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: November 2009

Abstract

We consider the identification of a Markov process $\{W_t, X_t^*\}$ when only $\{W_t\}$ is observed. In structural dynamic models, W_t includes the choice variables and observed state variables of an optimizing agent, while X_t^* denotes time-varying 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 time-varying 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 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

*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 Berkeley-Haas, BU, Clark, UC-Davis, 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, the "Econometrics of Industrial Organization" workshop at Toulouse, and SITE 2009 provided useful comments.

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 time-varying unobserved state variables (USV for short), which are observed by agents and affect their choice of Y_t, but unobserved by the econometrician. The economic importance of models with unobserved state variables has been recognized since the earliest papers on the structural estimation of dynamic optimization models. Two examples are:

[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 time-varying posterior means of workers' beliefs regarding their occupation-specific match values. The observed state variables M_t include job tenure and education level. ■

[2] **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 varies across years and 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. ■

These two early papers demonstrated that dynamic optimization problems with an unobserved process partly determining the state variables are indeed empirically tractable. Their authors (cf. Miller (1984, section V); Pakes and Simpson (1989)) also provided some discussion of the restrictions implied on the data by their models, thus highlighting how identification has been a concern since the earliest structural empirical applications of dynamic models with time-varying unobserved state variables. Obviously, the nonparametric identification of these complex nonlinear models has important practical relevance for empirical researchers, and our goal here is to provide identification results which apply to a broad class of Markovian dynamic models with unobserved state variables.

Our main result 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).[1] A general criticism of these methods is that they cannot accommodate unobserved state variables. In response, Aguirregabiria and Mira (2007), Buchinsky, Hahn, and Hotz (2004), and Houde and Imai (2006), among others, recently developed CCP-based estimation methodologies allowing for agent-specific unobserved heterogeneity, which is the special case where the latent X_t^* is time-invariant. Arcidiacono and Miller (2006) developed a CCP-based approach to estimate dynamic discrete models where X_t^* varies over time according to an exogenous 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 (2009) for nonclassical measurement models (where the measurement error is not assumed to be independent of the latent "true" variable) can be applied to Markovian dynamic models, and we use those results to establish nonparametric identification.

Kasahara and Shimotsu (2009, hereafter KS) consider the identification of dynamic models with discrete unobserved heterogeneity, where the latent variable X_t^* is time-invariant and discrete. KS demonstrate that the Markov law of motion $W_{t+1}|W_t, X^*$ is identified in this setting, using six periods of data. Relative to this, we consider a more general setting

[1]Subsequent methodological developments for CCP-based estimation include Aguirregabiria and Mira (2002), (2007), Pesendorfer and Schmidt-Dengler (2008), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), and Hong and Shum (2009).At the same time, Magnac and Thesmar (2002) and Bajari, Chernozhukov, Hong, and Nekipelov (2007) use the CCP logic to provide identification results for dynamic discrete-choice models.

[2]That is, X_t^* depends stochastically only on X_{t-1}^*, and not on any other variables. We relax this considerably; see the discussion following Assumption 1 below.

[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.

where the unobserved X_t^* is allowed to vary over time (as in the Miller and Pakes examples above), and is 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 are quite distinct from those in HKS, our results share some of HKS's intuition, 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.

2 Overview of assumptions

Consider a dynamic process $\{(W_T, X_T^*), \ldots, (W_t, X_t^*), ..., (W_1, X_1^*)\}_i$ for agent i. We assume that for each agent i, $\{(W_T, X_T^*), \ldots, (W_t, X_t^*), ..., (W_1, X_1^*)\}_i$ is an independent random draw from a bounded continuous distribution $f_{(W_T, X_T^*), \ldots, (W_t, X_t^*), \ldots, (W_1, X_1^*)}$. The researcher observes a panel dataset consisting of an i.i.d. random sample of $\{W_T, W_{T-1}, \ldots, W_1\}_i$, 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. This model subsumes the case of unobserved heterogeneity, in which X_t^* is fixed across all periods.

Next, we introduce our four assumptions. The first assumption below restricts attention

[4] Similarly, 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.

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.

Assumption 1. *(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 up to (but not including)* $t-1$.
(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 W_t. When $W_t = (Y_t, M_t)$, as before, Assumption 1 implies:

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

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 (the CCP) further simplifies to $f_{Y_t|M_t,X_t^*}$, because the Markovian laws of motion for (M_t, X_t^*) imply that the optimal policy function depends just on the current state variables. 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}. 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^*. While 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, it 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.[6] For the special case of unobserved heterogeneity, where

[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 Honoré (2000). Arellano (2003, chs. 7–8) considers linear panel models with lagged dependent variables and serially-correlated unobservables, which is also related to our framework.

[6] Most empirical applications of dynamic optimization models with unobserved state variables satisfy

$X_t^* = X_{t-1}^*$, $\forall t$, the limited feedback assumption is trivial. Finally, the limited feedback assumption places no restrictions on the law of motion for X_t^*, and allows X_t^* to depend stochastically on $X_{t-1}^*, Y_{t-1}, M_{t-1}$. ■

For this paper, we assume that the unobserved state variable X_t^* is scalar-valued, and is drawn from a continuous distribution.[7] An important role in the identification argument is played by many integral equalities which demonstrate the equivalence of multivariate density functions which contain the latent variable X_t^* as an argument (which are not identified directly in the data), and those containing only observed variables W_t (which are identified directly from the data). To avoid cumbersome repetition, we will express these integral equalities in the convenient notation of linear operators, which we introduce here.

Let R_1, R_2, R_3 denote three random variables, with support $\mathcal{R}_1$, $\mathcal{R}_2$, and $\mathcal{R}_3$, distributed with joint density $f_{R_1,R_2,R_3}(r_1, r_2, r_3)$ with support $\mathcal{R}_1 \times \mathcal{R}_2 \times \mathcal{R}_3$.[8] The linear operator L_{R_1,r_2,R_3} is a mapping from the $\mathcal{L}^p$-space of functions of R_3 to the $\mathcal{L}^p$ space of functions of R_1,[9] defined as [10]

$$\left(L_{R_1,r_2,R_3} h\right)(r_1) = \int f_{R_1,R_2,R_3}(r_1, r_2, r_3) h(r_3) dr_3; \quad h \in \mathcal{L}^p\left(\mathcal{R}_3\right),\ r_2 \in \mathcal{R}_2.$$

Similarly, we define the diagonal (or multiplication) operator

$$\left(D_{r_1|r_2,R_3} h\right)(r_3) = f_{R_1|R_2,R_3}(r_1|r_2, r_3) h(r_3); \quad h \in \mathcal{L}^p\left(\mathcal{R}_3\right),\ r_1 \in \mathcal{R}_1,\ r_2 \in \mathcal{R}_2.$$

In the next section, we show that our identification argument relies on 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.)[11] The next two assumptions ensure the validity and uniqueness of this decomposition.

---

the Markov and limited feedback conditions: examples from the industrial organization literature include Erdem, Imai, and Keane (2003), Crawford and Shum (2005), Das, Roberts, and Tybout (2007), Xu (2007), and Hendel and Nevo (2006).

[7] 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.

[8] Here, capital letters denote random variables, while lower-case letters denote realizations.

[9] For $1 \le p < \infty$, $\mathcal{L}^p(\mathcal{X})$ is the space of measurable real functions $h(\cdot)$ integrable in the L^p-norm, ie. $\int_{\mathcal{X}} |h(x)|^p d\mu(x) < \infty$, where μ is a measure on a σ-field in $\mathcal{X}$. One may also consider other classes of functions, such as bounded functions in $\mathcal{L}^1$, in the definition of an operator.

[10] Analogously, the operator $L_{R_1|r_2,R_3}$, corresponding to the conditional density $f_{R_1|R_2,R_3}$, is defined, for all functions $h \in \mathcal{L}^p(\mathcal{R}_3)$, and $r_2 \in \mathcal{R}_2$ as $\left(L_{R_1|r_2,R_3} h\right)(r_1) = \int f_{R_1|R_2,R_3}(r_1|r_2, r_3) h(r_3) dr_3$.

[11] 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.

Assumption 2. Invertibility: *There exists variable(s) $V \subseteq W$ such that*
(i) for any $w_t \in \mathcal{W}_t$, there exists a $w_{t-1} \in \mathcal{W}_{t-1}$ such that for any $(\overline{w}_t, \overline{w}_{t-1})$ in a neighborhood of (w_t, w_{t-1}) [12]*, $L_{V_{t-2},\overline{w}_{t-1},\overline{w}_t,V_{t+1}}$ is one-to-one;*
(ii) for any $w_t \in \mathcal{W}_t$, $L_{V_{t+1}|w_t,X_t^*}$ is one-to-one;
(iii) for any $w_{t-1} \in \mathcal{W}_t$, L_{V_{t-2},w_{t-1},V_t} is one-to-one.

Assumption 2 enables us to take inverses of certain operators, and is analogous to assumptions made in the nonclassical measurement error literature. Specifically, treating V_{t-2} and V_{t+1} as noisy "measurements" of the latent X_t^*, Assumption 2(i,ii) imposes the same restrictions between the measurements and the latent variable as Hu and Schennach (2008, Assumption 3) and Carroll, Chen, and Hu (2009, Assumption 2.4). Compared with these two papers, Assumption 2(iii) is an extra assumption we need because, in our dynamic setting, there is a second latent variable, X_{t-1}^*, in the Markov law of motion $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$. Below, we show that Assumption 2(ii) implies that pre-multiplication by the inverse operator $L^{-1}_{V_{t+1}|w_t,X_t^*}$ is valid, while 2(i,iii) imply that post-multiplication by, respectively, $L^{-1}_{V_{t+1},w_t,w_{t-1},V_{t-2}}$ and $L^{-1}_{V_t,w_{t-1},V_{t-2}}$ is valid.[13]

The statements in Assumption 2 are equivalent to *completeness* conditions which have recently been employed in the nonparametric IV literature: namely, 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 "completeness" condition: for any r_2,

$$\left(L_{R_1,r_2,R_3}h\right)(r_1) = \int f(r_1, r_2, r_3)h(r_3)dr_3 = 0 \text{ for all } r_1 \text{ implies } h(r_3) = 0 \text{ for all } r_3. \quad (2)$$

Completeness is a high-level condition, and special cases of it have been considered in, eg. Newey and Powell (2003), Blundell, Chen, and Kristensen (2007), d'Haultfoeuille (2009). However, sufficient conditions are not available for more general settings. Below, in Section 4, we will construct examples which satisfy the completeness requirements.

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.[14]

Assumption 2(ii) rules out models where X_t^* has a continuous support, but W_{t+1} contains

[12] A neighborhood of $w \in \mathbb{R}^k$ is defined as $\left\{\overline{w} \in \mathbb{R}^k : \|\overline{w} - w\|_E < r\right\}$ for some $r > 0$, where $\|\cdot\|_E$ is the Euclidean metric.

[13] Additional details are given in Section 2 of the online appendix (Hu and Shum (2009)).

[14] 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.

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: *For any $w_t \in \mathcal{W}_t$ and any $\overline{x}_t^* \neq \widetilde{x}_t^* \in \mathcal{X}_t^*$, there exists a $w_{t-1} \in \mathcal{W}_{t-1}$ satisfying Assumption 2(i) and there exists $(\overline{w}_t, \overline{w}_{t-1})$ in a neighborhood of (w_t, w_{t-1}) such that the density $f_{W_t|W_{t-1},X_t^*}$ satisfies :*
(i) $0 < k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*) < \infty$ for any $x_t^* \in \mathcal{X}_t^*$;
(ii) $k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, \overline{x}_t^*) \neq k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, \widetilde{x}_t^*)$,
where $k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*) = \frac{f_{w_t|w_{t-1},x_t^*} f_{\overline{w}_t|\overline{w}_{t-1},x_t^*}}{f_{\overline{w}_t|w_{t-1},x_t^*} f_{w_t|\overline{w}_{t-1},x_t^*}}$ with $f_{w_t|w_{t-1},x_t^*} = f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*)$.

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 Eq. (12) below shows, the eigenvalues in this decomposition involve the density $f_{W_t|W_{t-1},X_t^*}$, and conditions (i) and (ii) are restrictions on this density which guarantee that these eigenvalues are, respectively, bounded and distinct across all values of x_t^*.In turn, this ensures that the corresponding eigenfunctions are linearly independent, so that the spectral decomposition is unique.[15] ■

Assumption 4. Monotonicity and normalization: *For any $w_t \in \mathcal{W}_t$, 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^*. We normalize $x_t^* = G\left[f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)\right]$.*

The eigenfunctions in the aforementioned spectral decomposition correspond to the densities $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$, for all values of 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 restrictions deriving from the economic or stochastic structure 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 $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ is monotonic in x_t^*, given w_t. Given the monotonicity restriction, we can normalize X_t^* by setting, $x_t^* = G\left[f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)\right]$

[15] In the case where $W_t = (Y_t, M_t)$ and $f_{W_t|W_{t-1},X_t^*} = f_{Y_t|M_t,X_t^*} \cdot f_{M_t|Y_{t-1},M_{t-1},X_t^*}$, Assumption 3 simplifies further. Specifically, because the CCP term $f_{Y_t|M_t,X_t^*}$ does not contain W_{t-1}, Eq. (12) below implies that the CCP term cancels out in the expression of eigenvalues in the spectral decomposition, so that Assumption 3 imposes restrictions only on the second term $f_{M_t|Y_{t-1},M_{t-1},X_t^*}$. See additional discussion in Example 2 below.

without loss of generality.[16] 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,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 $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$,*

$$L_{V_{t+1},w_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*,w_{t-1},V_{t-2}}. \tag{3}$$

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

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

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 $t \in \{3,\ldots,T-1\}$, Assumptions 1, 2, 3, 4 imply that the density $f_{V_{t+1},W_t,W_{t-1},V_{t-2}}$ uniquely determines 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

[16] To be clear, the monotonicity assumption here is a model restriction, and not without loss of generality; if it were false, our identification argument would not recover the correct CCP's and laws of motion for the underlying model. See Matzkin (2003) and Hu and Schennach (2008) for similar uses of monotonicity restrictions in the context of nonparametric identification problems.

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)

For each w_t, choose a w_{t-1} to satisfy Assumptions 2(i) and 3(i), and pick $(\bar{w}_t, \bar{w}_{t-1})$ within a neighborhood of (w_t, w_{t-1}), such that $w_t \neq \bar{w}_t$ and $w_{t-1} \neq \bar{w}_{t-1}$. 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. (3) at the four pairs of points (w_t, w_{t-1}), $(\bar{w}_t, w_{t-1})$, $(w_t, \bar{w}_{t-1})$, $(\bar{w}_t, \bar{w}_{t-1})$, each pair of equations will share one operator in common. Specifically:

$$\begin{aligned}
\text{for } (w_t, w_{t-1}) \quad &: \quad L_{V_{t+1},w_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} L_{X^*_t,w_{t-1},V_{t-2}}, && (5)\\
\text{for } (\overline{w}_t, w_{t-1}) \quad &: \quad L_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|\overline{w}_t,X^*_t} D_{\overline{w}_t|w_{t-1},X^*_t} L_{X^*_t,w_{t-1},V_{t-2}}, && (6)\\
\text{for } (w_t, \overline{w}_{t-1}) \quad &: \quad L_{V_{t+1},w_t,\overline{w}_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X^*_t} D_{w_t|\overline{w}_{t-1},X^*_t} L_{X^*_t,\overline{w}_{t-1},V_{t-2}}, && (7)\\
\text{for } (\overline{w}_t, \overline{w}_{t-1}) \quad &: \quad L_{V_{t+1},\overline{w}_t,\overline{w}_{t-1},V_{t-2}} = L_{V_{t+1}|\overline{w}_t,X^*_t} D_{\overline{w}_t|\overline{w}_{t-1},X^*_t} L_{X^*_t,\overline{w}_{t-1},V_{t-2}}. && (8)
\end{aligned}$$

Assumptions 2(ii) and 3(i) imply that we can solve for $L_{X^*_t,w_{t-1},V_{t-2}}$ from Eq. (6) as

$$D^{-1}_{\overline{w}_t|w_{t-1},X^*_t} L^{-1}_{V_{t+1}|\overline{w}_t,X^*_t} L_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}} = L_{X^*_t,w_{t-1},V_{t-2}}.$$

Plugging in this expression to Eq. (5) leads to

$$L_{V_{t+1},w_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} D^{-1}_{\overline{w}_t|w_{t-1},X^*_t} L^{-1}_{V_{t+1}|\overline{w}_t,X^*_t} L_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}}.$$

Lemma 1 of Hu and Schennach (2008) shows that, given Assumption 2(i), we can postmultiply by $L^{-1}_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}}$, to obtain:

$$\begin{aligned}
\mathbf{A} \quad &\equiv \quad L_{V_{t+1},w_t,w_{t-1},V_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}}\\
&= \quad L_{V_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} D^{-1}_{\overline{w}_t|w_{t-1},X^*_t} L^{-1}_{V_{t+1}|\overline{w}_t,X^*_t}. && (9)
\end{aligned}$$

Similar manipulations of Eqs. (7) and Eq. (8) lead to

$$\begin{aligned}
\mathbf{B} \quad &\equiv \quad L_{V_{t+1},\overline{w}_t,\overline{w}_{t-1},V_{t-2}} L^{-1}_{V_{t+1},w_t,\overline{w}_{t-1},V_{t-2}}\\
&= \quad L_{V_{t+1}|\overline{w}_t,X^*_t} D_{\overline{w}_t|\overline{w}_{t-1},X^*_t} D^{-1}_{w_t|\overline{w}_{t-1},X^*_t} L^{-1}_{V_{t+1}|w_t,X^*_t}. && (10)
\end{aligned}$$

Assumption 2(i) guarantees that such $\overline{w}_t$, w_{t-1}, and $\overline{w}_{t-1}$ always exist for any w_t. Finally, we postmultiply Eq. (9) by Eq. (10) to obtain

$$\begin{aligned}
\mathbf{AB} &= L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} \left(L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*} L_{V_{t+1}|\overline{w}_t,X_t^*} \right) \times \\
&\quad \times D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|w_t,X_t^*} \\
&= L_{V_{t+1}|w_t,X_t^*} \left(D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*} \right) L^{-1}_{V_{t+1}|w_t,X_t^*} \\
&\equiv L_{V_{t+1}|w_t,X_t^*} D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|w_t,X_t^*}, \qquad \text{where} \qquad (11)
\end{aligned}$$

$$\begin{aligned}
\left(D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} h \right)(x_t^*) &= \left(D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*} h \right)(x_t^*) \\
&= \frac{f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1},x_t^*) f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|\overline{w}_{t-1},x_t^*)}{f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|w_{t-1},x_t^*) f_{W_t|W_{t-1},X_t^*}(w_t|\overline{w}_{t-1},x_t^*)} h(x_t^*) \qquad (12) \\
&\equiv k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},x_t^*\right) h\left(x_t^*\right).
\end{aligned}$$

This equation implies that the observed operator **AB** on the left hand side of Eq. (11) 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. (11) is similar to the decomposition in Hu and Schennach (2008) or Carroll, Chen, and Hu (2009).

Assumption 3 ensures that this decomposition is unique. Specifically, Eq. (11) implies that the operator **AB** 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 **AB** 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. (11) itself does not imply that the eigenvalues $k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},x_t^*\right)$ 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. 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 distinct across all x_t^*.[17] When w_t (resp. w_{t-1}) is close to $\bar{w}_t$ (resp. $\bar{w}_{t-1}$), Eq. (12)

[17] Specifically, the operators **AB** corresponding 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

implies that the logarithm of the eigenvalues in this decomposition can be represented as a second-order derivative of the log-density $f_{W_t|W_{t-1},X_t^*}$. Therefore, a sufficient condition for 3(ii) 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}).

Second, the eigenfunctions $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ in the spectral decomposition (11) 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 our arguments are nonparametric, and X_t^* is unobserved, we need an additional monotonicity condition, in Assumption 4, to pin down the value of X_t^* relative of the observed variables. This was discussed earlier, in the remarks following Assumption 4.

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. (11). *Q.E.D.*

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.[18] Hence, by Lemma 2, we have shown the following result:

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 $t \in \{4, \dots, T-1\}$*

$f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ 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., $k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*) \neq k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, \widetilde{x}_t^*)$.

[18] 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.

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: *the Markov law of motion of* (W_t, X^*_t) *is time-invariant:* $f_{W_t,X^*_t|W_{t-1},X^*_{t-1}} = f_{W_2,X^*_2|W_1,X^*_1}$, $\forall\ 2 \le t \le T$.

Stationarity is usually maintained in infinite-horizon dynamic programming 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.[19]

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

[19] 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.

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 direct calculation, as in the examples here. 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, we present two examples of dynamic models here. 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.[20]

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 binary scalars: $\forall t,\ \text{supp}X_t^* = \text{supp}W_t \equiv \{0,1\}$. This is the simplest example of the models considered in our framework. One example of such a 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. In this model, moral hazard in driving behavior and experience rating in insurance pricing imply that the laws of motion for both W_t and X_t^* should exhibit state dependence:

$$\Pr(W_t = 1|w_{t-1}, x_t^*, x_{t-1}^*) = p(w_{t-1}, x_t^*); \quad \Pr(X_t^* = 1|x_{t-1}^*, w_{t-1}) = q(x_{t-1}^*, w_{t-1}). \tag{13}$$

These laws of motion satisfy Assumption 1. Previously, KS also analyzed the identification of dynamic discrete models with unobserved variables, but they only considered models where the unobserved variables X^* were time-invariant. In contrast, even in the simple example here, we allow X_t^* to vary over time, so that this model falls outside KS's framework.[21]

Relative to the continuous case presented beforehand, some simplifications obtain in this finite-dimensional example. Notationally, the linear operators in the previous section reduce to matrices, with the L operators in the main proof corresponding to $2{\times}2$ square matrices, and the D operators are $2{\times}2$ diagonal matrices. Specifically, for binary random

[20] A third example, based on Rust (1987), is in the supplemental material (Hu and Shum (2009)).

[21] In section 3 of the online appendix, we provide a more detailed discussion.

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 $(R_1 = i, R_2 = r_2, R_3 = j)$, for $i, j \in \{0, 1\}$.

Assumptions 2, 3, and 4 are quite transparent to interpret in the matrix setting. Assumption 2 implies the invertibility of certain matrices. As shown in the proof of Lemma 3, our identification results require there exist at least four different points in the support of (W_t, W_{t-1}). In this simple dichotomous case, that means Assumptions 2(i) and 3(i) hold for all the possible values of (w_t, w_{t-1}). From Lemma 1, the following matrix equality holds, for all values of (w_t, w_{t-1}):

$$L_{W_{t+1},w_t,w_{t-1},W_{t-2}} = L_{W_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*,w_{t-1},W_{t-2}}. \tag{14}$$

Assumption 2(i) implies that the square matrix $L_{W_{t-2},w_{t-1},w_t,W_{t+1}} = L^T_{W_{t+1},w_t,w_{t-1},W_{t-2}}$ is invertible, which implies that $L_{W_{t+1},w_t,w_{t-1},W_{t-2}}$ is also invertible. Hence, by Eq. (14), $L_{W_{t+1}|w_t,X_t^*}$ 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, Assumption 2(ii) is redundant, because it is implied by 2(i).

Furthermore, Assumption 2(iii) is also implied by 2(i). Specifically, $L_{W_{t-2},w_{t-1},W_t} = L^T_{W_t,w_{t-1},W_{t-2}}$ with $L_{W_t,w_{t-1},W_{t-2}} = L_{W_t|w_{t-1},X^*_{t-1}} L_{X^*_{t-1},w_{t-1},W_{t-2}}$. By Assumption 2(i), $L_{W_t|w_{t-1},X^*_{t-1}}$ is invertible. Since $L_{X_t^*,w_{t-1},W_{t-2}}$ was shown above to be invertible, the matrix $L_{W_t,w_{t-1},W_{t-2}}$ is invertible, and hence so is L_{W_{t-2},w_{t-1},W_t}.

Assumption 3 puts restrictions on the eigenvalues in the spectral decomposition of the **AB** operator. In the discrete case, **AB** 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 (w_t, w_{t-1}), these assumptions can be verified, by directly diagonalizing the **AB** matrix.

In this discrete case, Assumption 4 is to an "ordering" assumption on the columns of the $L_{W_{t+1}|w_t,X_t^*}$ matrix, which are the eigenvectors of **AB**. This is because, for a matrix diagonalization $T = SDS^{-1}$, where D is diagonal, and T and S are square matrices, any permutation of the eigenvalues (the diagonal elements in D) and their corresponding eigenvectors (the columns in S) results in the same diagonal representation of T.

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 orderings in recovering $L_{W_{t+1}|w_t,X_t^*}$ and $L_{W_t|w_{t-1},X^*_{t-1}}$. By doing this, we cannot pin down the exact value of X_t^* or X^*_{t-1}, but the recovered density of $W_t, X_t^*|W_{t-1}, X^*_{t-1}$ is still 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

$w_{t-1} \in \mathcal{W}_{t-1}$ is consistent with the true, but unknown ordering of X_t^* and X_{t-1}^*.[22]

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. For this example, one reasonable monotonicity restriction is

$$\text{for } w_t = \{0,1\}: \quad \mathbb{E}[W_{t+1}|w_t, X_t^* = 1] < \mathbb{E}[W_{t+1}|w_t, X_t^* = 0] \tag{15}$$

The restriction (15) implies that future claims W_{t+1} occur less frequently with higher effort today, and imposes additional restrictions on the the $p(\cdots)$ and $q(\cdots)$ functions in (13).[23]

To see how this restriction orders the eigenvectors, and pins down the value of X_t^*, note that $\mathbb{E}[W_{t+1}|w_t, X_t^*] = f(W_{t+1} = 1|w_t, X_t^*)$, which is the second component of each eigenvector. Therefore, the monotonicity restriction (15) implies that the eigenvectors (and their corresponding eigenvalues) should be ordered such that their second components are decreasing, from left to right. Given this ordering, we assign a value of $X_t^* = 0$ to the eigenvector in the first column, and $X_t^* = 1$ to the eigenvector in the second column.

4.2 Example 2: generalized investment model

For the second 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).[24] In this model, $W_t = (Y_t, M_t)$, where Y_t is a firm's R&D in year t, and M_t is the product's installed base. The unobserved state variable X_t^* is the firm's product quality.

Product quality $X_t^* \in \mathbb{R}$ evolves as follows:

$$X_t^* = 0.8X_{t-1}^* + 0.1\psi\left(Y_{t-1}\right) + 0.1\nu_t;. \tag{16}$$

In the above, $\nu_t \in \mathbb{R}$ is a standard normal shock, distributed independently over t, and $\psi'(\cdot) > 0$.

Installed base evolves as:

$$M_{t+1} - M_t = M_t \exp(\eta_{t+1} + X_{t+1}^*) \tag{17}$$

where $\eta_{t+1} \in \mathbb{R}$ is a random shock following the extreme value distribution $f_{\eta_{t+1}}(\eta) =$

[22] We thank Thierry Magnac for this insight.

[23] See Hu (2008) for a number of other alternative ordering assumptions for the discrete case.

[24] See Hu and Shum (2009, Section 1.2) for additional discussion of dynamic investment models.

$\exp(\eta - e^{\eta})$ for $\eta \in \mathbb{R}$, independently across t.[25] Eq. (17) implies that, *ceteris paribus*, product quality raises installed base.

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

$$\begin{aligned} Y_t &= Y^*(M_t, X_t^*, \gamma_t) \\ &= \text{argmax}_{0 \le y \le \bar{I}} \big[\underbrace{\Pi(M_t, X_t^*)}_{\text{profits}} - \underbrace{\gamma_t}_{\text{shock}} \cdot \underbrace{Y_t^2}_{\text{R\&D cost}} + \beta \mathbb{E} \underbrace{V(M_{t+1}, X_{t+1}^*, \gamma_{t+1})}_{\text{value fxn}} \big] \end{aligned} \tag{18}$$

$\bar{I}$ is a cap on per-period R&D, and γ_t is a shock to R&D costs. We assume that $\gamma_t \in (0, +\infty)$ follows a standard exponential distribution independently across t. Therefore, the RHS of Eq. (18) is supermodular in Y_t and $-\gamma_t$, for all (M_t, X_t^*). Accordingly, for fixed (M_t, X_t^*), the firm's optimal R&D investment Y_t^* is monotonically decreasing in γ_t, and take values in $(0, \bar{I}]$.

We verify the assumptions out of order, leaving the most involved Assumption 2 to the end. Since we focus here on the stationary case, without loss of generality we label the four observed periods of data W_t as $t = 1, 2, 3, 4$.

Assumption 1 is satisfied for this model. **Assumption 3** contains two restrictions on the density $f_{W_3|W_2,X_3^*}$, which factors as

$$f_{W_3|W_2,X_3^*} = f_{Y_3|M_3,X_3^*} \cdot f_{M_3|M_2,X_3^*}. \tag{19}$$

The first term in Eq. (19) is the density of R&D Y_3. Because the first term is not a function of M_2, Eq. (12) implies that the investment density $f_{Y_3|M_3,X_3^*}$ cancels out from the numerator and denominator of the eigenvalues in the spectral decomposition. Hence, to ensure that the eigenvalues are distinct, we only require $f_{Y_3|M_3,X_3^*} > 0$ for all X_3^*. Given the discussions above, conditional on (M_3, X_3^*), investment Y_3 will be monotonically decreasing in the shock γ_3. Since, by assumption, the density of γ_3 is nonzero for $\gamma_t > 0$, so will the conditional density $f_{Y_3|M_3,X_3^*} > 0$ along its support $(0, \bar{I}]$, for all (M_3, X_3^*), as required.

The second term $f_{M_3|M_2,X_3^*}$ is the law of motion for installed base which, by assumption, is an extreme value distribution

$$f_{M_3|M_2,X_3^*}(m_3|m_2, x_3^*) = \frac{1}{(m_3 - m_2)} \exp\left[\varphi(m_3, m_2) - x_3^* - e^{\varphi(m_3,m_2) - x_3^*}\right].$$

[25] Furthermore, the characteristic function of $f_{\eta_{t+1}}$ is $\phi_{\eta_{t+1}}(\tau) = \Gamma(1 + i\tau)$, which is nonzero for any $\tau \in \mathbb{R}$.

where $\varphi(m_3, m_2) = \ln \frac{m_3 - m_2}{m_2}$. We then have

$$k\left(w_3, \overline{w}_3, w_2, \overline{w}_2, x_3^*\right) = \exp\left(-e^{-x_3^*}\left[\frac{-\left(\overline{m}_3 - m_3\right)\left(\overline{m}_2 - m_2\right)}{m_2 \overline{m}_2}\right]\right)$$

In order to make all the eigenvalues bounded, we may take $(\overline{m}_3, \overline{m}_2) = (m_3 - \Delta m, m_2 + \Delta m)$ in the neighborhood of (m_3, m_2) so that the RHS equals $\exp\left(-e^{-x_3^*}\frac{(\Delta m)^2}{m_2 \overline{m}_2}\right) \in (0, \infty)$. That means all the eigenvalues are nonzero and bounded, which implies Assumption 3(i) holds. In other words, for any m_3, we may always find $(m_2, \overline{m}_3, \overline{m}_2)$ such that $k\left(w_3, \overline{w}_3, w_2, \overline{w}_2, x_3^*\right)$ is nonzero and finite for any x_3^* on the real line. Moreover, the eigenvalues $k\left(w_3, \overline{w}_3, w_2, \overline{w}_2, x_3^*\right)$ are monotonic in x_3^* for any given $(w_3, \overline{w}_3, w_2, \overline{w}_2)$, which implies Assumption 3(ii) holds.

For **Assumption 4**, note $\mathbb{E}[\ln \frac{M_4 - m_3}{m_3} | m_3, y_3, x_3^*] = \mathbb{E}[\eta_4] + \mathbb{E}[X_4^* | x_3^*, y_3]$. Because the law of motion for product quality implies that $\mathbb{E}[X_4^* | x_3^*, y_3]$ is monotonic in x_3^*, we may take G to be $x_3^* = \mathbb{E}[\ln \frac{M_4 - m_3}{m_3} | m_3, y_3, x_3^*]$.

Finally, **Assumption 2** contains three injectivity assumptions. For the V_t variables in Assumption 2, we use $V_t = M_t$, for all periods t.[26]

Here, we provide sufficient conditions for Assumption 2, in the context of this investment model. We exploit the fact that the laws of motion for this model (cf. Eqs. (16) and (17)) are either linear or log-linear to apply results from the convolution literature, for which operator invertibility has been studied in detail.

For Assumption 2, we need to establish the injectivity of the operators L_{M_1, w_2, w_3, M_4}, $L_{M_4|w_3, X_3^*}$, and L_{M_1, w_2, M_3}. We start by showing the injectivity of L_{M_4, w_3, w_2, M_1}, $L_{M_4|w_3, X_3^*}$, and L_{M_3, w_2, M_1}. As shown in the proof of Lemma 2, Assumption 1 implies that

$$\begin{aligned}
L_{M_4, w_3, w_2, M_1} &= L_{M_4|w_3, X_3^*} D_{w_3|w_2, X_3^*} L_{X_3^*, w_2, M_1} \\
&= L_{M_4|w_3, X_3^*} D_{w_3|w_2, X_3^*} L_{X_3^*|w_2, X_2^*} L_{X_2^*, w_2, M_1} \qquad (20) \\
L_{M_3, w_2, M_1} &= L_{M_3|w_2, X_2^*} L_{X_2^*, w_2, M_1}. \qquad (21)
\end{aligned}$$

Furthermore, we have $L_{M_4|w_3, X_3^*} = L_{M_4|w_3, X_4^*} L_{X_4^*|w_3, X_3^*}$.

Hence, the injectivity of L_{M_4, w_3, w_2, M_1}, $L_{M_4|w_3, X_3^*}$, and L_{M_3, w_2, M_1} is implied by the injectivity of $L_{M_4|w_3, X_4^*}$, $D_{w_3|w_2, X_3^*}$, $L_{X_3^*|w_2, X_2^*}$ and $L_{X_2^*, w_2, M_1}$.[27] It turns out that assumptions we have made already for this example ensure that three of these operators are injective.

[26] Levinsohn and Petrin (2003) and Ackerberg, Benkard, Berry, and Pakes (2007) note that, with fixed costs to R&D, $Y_t = 0$ for many values of (M_t, X_t^*), and hence may not provide enough information on X_t^*.

[27] By stationarity, the operators $L_{M_4|w_3, X_3^*}$ and $L_{M_3|w_2, X_2^*}$ are the same, and do not need to be considered separately. Note that our notion of stationarity here is distinct from the notion of covariance-stationarity for stochastic processes. Indeed, as defined in Eq. (17), the M_t process may not be covariance-stationarity, but the law of motion $f_{M_4|w_3, X_3^*}$ is still time-invariant.

We discuss each case in turn.

(i) For the diagonal operator $D_{w_3|w_2,X_3^*}$, its a kernel function is $f_{w_3|w_2,X_3^*} = f_{y_3|m_3,X_3^*} f_{m_3|m_2,X_3^*}$. In the discussion on Assumption 3(i) above, we have shown that $f_{y_3|m_3,X_3^*}$ is nonzero along its support and that $f_{m_3|m_2,X_3^*}$ is nonzero for any (m_3, m_2, x_3^*) in the support. Therefore, The operator $D_{w_3|w_2,X_3^*}$ is injective. Moreover, this argument holds for any $(\overline{w}_t, \overline{w}_{t-1})$ in a neighborhood of (w_t, w_{t-1}).

(ii) For $L_{M_4|w_3,X_4^*}$, we use Eq. (17) whereby, for every (y_3, m_3), M_4 is a convolution of X_4^*, ie. $\ln[M_4 - M_3] - \ln M_3 = X_4^* + \eta_4$. As is well-known, as long as the characteristic function of η_4 has no real zeros, which is satisfied by the assumed extreme value distribution, the corresponding operator is injective, and therefore, invertible.

(iii) Similarly, for fixed w_2, X_3^* is a convolution of X_2^*, ie. $X_3^* = 0.8X_2^* + 0.1\psi(Y_2) + 0.1\nu_3$ (cf. Eq. (16). Hence, $L_{X_3^*|w_2,X_2^*}$ is injective if the characteristic function of ν_3 has no real zeros, which is satisfied by the assumed normal distribution.

(iv) For the last operator, corresponding to the density $f_{X_2^*,w_2,M_1}$, the model assumptions do not allow us to establish injective directly. This is because this joint density confounds both the structural components (laws of motions) in the model and the initial condition $f_{X_1^*,M_1}$. Thus in general, injectivity of this operator is not verifiable only based on the assumptions made thus far about the laws of motion for the state variables.

However, in the special case where product quality X_t^* evolves exogenously – that is, $\psi(\cdot) = 0$ in Eq. (16) – it turns out that an additional assumption on the initial values of the state variables (X_1^*, M_1) suffices to ensure injectivity of the operator $L_{X_2^*,w_2,M_1}$:

Claim 1: If $\psi(\cdot) = 0$ in Eq. (16), and the initial values of the state variables (X_1^*, M_1) are independently distributed, the operator $L_{X_2^*,w_2,M_1}$ is injective.

Proof: in Appendix B

Up to this point, we have shown that the injectivity of L_{M_4,w_3,w_2,M_1}, $L_{M_4|w_3,X_3^*}$, and L_{M_1,w_2,M_3}. It turns out that this implies injectivity of L_{M_1,w_2,w_3,M_4} and L_{M_1,w_2,M_3}, as required by Assumption 2:

Claim 2: L_{M_1,w_2,w_3,M_4} and L_{M_1,w_2,M_3} are injective.

Proof: in Appendix B

The assumptions underlying Claim 1, particularly the assumption that X_t^* evolves exogenously, are doubtlessly strong. However, we stress here that these are sufficient conditions, and are not necessary for the general results. Moreover, a large class of investment models (eg. Olley and Pakes (1996), Levinsohn and Petrin (2003)) assume that the unobserved variable X_t^* (denoting productivity) evolves exogenously. Finally, these assumptions are needed only in this example because we assume X_t^* to be continuous-valued. As Example 1 above demonstrates, when X_t^* is discrete, we can verify the identification assumptions even when the evolution of X_t^* depends on past values of the observed variables w_{t-1}.

5 Concluding remarks

We have considered the identification of a first-order Markov process $\{W_t, X_t^*\}$ when only $\{W_t\}$ 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 stationarity, identification of $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ obtains with only four observations $W_{t+1},\ldots,W_{t-2}$. Once $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is identified, nonparametric identification of the remaining parts of the models – particularly, the per-period utility functions – can proceed by applying the results in Magnac and Thesmar (2002) and Bajari, Chernozhukov, Hong, and Nekipelov (2007), who considered dynamic models without unobserved state variables X_t^*.

For a general k-th order Markov process ($k < \infty$), it can be shown that the $3k+2$ observations $W_{t+k},\ldots,W_{t-2k-1}$ can identify the 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 in this paper.

We have only considered the case where the unobserved state variable X_t^* is scalar-valued. The case where X_t^* is a multivariate process, which may apply to dynamic game settings, presents some serious challenges. Specifically, 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) discuss the difficulties with multivariate unobserved state variables in the context of dynamic investment models.

Finally, this paper has focused on identification, but not estimation. In ongoing work, we are using our identification results to guide the estimation of dynamic models with unobserved state variables. This would complement recent papers on the estimation of parametric dynamic models with unobserved state variables, using non-CCP-based approaches.[28]

[28]Imai, Jain, and Ching (2009) and Norets (2006) consider Bayesian estimation, and Fernandez-Villaverde and Rubio-Ramirez (2007) consider efficient simulation estimation based on particle filtering.

APPENDIX 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

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

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

$$
\begin{aligned}
f_{W_{t+1},W_t,W_{t-1},W_{t-2}} &= \int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*} \left(\int f_{X_t^*,X_{t-1}^*,W_{t-1},W_{t-2}} dx_{t-1}^* \right) dx_t^* \\
&= \int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*} f_{X_t^*,W_{t-1},W_{t-2}} dx_t^*. \qquad (22)
\end{aligned}
$$

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

$$
L_{W_{t+1},w_t,w_{t-1},W_{t-2}} = L_{W_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*,w_{t-1},W_{t-2}}. \qquad (23)
$$

For the variable(s) $V_t \subseteq W_t$, for all periods t, introduced in Assumption 2, Eq. (23) implies that the joint density of $\{V_{t+1}, W_t, W_{t-1}, V_{t-2}\}$ is expressed in operator notation as $L_{V_{t+1},w_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*,w_{t-1},V_{t-2}}$, as postulated by Lemma 1. *Q.E.D.*

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

$$
\begin{aligned}
f_{V_{t+1},W_t,W_{t-1},V_{t-2}} &= \int f_{V_{t+1}|W_t,X_t^*} f_{W_t,X_t^*,W_{t-1},V_{t-2}} dx_t^* \\
f_{W_t,X_t^*,W_{t-1},V_{t-2}} &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*,W_{t-1},V_{t-2}} dx_{t-1}^*. \qquad (24)
\end{aligned}
$$

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

$$\begin{aligned} L_{V_{t+1},w_t,w_{t-1},V_{t-2}} &= L_{V_{t+1}|w_t,X_t^*} L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*,w_{t-1},V_{t-2}} \\ \Leftrightarrow L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*,w_{t-1},V_{t-2}} &= L^{-1}_{V_{t+1}|w_t,X_t^*} L_{V_{t+1},w_t,w_{t-1},V_{t-2}}. \end{aligned} \tag{25}$$

where the second line uses Assumption 2(ii). Next, we eliminate $L_{X_{t-1}^*,w_{t-1},V_{t-2}}$ from the above. Again using Assumption 1, we have

$$f_{V_t,W_{t-1},V_{t-2}} = \int f_{V_t|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*,W_{t-1},V_{t-2}} dx_{t-1}^* \tag{26}$$

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

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

where the right-hand side applies Assumption 2(ii). Hence, substituting the above into Eq. (25), we obtain the desired representation

$$\begin{aligned} & L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L^{-1}_{V_t|w_{t-1},X_{t-1}^*} L_{V_t,w_{t-1},V_{t-2}} = L^{-1}_{V_{t+1}|w_t,X_t^*} L_{V_{t+1},w_t,w_{t-1},V_{t-2}} \\ \Rightarrow \quad & L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L^{-1}_{V_t|w_{t-1},X_{t-1}^*} = L^{-1}_{V_{t+1}|w_t,X_t^*} L_{V_{t+1},w_t,w_{t-1},V_{t-2}} L^{-1}_{V_t,w_{t-1},V_{t-2}} \\ \Rightarrow \quad & L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} = L^{-1}_{V_{t+1}|w_t,X_t^*} L_{V_{t+1},w_t,w_{t-1},V_{t-2}} L^{-1}_{V_t,w_{t-1},V_{t-2}} L_{V_t|w_{t-1},X_{t-1}^*}. \end{aligned} \tag{27}$$

The second line applies Assumption 2(iii) to postmultiply by $L^{-1}_{V_t,w_{t-1},V_{t-2}}$, while in the third line, we postmultiply both sides by $L_{V_t|w_{t-1},X_{t-1}^*}$. *Q.E.D.*

Proof: (Corollary 1)

From Lemma 3, $f_{V_t|W_{t-1},X_{t-1}^*}$ is identified from density $f_{V_t,W_{t-1},W_{t-2},V_{t-3}}$. The equality $f_{V_t,W_{t-1}} = \int f_{V_t|W_{t-1},X_{t-1}^*} f_{W_{t-1},X_{t-1}^*} dx_{t-1}^*$ implies that, for any $w_{t-1} \in \mathcal{W}_t$,

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

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

Proof: (Corollary 3)

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). Because $f_{V_{t-1},W_{t-2}} = \int f_{V_{t-1}|W_{t-2},X_{t-2}^*} f_{W_{t-2},X_{t-2}^*} dx_{t-2}^*$, the same argument as in the proof of Corollary 1 then implies that f_{W_{t-2},X_{t-2}^*} is identified from the observed density $f_{V_{t-1},W_{t-2}}$. *Q.E.D.*

6 APPENDIX B: Proofs of Claims for Example 2

Here we provide the proofs for Claims 1 and 2 in example 2. We start with a general lemma regarding integral operators based on a convolution form. We consider the basic convolution case where $X = Z + \epsilon$ with $Z \in \mathbb{R}$, $\epsilon \in \mathbb{R}$, and $Z \perp \epsilon$. The independence between Z and ϵ implies that $f_{X|Z}(x|z) = f_\epsilon(x-z)$. Given $\left(L_{X|Z}h\right)(x) = \int f_\epsilon(x-z)\, h(z)\, dz$, we define correspondingly

$$\left(L_{X|Z}^* h\right)(z) = \int f_\epsilon(x-z)\, h(x)\, dx.$$

Notice that $L_{X|Z}^*$ maps functions of X to those of Z. The following lemma is useful for what follows.

Lemma 4. *Suppose that i) the kernel of operator $L_{X|Z}$ is $f_\epsilon(x-z)$; ii) the Fourier transform of f_ϵ does not vanish on the real line. Then, operators $L_{X|Z}$ and $L_{X|Z}^*$ are injective.*

Proof: (see the online appendix)

Proof of Claim 1: The operator $L_{X_2^*,w_2,M_1}$ has kernel function

$$\begin{aligned} f_{X_2^*,w_2,M_1} &= f_{X_2^*,y_2,m_2,M_1} \\ &= f_{y_2|m_2,X_2^*} f_{m_2|X_2^*,M_1} \int f_{X_2^*|X_1^*} f_{X_1^*,M_1} dX_1^*, \end{aligned}$$

Note that in the second line, we have utilized the "special case" restriction that $\psi(\cdot) = 0$ in Eq. (16) so that the density of $f_{X_2^*,X_1^*,M_1}$ can be factored as $f_{X_2^*|X_1^*} \cdot f_{X_1^*,M_1}$. This equation shows that the property of the operator $L_{X_2^*,w_2,M_1}$ depends on not only the structure of the model but also the initial condition $f_{X_1^*,M_1}$. In the special case where $f_{X_1^*,M_1} = f_{X_1^*} f_{M_1}$,which means the observed and the unobserved state variables are independent at the first period of observation, we have

$$f_{X_2^*,w_2,M_1} = f_{y_2|m_2,X_2^*} f_{X_2^*} f_{m_2|X_2^*,M_1} f_{M_1},$$

We then have the corresponding operator equation

$$L_{X_2^*,w_2,M_1} = D_{y_2|m_2,X_2^*} D_{X_2^*} L_{m_2|X_2^*,M_1} D_{M_1}.$$

Given that all the densities in the diagonal operators are nonzero and bounded, we then need to show the injectivity of $L_{m_2|X_2^*,M_1}$. In fact, $f_{m_2|X_2^*,M_1}$ for any given m_2 contains a convolution form:

$$f_{m_2|X_2^*,M_1}(m_2|x_2^*,m_1) = \frac{m_1}{m_2}\left|\frac{\partial}{\partial m_1}\varphi(m_2,m_1)\right| f_{\eta_2}\left(\varphi(m_2,m_1) - x_2^*\right),$$

where $\varphi(m_2,m_1) = \ln\frac{m_2-m_1}{m_1}$ and $\frac{m_1}{m_2} \in (0,1)$. Therefore, applying Lemma 4 with $z = x_2^*$ and $x = \varphi(m_2,m_1)$, we obtain that $L_{X|Z}^*$ is injective, which implies $L_{m_2|X_2^*,M_1}$ is injective because the extra term $\frac{m_1}{m_2} \in (0,1)$ on the RHS corresponds to an injective diagonal operator. Thus, the operator $L_{X_2^*,w_2,M_1}$ is also injective. That concludes our proof of injectivity of L_{M_4,w_3,w_2,M_1}. ■

Proof of Claim 2: Here we show the injectivity of L_{M_1,w_2,w_3,M_4} given the arguments made so far. We have for fixed (w_2,w_3)

$$\begin{aligned}
f_{M_1,w_2,w_3,M_4} &= f_{M_4|w_3,X_3^*} f_{w_3|w_2,X_3^*} f_{X_3^*,w_2,M_1}. \\
&= f_{M_4|w_3,X_4^*} f_{X_4^*|w_3,X_3^*} f_{w_3|w_2,X_3^*} f_{X_3^*|w_2,X_2^*} f_{X_2^*,w_2,M_1} \\
&= f_{M_4|w_3,X_4^*} f_{X_4^*|w_3,X_3^*} f_{w_3|w_2,X_3^*} f_{X_3^*|w_2,X_2^*} f_{y_2|m_2,X_2^*} f_{X_2^*} f_{m_2|X_2^*,M_1} f_{M_1}.
\end{aligned}$$

Therefore, the equivalent operator equation is

$$\begin{aligned}
L_{M_1,w_2,w_3,M_4} &= L_{M_1,y_2,m_2,y_3,m_3,M_4} \\
&= D_{M_1} L_{m_2|M_1,X_2^*} D_{X_2^*} D_{y_2|m_2,X_2^*} L^*_{X_3^*|w_2,X_2^*} D_{w_3|w_2,X_3^*} L^*_{X_4^*|w_3,X_3^*} L^*_{M_4|w_3,X_4^*}.
\end{aligned}$$

Every L operator in the equation above is based on a convolution form. For example, $L_{m_2|M_1,X_2^*}$ has a convolution kernel. We let $z = x_2^*$ and $x = \varphi(m_2,m_1)$, the injectivity of the operator $L_{m_2|M_1,X_2^*}$ is implied by that of $L_{X|Z}$ in Lemma 4. The same argument applies to other operators. Thus, L_{M_1,w_2,w_3,M_4} is also injective.

A completely analogous argument can be made for L_{M_1,w_2,M_3}, which we omit for brevity.

■

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. Honoré (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.

Carroll, R., X. Chen, and Y. Hu (2009): "Identification and estimation of nonlinear models using two samples with nonclassical measurement errors," *Journal of Nonparametric Statistics*, forthcoming.

Chen, X., H. Hong, and D. Nekipelov (2007): "Measurement Error Models," manuscript, Yale 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. (2009): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, forthcoming.

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.

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," *Review of Economic Studies*, 74, 1059–1087.

Hendel, I., and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Stockpiling Behavior," *Econometrica*, 74, 1637–1674.

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 (2009): "Pairwise-Difference Estimation of a Dynamic Optimization Model," *Review of Economic Studies*, forthcoming.

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 (2009): "Nonparametric Identification of Dynamic Models with Unobserved State Variables, Supplemental Material," available at *http://www.hss.caltech.edu/~mshum/papers/onlineapp.pdf*.

Imai, S., N. Jain, and A. Ching (2009): "Bayesian Estimation of Dynamic Discrete Choice Models," *Econometrica*, 77, 1865–1899.

Kasahara, H., and K. Shimotsu (2009): "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choice," *Econometrica*, 77, 135–175.

Levinsohn, J., and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–342.

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.

Olley, S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

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

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*, 38, 373–399.

Pakes, A., and M. Simpson (1989): "Patent Renewal Data," *Brookings Papers on Economic Activity: Microeconomics*, pp. 331–403.

Pesendorfer, M., and P. Schmidt-Dengler (2008): "Asymptotic Least Squares Estimators for Dynamic Games," *Review of Economic Studies*, 75, 901–928.

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.

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, 201–23.

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