (2008): "Theory and Empirical Work in Imperfectly Competitive Markets," Fisher-Schultz Lecture at 2005 Econometric Society World Congress (London, England).

Pakes, A., M. Ostrovsky, and S. Berry (2007): "Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry/Exit Examples)," *RAND Journal of Economics*, 37.

Pesendorfer, M., and P. Schmidt-Dengler (2007): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, forthcoming.

Rust, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55, 999–1033.

——— (1994): "Structural Estimation of Markov Decision Processes," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle, and D. McFadden, pp. 3082–146. North Holland.

Ryan, S. (2006): "The Costs of Environmental Regulation in a Concentrated Industry," manuscript, MIT.

Schennach, S. (2004): "Estimation of nonlinear models with measurement error," *Econometrica*, 72, 33–76.

——— (2007): "Instrumental variable estimation of nonlinear errors-in-variables models," *Econometrica*, 75, 201–239.

Xu, D. (2007): "A Structural Empirical Model of R&D, Firm Heterogeneity, and Industry Evolution," Manuscript, New York University.