Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Nonparametric identification of dynamic models with unobserved state variables[☆]

Yingyao Hu[a,*], Matthew Shum[b]

[a] Johns Hopkins University, United States
[b] California Institute of Technology, United States

ARTICLE INFO

Article history:
Received 27 November 2010
Received in revised form
29 September 2011
Accepted 30 May 2012
Available online xxxx

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 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 (USVs), 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

☆ 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, SITE 2009, and the 2010 "Five Star Forum" at Renmin University's Hanqing Institute provided useful comments.
* Corresponding author. Tel.: +1 410 516 7610; fax: +1 410 516 7600.
E-mail addresses: yhu@jhu.edu (Y. Hu), mshum@caltech.edu (M. Shum).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.05.023

models with 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 time-varying 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), the law of motion $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ 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 CCPs 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 et al. (1994).[1] A general criticism of these methods is that they cannot accommodate unobserved state variables. In response, Aguirregabiria and Mira (2007), Buchinsky et al. (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. Similarly, Kasahara and Shimotsu (2009, hereafter KS) consider the identification of dynamic models with discrete unobserved heterogeneity, where the latent variable $X_t^* = X^*$ 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 these papers, we consider a more general setting where the unobserved X_t^* is allowed to vary over time (as in the Miller and Pakes examples above), and can evolve depending on past values of the observed variables W_{t-1}. Our focus is on the identification of such models. Our identification approach is novel because it is based on recent econometric results in nonlinear measurement error models.[2] Specifically, we show that the identification results in Hu and Schennach (2008) and Carroll et al. (2010) 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.

Our results extend nonparametric identification to classes of models not covered in the existing identification literature. When the unobserved state variable X_t^* is discrete, our results cover cases where X_t^* is time-varying and can evolve depending on past values of the observed variables W_{t-1}. This is new in the literature. When X_t^* is continuous, however, our identification results require high-level "completeness" assumptions which are difficult to verify in practice. One worked-out example (in Section 4.2) shows that these completeness assumptions are implied by independent initial conditions, in addition to other restrictions on the laws of motion of the state variables: while this is new ground, these restrictions are nevertheless strong. Because of this, when X_t^* is continuous, we see our results more as a useful starting point, rather than a final word on the subject.

Our identification approach is quite distinct from other recent papers which have studied identification and estimation of dynamic models with unobserved and time-varying state variables. 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 discrete Markov process.[3] Henry et al. (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, they share a common starting point in that 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 et al. (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 $(W_t, X_t^*|W_{t-1}, X_{t-1}^*)$. 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 two 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, Appendix A contains the proofs for several lemmas and corollaries.

2. Overview of assumptions

We assume that for each agent i, $\{(W_T, X_T^*), \ldots, (W_t, X_t^*), \ldots, (W_1, X_1^*)\}_i$ 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 non-stationary 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 special 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 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 $\Omega_{<t-1} \equiv \{W_{t-2}, \ldots, W_1, X_{t-2}^*, \ldots, X_1^*\}$, the history up to (but not including) $t-1$.
(ii) *Limited feedback*: $f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} = f_{W_t|W_{t-1},X_t^*}$.

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

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

[3] That is, X_t^* is discrete-valued, and depends stochastically only on X_{t-1}^*, and not on any other variables. We relax this in Section 4.1.

[4] Similarly, Bouissou et al. (1986) exploit the Markov restrictions on a stochastic process X to formulate tests for the noncausality of another process Y on X.

Assumption 1(i), a first-order Markov assumption, is satisfied for Markovian dynamic decision models (cf. Rust, 1994). Assumption 1(ii) is a "limited feedback" assumption, which 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:

$$\begin{aligned} f_{W_t,X^*_t|W_{t-1},X^*_{t-1},\Omega_{<t-1}} &= f_{W_t,X^*_t|W_{t-1},X^*_{t-1}} \\ &= f_{W_t|W_{t-1},X^*_t,X^*_{t-1}} f_{X^*_t|W_{t-1},X^*_{t-1}}, \end{aligned} \tag{2}$$

and

$$\begin{aligned} f_{W_t|W_{t-1},X^*_t,X^*_{t-1}} &= f_{W_t|W_{t-1},X^*_t} \\ &= f_{Y_t,M_t|Y_{t-1},M_{t-1},X^*_t} \\ &= f_{Y_t|M_t,Y_{t-1},M_{t-1},X^*_t} f_{M_t|Y_{t-1},M_{t-1},X^*_t}. \end{aligned} \tag{3}$$

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 further simplifies to $f_{Y_t|M_t,X^*_t}$ (the CCP), 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 $X^*_t = X^*_{t-1}, \ \forall t$, 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 $X^*_{t-1}, Y_{t-1}, M_{t-1}$. □

In this paper, we assume that the unobserved state variable X^*_t is scalar-valued, and is drawn from a continuous distribution.[7] Since W_t is usually a vector, we first reduce the dimensionality of W_t by defining

$$V_t \equiv g_t(W_t), \tag{4}$$

where the function $g_t : \mathbb{R}^d \to \mathbb{R}$ is known with d being the dimension of W_t. We treat V_{t-2} and V_{t+1} as noisy "measurements" of the latent X^*_t and use the identification strategies in Carroll et al. (2010, Assumption 2.4) to achieve the nonparametric identification of our model. Before we introduce our identification assumptions, we connect our model to the existing nonclassical measurement error models in Hu and Schennach (2008) and Carroll et al. (2010).

Hu and Schennach (2008) consider a framework where three observed measurements (X, Y, Z) are conditionally independent given a latent variable X^*. In other words, the four variables (X, Y, Z, X^*) satisfy

$$f_{X,Y,Z} = \int f_{X|X^*} f_{Y|X^*} f_{X^*,Z} dx^*. \tag{5}$$

They use a spectral decomposition technique to show that under reasonable assumptions all the elements $f_{X|X^*}, f_{Y|X^*}, f_{X^*,Z}$ are nonparametrically identified from $f_{X,Y,Z}$. Besides the conditional independence, other key assumptions include that (i) the linear operators corresponding to density functions $f_{X|X^*}$ and $f_{Z|X}$ are injective; (ii) the eigenvalues corresponding to $f_{Y|X^*}$ are distinctive; (iii) the measurement error distribution $f_{X|X^*}$ satisfies a zero location assumption.

Carroll et al. (2010) consider the identification of a model $f_{Y|X^*,Z}$ with a latent X^* using two survey samples $\{X, Y, Z, S\}$, where S is a binary indicator for the two samples. They assume the three observables (X, Y, Z) in the two samples satisfy

$$f_{X,Y,Z,S} = \int f_{X|X^*,S} f_{X^*,Z,S} f_{Y|X^*,Z} dx^*, \tag{6}$$

where $f_{Y|X^*,Z}$ is the model of interest, $f_{X|X^*,S}$ is the measurement error distribution in the two samples, and $f_{X^*,Z,S}$ is the joint distribution of explanatory variables in the two samples. Since the results in Hu and Schennach (2008) do not directly apply to this framework, Carroll et al. (2010) use a clever trick to extend the spectral decomposition technique to this framework to show the nonparametric identification of all the elements on the RHS. Besides the conditional independence, other key assumptions include that (i) the linear operators corresponding to $f_{X|X^*,S}$ and $f_{X,Y,Z,S}$ are injective; (ii) the eigenvalues are distinctive; (iii) the measurement error distribution $f_{X|X^*,S}$ satisfies a zero location assumption.

Under Assumption 1, our paper considers a framework as follows

$$\begin{aligned} f_{W_{t+1},W_t,W_{t-1},W_{t-2}} &= \int f_{W_{t+1}|W_t,X^*_t} f_{W_t|W_{t-1},X^*_t} f_{X^*_t,W_{t-1},W_{t-2}} dx^*_t \\ &= \int f_{W_{t+1}|W_t,X^*_t} f_{W_t,W_{t-1},X^*_t} f_{W_{t-2}|X^*_t,W_{t-1}} dx^*_t. \end{aligned} \tag{7}$$

Comparing Eqs. (6) and (7), we may use the identification strategy in Carroll et al. (2010) with $\left(W_{t+1}, W_t, W_{t-1}, W_{t-2}, X^*_t\right)$ corresponding to (X, S, Z, Y, X^*), respectively. This consists of the key step of our identification. We also make assumptions corresponding to those in Carroll et al. (2010).[8]

Denote the supports of V_t and W_t as $\mathcal{V}_t$ and $\mathcal{W}_t$, respectively.[9] The linear operator $L_{V_{t-2},\overline{w}_{t-1},\overline{w}_t,V_{t+1}}$ is a mapping from the $\mathcal{L}^p$-space of functions of V_{t+1} to the $\mathcal{L}^p$ space of functions of V_{t-2},[10]

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

7 A discrete distribution for X^*_t, which is assumed in many applied settings (e.g. Arcidiacono and Miller, 2006) is a special case, which we will consider as an example in Section 4.1.

8 Shiu and Hu (2010) use the identification results in Hu and Schennach (2008) in the measurement error literature to identify a panel data model. The identification strategy only requires three periods of data. The limited feedback assumption in our paper is more general than the one used in Shiu and Hu (2010) so that we require five periods of data and Shiu and Hu (2010) need three periods of data in the comparable setting. Their assumptions are tailored for panel data models. Our framework is more suitable for IO models, where the conditional independence assumptions are based on empirical IO models.

9 Here, capital letters denote random variables, while lower-case letters denote realizations.

10 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, i.e. $\int_{\mathcal{X}} |h(x)|^p d\mu(x) < \infty$, 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.

Please cite this article in press as: Hu, Y., Shum, M., Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics (2012), doi:10.1016/j.jeconom.2012.05.023

defined as[11]

$$\left(L_{V_{t-2},\overline{w}_{t-1},\overline{w}_t,V_{t+1}}h\right)(v_{t-2}) = \int f_{V_{t-2},W_{t-1},W_t,V_{t+1}}(v_{t-2},\overline{w}_{t-1},\overline{w}_t,v_{t+1})h(v_{t+1})dv_{t+1}; \quad h\in\mathcal{L}^p(\mathcal{V}_{t+1}),\ \overline{w}_{t-1}\in\mathcal{W}_{t-1},\ \overline{w}_t\in\mathcal{W}_t. \tag{8}$$

Similarly, we define the diagonal (or multiplication) operator

$$\left(D_{\overline{w}_t|\overline{w}_{t-1},x_t^*}h\right)(x_t^*) = f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|\overline{w}_{t-1},x_t^*)h(x_t^*); \quad h\in\mathcal{L}^p\left(\mathcal{X}_t^*\right),\overline{w}_{t-1}\in\mathcal{W}_{t-1},\overline{w}_t\in\mathcal{W}_t. \tag{9}$$

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.)[12] The next two assumptions ensure the validity and uniqueness of this decomposition.

Assumption 2. *Invertibility*. There exists variable(s) V_t such that

(i) for any $w_t \in \mathcal{W}_t$, there exist a $w_{t-1}\in\mathcal{W}_{t-1}$ and a neighborhood $\mathcal{N}^r$ around (w_t,w_{t-1})[13] such that, for any $(\overline{w}_t,\overline{w}_{t-1})\in\mathcal{N}^r$, $L_{V_{t-2},\overline{w}_{t-1},\overline{w}_t,V_{t+1}}$ 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 $w_{t-1}\in\mathcal{W}_{t-1}$, 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) impose the same restrictions between the measurements and the latent variable as Hu and Schennach (2008, Assumption 3) and Carroll et al. (2010, 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.[14]

The statements in Assumption 2 are equivalent to *completeness* conditions which have recently been employed in the nonparametric IV literature: namely, an operator $L_{V_{t-2},\overline{w}_{t-1},\overline{w}_t,V_{t+1}}$ is one-to-one if the corresponding density function $f_{V_{t-2},W_{t-1},W_t,V_{t+1}}$ satisfies a "completeness" condition: for any $(\overline{w}_{t-1},\overline{w}_t)$,

$$\int f_{V_{t-2},W_{t-1},W_t,V_{t+1}}(v_{t-2},\overline{w}_{t-1},\overline{w}_t,v_{t+1})h(v_{t+1})dv_{t+1}=0 \quad \text{for all } v_{t-2} \text{ implies } h(v_{t+1})=0 \text{ for all } v_{t+1}. \tag{10}$$

Completeness is a high-level condition, and special cases of it have been considered in, e.g. Newey and Powell (2003), Blundell et al. (2007) and d'Haultfoeuille (2011). 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 V_{t+1} defined in Eq. (4) is a function of W_{t+1}. Intuitively, by Assumption 2(ii), the variable V_{t+1} is a component of W_{t+1} which "transmits" information on the latent X_t^* conditional on W_t, the observables in the previous period. We consider suitable choices of V_{t+1} for specific examples in Section 4.[15] Assumption 2(ii) also rules out models where X_t^* has a continuous support, but W_{t+1} contains only discrete components. In this case, there is no V_{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 $w_{t-1}\in\mathcal{W}_{t-1}$ and corresponding neighborhood $\mathcal{N}^r$ satisfying Assumption 2(i) such that, for some $(\overline{w}_t,\overline{w}_{t-1})\in\mathcal{N}^r$ with $\overline{w}_t\neq w_t,\ \overline{w}_{t-1}\neq w_{t-1}$:

(i) $0<k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},x_t^*\right)<C<\infty$ for any $x_t^*\in\mathcal{X}_t^*$ and some constant C;
(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

$$k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},x_t^*\right) = \frac{f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1},x_t^*)f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|\overline{w}_{t-1},x_t^*)}{f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|w_{t-1},x_t^*)f_{W_t|W_{t-1},X_t^*}(w_t|\overline{w}_{t-1},x_t^*)}. \tag{11}$$

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. (45) shows, the $k(\cdots)$ function in the assumption corresponds to the eigenvalues in this decomposition, so that conditions (i) and (ii) 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.[16] □

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, to "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,

11 Analogously, the operator $L_{V_{t+1}|w_t,X_t^*}$, corresponding to the conditional density $f_{V_{t+1}|W_t,X_t^*}$, is defined, for all functions $h\in\mathcal{L}^p\left(\mathcal{X}_t^*\right)$, and $w_t\in\mathcal{W}_t$ as $\left(L_{V_{t+1}|w_t,X_t^*}h\right)(v_{t+1}) = \int f_{V_{t+1}|W_t,X_t^*}(v_{t+1}|w_t,x_t^*)h(x_t^*)dx_t^*$.

12 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 the corresponding matrix. This discrete case is discussed in detail in Section 4.1.

13 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.

14 Additional details are given in Section 3 of the online appendix (Hu and Shum, 2009).

15 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.

16 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. (45) 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.

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.[17] 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 $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$,*

$$L_{V_{t+1},w_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*,w_{t-1},V_{t-2}}. \tag{12}$$

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

$$L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} = L^{-1}_{V_{t+1}|w_t,X_t^*} L_{V_{t+1},w_t,w_{t-1},V_{t-2}} L^{-1}_{V_t,w_{t-1},V_{t-2}} L_{V_t|w_{t-1},X_{t-1}^*}. \tag{13}$$

Proofs. In Appendix A. □

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 1–4 imply that the density $f_{V_{t+1},W_t,W_{t-1},V_{t-2}}$ uniquely determines the density $f_{V_{t+1}|W_t,X_t^*}$.*

Proofs. In Appendix A. □

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.[18] Hence, we have shown the following result.

Theorem 1 (*Identification of Markov Law of Motion, Non-Stationary Case*). *Under Assumptions 1–4, the density $f_{W_{t+1},W_t,W_{t-1},W_{t-2},W_{t-3}}$ for any $t \in \{4,\ldots,T-1\}$ 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 (e.g. Hotz et al., 1994 and Bajari et al., 2007a) require simulation of the Markov process $(W_t, X_t^*, W_{t+1}, X_{t+1}^*, W_{t+2}, X_{t+2}^*, \ldots)$ 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 Assumptions 1–4, the density $f_{W_{t+1},W_t,W_{t-1},W_{t-2},W_{t-3}}$ for any $t \in \{4,\ldots,T-1\}$ uniquely determines the density f_{W_{t-1},X_{t-1}^*}.*

Proof. In Appendix A. □

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 1–5, 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 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 1–5, 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 A. □

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

[17] 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 CCPs 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.

[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.

[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.

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 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. 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.

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: $\forall t,\ \mathrm{supp}X_t^* = \mathrm{supp}W_t \equiv \{0, 1\}$. This is the simplest example of the models considered in our framework. One example of such a model is a binary version of Abbring et al.'s (2008) "dynamic moral hazard" model of auto insurance. In that model, W_t is a binary indicator of claim 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:

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

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.

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.[20] 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, Assumption 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 ordering in recovering $L_{W_{t+1}|w_t,X_t^*}$ 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}^*).[21]

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

$$\begin{aligned}\text{for } w_t = \{0, 1\} :\ &\mathbb{E}[W_{t+1}|w_t, X_t^* = 1]\\&< \mathbb{E}[W_{t+1}|w_t, X_t^* = 0].\end{aligned} \tag{16}$$

The restriction (16) implies that future claims W_{t+1} occur less frequently with higher effort today, and imposes additional restrictions on the $p(\cdot)$ and $q(\cdot)$ functions in (14).[22]

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 (16) 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). This framework, in which firms make incremental "investment" decisions which affect the growth of an underlying "capital stock" variable, stems from the work of Ericson and Pakes (1995). More recently, such models have been usefully applied in empirical work in IO (Ryan, 2006), productivity (Xu, 2007; Collard-Wexler, 2006), and international trade (Dunne et al., 2006).[23] 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, which is unobserved by the econometrician but observed by the firm, and affects their R&D choices.

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

$$X_t^* = 0.8X_{t-1}^* + 0.2 \exp\left(\psi\left(Y_{t-1}\right)\right) \nu_t. \tag{17}$$

In the above, $\nu_t \in \mathbb{R}$ is a standard normal shock, distributed independently over t, and $\psi(\cdot) < \infty$, $\psi'(\cdot) > 0$. Eq. (17) implies $f_{X_t^*|Y_{t-1},M_{t-1},X_{t-1}^*} = f_{X_t^*|Y_{t-1},X_{t-1}^*}$.

[20] 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 $i, j \in \{0, 1\}$.

[21] We thank Thierry Magnac for this insight.

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

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

Please cite this article in press as: Hu, Y., Shum, M., Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics (2012), doi:10.1016/j.jeconom.2012.05.023

Installed base evolves as:

$$M_{t+1} = M_t[1 + \exp(\eta_{t+1} + X^*_{t+1})] \tag{18}$$

where $\eta_{t+1} \in \mathbb{R}$ is a random shock following the extreme value distribution, with density $f_{\eta_{t+1}}(\eta) = \exp(\eta - e^\eta)$ for $\eta \in \mathbb{R}$, independently across t. This law of motion also implies that $f_{M_{t+1}|Y_t,M_t,X^*_t,X^*_{t+1}} = f_{M_{t+1}|M_t,X^*_{t+1}}$. Eq. (18) implies that, *ceteris paribus*, product quality raises installed base. Moreover, we also assume that the initial installed base $M_1 > 0$, so that $M_t > 0$ for all t and, for a given M_t, $M_{t+1} \in (M_t, +\infty)$.

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

$$\begin{aligned} Y_t &= Y^*(M_t, X^*_t, \gamma_t) \\ &= \text{argmax}_{0\le y\le \bar{I}} \left[\underbrace{\Pi(M_t, X^*_t)}_{\text{profits}} - \underbrace{\gamma_t}_{\text{shock}} \cdot \underbrace{Y_t^2}_{\text{R\&D cost}} \right. \\ &\quad \left. + \beta\mathbb{E}\underbrace{V(M_{t+1}, X^*_{t+1}, \gamma_{t+1})}_{\text{value fxn}} \right] \end{aligned} \tag{19}$$

where $\bar{I}$ 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. The RHS of Eq. (19) 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

$$f_{W_3|W_2,X^*_3} = f_{Y_3|M_3,X^*_3} \cdot f_{M_3|M_2,X^*_3}. \tag{20}$$

The first term in Eq. (20) is the density of R&D Y_3. Because the first term is not a function of M_2, Eq. (45) 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 as follows:

$$\begin{aligned} &k\left(w_3, \overline{w}_3, w_2, \overline{w}_2, x^*_3\right) \\ &= \frac{f_{W_3|W_2,X^*_3}(w_3|w_2, x^*_3) f_{W_3|W_2,X^*_3}(\overline{w}_3|\overline{w}_2, x^*_3)}{f_{W_3|W_2,X^*_3}(\overline{w}_3|w_2, x^*_3) f_{W_3|W_2,X^*_3}(w_3|\overline{w}_2, x^*_3)} \\ &= \frac{f_{M_3|M_2,X^*_3}(m_3|m_2, x^*_3) f_{M_3|M_2,X^*_3}(\overline{m}_3|\overline{m}_2, x^*_3)}{f_{M_3|M_2,X^*_3}(\overline{m}_3|m_2, x^*_3) f_{M_3|M_2,X^*_3}(m_3|\overline{m}_2, x^*_3)}. \end{aligned} \tag{21}$$

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 $\gamma_3 > 0$, so also 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 with density

$$\begin{aligned} f_{M_3|M_2,X^*_3}(m_3|m_2, x^*_3) &= \frac{1}{(m_3 - m_2)} \exp\left[\log\left(\frac{m_3 - m_2}{m_2}\right) \right. \\ &\quad \left. - x^*_3 - e^{\log\left(\frac{m_3-m_2}{m_2}\right) - x^*_3}\right] \\ &= \frac{e^{-x^*_3}}{m_2} \exp\left(-e^{-x^*_3}\left[\frac{m_3 - m_2}{m_2}\right]\right). \end{aligned} \tag{22}$$

Plugging this into Eq. (21), we obtain an expression for the eigenvalues

$$\begin{aligned} &k\left(w_3, \overline{w}_3, w_2, \overline{w}_2, x^*_3\right) \\ &= \exp\left(-e^{-x^*_3}\left[\frac{-\left(\overline{m}_3 - m_3\right)\left(\overline{m}_2 - m_2\right)}{m_2\overline{m}_2}\right]\right). \end{aligned} \tag{23}$$

For given m_3, we can pick a finite and nonzero m_2,[24] and set $(\overline{m}_3, \overline{m}_2) = (m_3 - \Delta, m_2 + \Delta)$, with Δ nonzero and small. At these values, the eigenvalues in Eq. (23) simplify to $\exp(-e^{-x^*_3}[\frac{\Delta^2}{m_2(m_2+\Delta)}])$ so that, for fixed m_3, and $x^*_3 \in \mathbb{R}$, $0 < k\left(w_3, \overline{w}_3, w_2, \overline{w}_2, x^*_3\right) < 1$, which satisfies Assumption 3(i). Moreover, the eigenvalues in Eq. (23) are monotonic in x^*_3 for any given $(w_3, \overline{w}_3, w_2, \overline{w}_2)$, which implies Assumption 3(ii).

To verify Assumption 4, we set $V_t = M_t$ for all t. Note $\mathbb{E}[\log \frac{M_4-m_3}{m_3}|m_3, y_3, x^*_3] = \mathbb{E}[\eta_4] + \mathbb{E}[X^*_4|x^*_3, y_3]$. Because the law of motion for product quality $X^*_4 = 0.8X^*_3 + 0.2\exp(\psi(Y_3))\nu_4$ implies that $\mathbb{E}[X^*_4|x^*_3, y_3]$ is monotonic in x^*_3, we set the functional G to be $x^*_3 = \mathbb{E}[\log \frac{M_4-m_3}{m_3}|m_3, y_3, x^*_3]$.

Finally, Assumption 2 contains three injectivity assumptions. As before, we use $V_t = M_t$, for all periods t. 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. (17) and (18)) 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, it is sufficient 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} for any (w_2, w_3) in the support. 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 1, Assumption 1 implies that

$$\begin{aligned} L_{M_4,w_3,w_2,M_1} &= L_{M_4|w_3,X^*_3} D_{w_3|w_2,X^*_3} L_{X^*_3,w_2,M_1} \\ &= L_{M_4|w_3,X^*_3} D_{w_3|w_2,X^*_3} L_{X^*_3|w_2,X^*_2} L_{X^*_2,w_2,M_1} \end{aligned} \tag{24}$$

$$L_{M_3,w_2,M_1} = L_{M_3|w_2,X^*_2} L_{X^*_2,w_2,M_1}. \tag{25}$$

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 $L_{X^*_2,w_2,M_1}$.[25] It turns out that assumptions we have made already for this example ensure that three of these operators are injective. We discuss each case in turn.

(i) The diagonal operator $D_{w_3|w_2,X^*_3}$ has kernel function $f_{w_3|w_2,X^*_3} = f_{y_3|m_3,X^*_3} f_{m_3|m_2,X^*_3}$. In the discussion on Assumption 3(i) above, we showed 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 $\left(m_3, m_2, x^*_3\right)$ in the support. Therefore, $D_{w_3|w_2,X^*_3}$ is injective.

(ii) For $L_{M_4|w_3,X^*_4}$, we use Eq. (18) whereby, for every (y_3, m_3), M_4 is a convolution of X^*_4, i.e. $\log[M_4 - M_3] - \log M_3 = X^*_4 + \eta_4$. We have

$$g(m_4) \equiv \left(L_{M_4|w_3,X^*_4} h\right)(m_4)$$

[24] In verifying Assumption 2(i) below, we show that the assumption holds for all (w_3, w_2), so that the neighborhood $\mathcal{N}^r$ is unrestricted. Hence, in verifying Assumption 3(i) here, we can pick any m_2, and also pick any other point $(\overline{m}_3, \overline{m}_2)$ as needed.

[25] 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. Our notion of stationarity here is distinct from the notion of covariance-stationarity for stochastic processes. Indeed, as defined in Eq. (18), the M_t process may not be covariance-stationary, but the law of motion $f_{M_4|w_3,X^*_4}$ is still time-invariant.

$$= \int_{-\infty}^{\infty} f_{M_4|w_3,X_4^*}\left(m_4|w_3,x_4^*\right) h(x_4^*)dx_4^*$$
$$= \int_{-\infty}^{\infty} \frac{1}{m_4-m_3} f_{\eta_4}\left(\log\left(\frac{m_4-m_3}{m_3}\right)-x_4^*\right) h(x_4^*)dx_4^*$$
$$= \frac{1}{m_4-m_3}\int_{-\infty}^{\infty} f_{\eta_4}\left(\varphi_4-x_4^*\right) h(x_4^*)dx_4^*,$$
$$\left[\varphi_4 \equiv \log\left(\frac{m_4-m_3}{m_3}\right)\right]$$
$$\equiv \frac{1}{m_4-m_3}\times\left(L_{\varphi_4|X_4^*}h\right)(\varphi_4). \quad (26)$$

Since the function $\frac{1}{m_4-m_3}$ is nonzero, $g(m_4) = 0$ for any $m_4 \in (m_3,\infty)$ implies $\left(L_{\varphi_4|X_4^*}h\right)(\varphi_4) = 0$ for any $\varphi_4 \in \mathbb{R}$, where the kernel of the operator $L_{\varphi_4|X_4^*}$ has a convolution form $f_{\eta_4}\left(\varphi_4 - x_4^*\right)$. As shown in Lemma 4, as long as the characteristic function of η_4 has no real zeros, which is satisfied by the assumed extreme value distribution,[26]the corresponding operator $L_{\varphi_4|X_4^*}$ is injective. Therefore, $\left(L_{\varphi_4|X_4^*}h\right)(\varphi_4) = 0$ for any $\varphi_4 \in \mathbb{R}$ implies $h\left(x_4^*\right) = 0$ for any $x_4^* \in \mathbb{R}$. Thus, the operator $L_{M_4|w_3,X_4^*}$ is injective.

(iii) Similarly, for fixed w_2, X_3^* is a convolution of X_2^*, i.e. $X_3^* = 0.8X_2^* + 0.2\exp(\psi(Y_2))\nu_3$ (cf. Eq. (17)). By an argument similar to that for the previous operator, we can show that $L_{X_3^*|w_2,X_2^*}$ is injective.

(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 injectivity directly. This is because this joint density confounds both the structural components (laws of motion) in the model and the initial condition $f_{X_1^*,M_1}$. Thus in general, injectivity of this operator is not verifiable based only 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[27] – that is, $\psi(\cdot) = 0$ in Eq. (17) – it turns out that an additional independence assumption on the initial values of the state variables (X_1^*, M_1), i.e., $f_{X_1^*,M_1} = f_{X_1^*}f_{M_1}$, suffices to ensure injectivity of the operator $L_{X_2^*,w_2,M_1}$:

Claim 1. *If $\psi(\cdot) = 0$ in Eq. (17), 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 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}. 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 restrictions on the joint distribution of the initial values of the state variables (X_1^*, M_1), may appear artificial. However, given the recursive nature of Markovian dynamic optimization models, we believe that restrictions on initial conditions will be, generally, unavoidable in verifying completeness. However, the exact nature of the restrictions will differ on an example-by-example basis. Here, only restrictions on the initial distribution of the state variables (X_1^*, M_1) were required. At the same time, we also reiterate that these are sufficient conditions, and may not be necessary for the general results. In Appendices, we provide and discuss a *necessary* condition for operators to satisfy completeness, which allows for very general and flexible classes of joint densities.

5. Concluding remarks

We have considered the identification of a first-order Markov process $\left\{W_t, X_t^*\right\}$ when only $\{W_t\}$ is observed. Under nonstationarity, 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 et al. (2007b), 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 et al. (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]

Appendix A. Proofs

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

$$\iint f_{W_{t+1},W_t,X_t^*,X_{t-1}^*,W_{t-1},W_{t-2}}dx_t^*dx_{t-1}^*$$
$$= \iint f_{W_{t+1}|W_t,W_{t-1},W_{t-2},X_t^*,X_{t-1}^*} f_{W_t,X_t^*|W_{t-1},W_{t-2},X_{t-1}^*} \times f_{X_{t-1}^*,W_{t-1},W_{t-2}}dx_t^*dx_{t-1}^*$$
$$= \iint f_{W_{t+1}|W_t,X_t^*} f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*,W_{t-1},W_{t-2}}dx_t^*dx_{t-1}^*$$
$$= \iint f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} f_{X_t^*|W_{t-1},X_{t-1}^*} \times f_{X_{t-1}^*,W_{t-1},W_{t-2}}dx_t^*dx_{t-1}^*$$
$$= \iint f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} f_{X_t^*|W_{t-1},W_{t-2},X_{t-1}^*}$$

[26] The characteristic function for η_4 is $\phi_{\eta_4}(\tau) = \Gamma(1+i\tau)$, which is nonzero for any $\tau \in \mathbb{R}$.

[27] A large class of investment models (e.g. Olley and Pakes, 1996, Levinsohn and Petrin, 2003) assume that the unobserved variable X_t^* (denoting productivity) evolves exogenously.

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

Please cite this article in press as: Hu, Y., Shum, M., Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics (2012), doi:10.1016/j.jeconom.2012.05.023

$$\times f_{X^*_{t-1},W_{t-1},W_{t-2}} dx^*_t dx^*_{t-1}$$
$$= \iint f_{W_{t+1}|W_t,X^*_t} f_{W_t|W_{t-1},X^*_t,X^*_{t-1}} f_{X^*_t,X^*_{t-1},W_{t-1},W_{t-2}} dx^*_t dx^*_{t-1}. \quad (27)$$

(We omit the arguments in the density functions as long as doing this does not cause confusion.) Assumption 1(ii) then implies

$$f_{W_{t+1},W_t,W_{t-1},W_{t-2}} = \int f_{W_{t+1}|W_t,X^*_t} f_{W_t|W_{t-1},X^*_t} \times \left(\int f_{X^*_t,X^*_{t-1},W_{t-1},W_{t-2}} dx^*_{t-1} \right) dx^*_t$$
$$= \int f_{W_{t+1}|W_t,X^*_t} f_{W_t|W_{t-1},X^*_t} f_{X^*_t,W_{t-1},W_{t-2}} dx^*_t. \quad (28)$$

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

$$L_{W_{t+1},w_t,w_{t-1},W_{t-2}} = L_{W_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} L_{X^*_t,w_{t-1},W_{t-2}}. \quad (29)$$

For the variable(s) $V_t \subseteq W_t$, for all periods t, introduced in Assumption 2, Eq. (29) 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. □

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

$$f_{V_{t+1},W_t,W_{t-1},V_{t-2}} = \int f_{V_{t+1}|W_t,X^*_t} f_{W_t,X^*_t,W_{t-1},V_{t-2}} dx^*_t$$
$$f_{W_t,X^*_t,W_{t-1},V_{t-2}} = \int f_{W_t,X^*_t|W_{t-1},X^*_{t-1}} f_{X^*_{t-1},W_{t-1},V_{t-2}} dx^*_{t-1}. \quad (30)$$

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

$$L_{V_{t+1},w_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X^*_t} L_{w_t,X^*_t,w_{t-1},V_{t-2}} \quad (31)$$
$$L_{w_t,X^*_t,w_{t-1},V_{t-2}} = L_{w_t,X^*_t|w_{t-1},X^*_{t-1}} L_{X^*_{t-1},w_{t-1},V_{t-2}}.$$

Substituting the second line into the first, we get

$$L_{V_{t+1},w_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X^*_t} L_{w_t,X^*_t|w_{t-1},X^*_{t-1}} L_{X^*_{t-1},w_{t-1},V_{t-2}}$$
$$\Leftrightarrow L_{w_t,X^*_t|w_{t-1},X^*_{t-1}} L_{X^*_{t-1},w_{t-1},V_{t-2}} = L^{-1}_{V_{t+1}|w_t,X^*_t} L_{V_{t+1},w_t,w_{t-1},V_{t-2}} \quad (32)$$

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

$$f_{V_t,W_{t-1},V_{t-2}} = \int f_{V_t|W_{t-1},X^*_{t-1}} f_{X^*_{t-1},W_{t-1},V_{t-2}} dx^*_{t-1} \quad (33)$$

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

$$L_{V_t,w_{t-1},V_{t-2}} = L_{V_t|w_{t-1},X^*_{t-1}} L_{X^*_{t-1},w_{t-1},V_{t-2}}$$
$$\Rightarrow L_{X^*_{t-1},w_{t-1},V_{t-2}} = L^{-1}_{V_t|w_{t-1},X^*_{t-1}} L_{V_t,w_{t-1},V_{t-2}} \quad (34)$$

where the right-hand side applies Assumption 2(ii). Hence, substituting the above into Eq. (32), we obtain the desired representation

$$L_{w_t,X^*_t|w_{t-1},X^*_{t-1}} L^{-1}_{V_t|w_{t-1},X^*_{t-1}} L_{V_t,w_{t-1},V_{t-2}}$$
$$= L^{-1}_{V_{t+1}|w_t,X^*_t} L_{V_{t+1},w_t,w_{t-1},V_{t-2}}$$
$$\Rightarrow L_{w_t,X^*_t|w_{t-1},X^*_{t-1}} L^{-1}_{V_t|w_{t-1},X^*_{t-1}}$$
$$= L^{-1}_{V_{t+1}|w_t,X^*_t} L_{V_{t+1},w_t,w_{t-1},V_{t-2}} L^{-1}_{V_t,w_{t-1},V_{t-2}}$$
$$\Rightarrow L_{w_t,X^*_t|w_{t-1},X^*_{t-1}} = L^{-1}_{V_{t+1}|w_t,X^*_t} L_{V_{t+1},w_t,w_{t-1},V_{t-2}}$$
$$L^{-1}_{V_t,w_{t-1},V_{t-2}} L_{V_t|w_{t-1},X^*_{t-1}}. \quad (35)$$

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

Proof of Lemma 3. For each w_t, choose a w_{t-1} and a neighborhood $\mathcal{N}^r$ around (w_t, w_{t-1}) to satisfy Assumptions 2(i) and 3, and pick a $(\overline{w}_t, \overline{w}_{t-1})$ within the neighborhood $\mathcal{N}^r$ to satisfy Assumption 3. Because $(\overline{w}_t, \overline{w}_{t-1}) \in \mathcal{N}^r$, also $(\overline{w}_t, w_{t-1}), (\overline{w}_t, w_{t-1}) \in \mathcal{N}^r$. 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. (12) at the four pairs of points (w_t, w_{t-1}), $(\overline{w}_t, w_{t-1})$, $(w_t, \overline{w}_{t-1})$, $(\overline{w}_t, \overline{w}_{t-1})$, each pair of equations will share one operator in common. Specifically:

$$\text{for } (w_t, w_{t-1}) : L_{V_{t+1},w_t,w_{t-1},V_{t-2}}$$
$$= L_{V_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} L_{X^*_t,w_{t-1},V_{t-2}}, \quad (36)$$
$$\text{for } (\overline{w}_t, w_{t-1}) : L_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}}$$
$$= L_{V_{t+1}|\overline{w}_t,X^*_t} D_{\overline{w}_t|w_{t-1},X^*_t} L_{X^*_t,w_{t-1},V_{t-2}}, \quad (37)$$
$$\text{for } (w_t, \overline{w}_{t-1}) : L_{V_{t+1},w_t,\overline{w}_{t-1},V_{t-2}}$$
$$= L_{V_{t+1}|w_t,X^*_t} D_{w_t|\overline{w}_{t-1},X^*_t} L_{X^*_t,\overline{w}_{t-1},V_{t-2}}, \quad (38)$$
$$\text{for } (\overline{w}_t, \overline{w}_{t-1}) : L_{V_{t+1},\overline{w}_t,\overline{w}_{t-1},V_{t-2}}$$
$$= L_{V_{t+1}|\overline{w}_t,X^*_t} D_{\overline{w}_t|\overline{w}_{t-1},X^*_t} L_{X^*_t,\overline{w}_{t-1},V_{t-2}}. \quad (39)$$

Assumption 2(ii) implies that $L_{V_{t+1}|\overline{w}_t,X^*_t}$ is invertible. Moreover, Assumption 3(i) implies $f_{W_t|W_{t-1},X^*_t}(\overline{w}_t|w_{t-1}, x^*_t) > 0$ for all x^*_t so that $D_{\overline{w}_t|w_{t-1},X^*_t}$ is invertible. We can then solve for $L_{X^*_t,w_{t-1},V_{t-2}}$ from Eq. (37) as

$$D^{-1}_{\overline{w}_t|w_{t-1},X^*_t} L^{-1}_{V_{t+1}|\overline{w}_t,X^*_t} L_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}} = L_{X^*_t,w_{t-1},V_{t-2}}. \quad (40)$$

Plugging in this expression to Eq. (36) leads to

$$L_{V_{t+1},w_t,w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} D^{-1}_{\overline{w}_t|w_{t-1},X^*_t}$$
$$\times L^{-1}_{V_{t+1}|\overline{w}_t,X^*_t} L_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}}. \quad (41)$$

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

$$\mathbf{A} \equiv L_{V_{t+1},w_t,w_{t-1},V_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t,w_{t-1},V_{t-2}}$$
$$= L_{V_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} D^{-1}_{\overline{w}_t|w_{t-1},X^*_t} L^{-1}_{V_{t+1}|\overline{w}_t,X^*_t}. \quad (42)$$

Similar manipulations of Eqs. (38) and (39) lead to

$$\mathbf{B} \equiv L_{V_{t+1},\overline{w}_t,\overline{w}_{t-1},V_{t-2}} L^{-1}_{V_{t+1},w_t,\overline{w}_{t-1},V_{t-2}}$$
$$= L_{V_{t+1}|\overline{w}_t,X^*_t} D_{\overline{w}_t|\overline{w}_{t-1},X^*_t} D^{-1}_{w_t|\overline{w}_{t-1},X^*_t} L^{-1}_{V_{t+1}|w_t,X^*_t}. \quad (43)$$

Assumption 2(i) guarantees that, for any w_t, $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1})$ exist so that (9) and (10) are valid operations. Finally, we postmultiply Eq. (42) by Eq. (43) to obtain

$$\mathbf{AB} = L_{V_{t+1}|w_t,X^*_t} D_{w_t|w_{t-1},X^*_t} D^{-1}_{\overline{w}_t|w_{t-1},X^*_t}$$
$$\times \left(L^{-1}_{V_{t+1}|\overline{w}_t,X^*_t} L_{V_{t+1}|\overline{w}_t,X^*_t} \right)$$
$$\times D_{\overline{w}_t|\overline{w}_{t-1},X^*_t} D^{-1}_{w_t|\overline{w}_{t-1},X^*_t} L^{-1}_{V_{t+1}|w_t,X^*_t}$$
$$= L_{V_{t+1}|w_t,X^*_t} \left(D_{w_t|w_{t-1},X^*_t} D^{-1}_{\overline{w}_t|w_{t-1},X^*_t} D_{\overline{w}_t|\overline{w}_{t-1},X^*_t} \times D^{-1}_{w_t|\overline{w}_{t-1},X^*_t} \right) L^{-1}_{V_{t+1}|w_t,X^*_t}$$
$$\equiv L_{V_{t+1}|w_t,X^*_t} D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X^*_t} L^{-1}_{V_{t+1}|w_t,X^*_t}, \quad \text{where} \quad (44)$$

Please cite this article in press as: Hu, Y., Shum, M., Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics (2012), doi:10.1016/j.jeconom.2012.05.023

$$\begin{aligned}&\left(D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}h\right)\left(x_t^*\right)\\&=\left(D_{w_t|w_{t-1},X_t^*}D^{-1}_{\overline{w}_t|w_{t-1},X_t^*}D_{\overline{w}_t|\overline{w}_{t-1},X_t^*}D^{-1}_{w_t|\overline{w}_{t-1},X_t^*}h\right)\left(x_t^*\right)\\&=\frac{f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1},x_t^*)f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|\overline{w}_{t-1},x_t^*)}{f_{W_t|W_{t-1},X_t^*}(\overline{w}_t|w_{t-1},x_t^*)f_{W_t|W_{t-1},X_t^*}(w_t|\overline{w}_{t-1},x_t^*)}h(x_t^*)\\&\equiv k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},x_t^*\right)h\left(x_t^*\right).\end{aligned}\tag{45}$$

This equation implies that the observed operator **AB** on the left hand side of Eq. (44) 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. (44) is similar to the decomposition in Hu and Schennach (2008) or Carroll et al. (2010).

Assumption 3 ensures that this decomposition is unique. Specifically, Eq. (44) 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. (44) 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, and implies that for each w_t, we can find values $\overline{w}_t$, w_{t-1}, and $\overline{w}_{t-1}$ such that the eigenvalues are distinct across all x_t^*.[29,30]

Second, the eigenfunctions $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t,x_t^*)$ in the spectral decomposition (44) 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. (44). □

Proof of 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 $w_{t-1}\in\mathcal{W}_t$,

$$\begin{aligned}&f_{V_t,W_{t-1}=w_{t-1}}=L_{V_t|w_{t-1},X_{t-1}^*}f_{W_{t-1}=w_{t-1},X_{t-1}^*}\\&\Leftrightarrow f_{W_{t-1}=w_{t-1},X_{t-1}^*}=L^{-1}_{V_t|w_{t-1},X_{t-1}^*}f_{V_t,W_{t-1}=w_{t-1}}\end{aligned}\tag{46}$$

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

Proof of 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 $f_{V_{t-1},W_{t-2}}$. □

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, which is useful for what follows. 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 $f_{X|Z}(x|z)=f_\epsilon(x-z)$. We define the two operators

$$\begin{aligned}&\left(L_{X|Z}h\right)(x)=\int f_\epsilon(x-z)h(z)dz\\&\left(L^*_{X|Z}h\right)(z)=\int f_\epsilon(x-z)h(x)dx.\end{aligned}\tag{47}$$

Notice that $L^*_{X|Z}$ maps functions of X to those of Z.

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 of Lemma 4. We have

$$\begin{aligned}g(x)&\equiv\left(L_{X|Z}h\right)(x)\\&=\int f_\epsilon(x-z)h(z)dz.\end{aligned}\tag{48}$$

Let ϕ_g denote the Fourier transform of g, and ϕ_ϵ that of f_ϵ. We have for any $t\in\mathbb{R}$

$$\phi_g(t)=\phi_\epsilon(t)\phi_h(t).\tag{49}$$

Therefore, $\phi_g=0$ implies $\phi_h=0$ if $\phi_\epsilon(t)\neq0$ for any $t\in\mathbb{R}$, which is assumed by hypothesis. So $L_{X|Z}$ is injective.

Next, we show the injectivity of $L^*_{X|Z}$. We consider

$$\begin{aligned}\varphi(z)&\equiv\left(L^*_{X|Z}\psi\right)(z)\\&=\int f_\epsilon(x-z)\psi(x)dx\\&\equiv\int\kappa(z-x)\psi(x)dx\end{aligned}\tag{50}$$

where $\kappa(x)\equiv f_\epsilon(-x)$, i.e., $\phi_\kappa(t)=\phi_\epsilon(-t)$. We then have

$$\begin{aligned}\phi_\varphi(t)&=\phi_\kappa(t)\phi_\psi(t)\\&=\phi_\epsilon(-t)\phi_\psi(t).\end{aligned}\tag{51}$$

Again, $\phi_\varphi=0$ implies $\phi_\psi=0$ because $\phi_\epsilon(t)\neq0$ for any $t\in\mathbb{R}$. Thus, $L^*_{X|Z}$ is injective. □

Given this lemma, we proceed to prove the two claims from Example 2.

---

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

30 When w_t (resp. w_{t-1}) is close to $\overline{w}_t$ (resp. $\overline{w}_{t-1}$), Eq. (45) 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 $\frac{\partial^3}{\partial z_t\partial z_{t-1}\partial x_t^*}\log f_{W_t|W_{t-1},X_t^*}$ is continuous and nonzero, which implies that $\frac{\partial^2}{\partial z_t\partial z_{t-1}}\log f_{W_t|W_{t-1},X_t^*}$ is monotonic in x_t^* for any (w_t,w_{t-1}), where z_t is the continuous component of w_t.

Please cite this article in press as: Hu, Y., Shum, M., Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics (2012), doi:10.1016/j.jeconom.2012.05.023

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

$$\begin{aligned} f_{X_2^*,w_2,M_1} &= \iint f_{X_2^*,y_2,m_2,X_1^*,Y_1,M_1} dy_1 dx_1^* \\ &= f_{y_2|m_2,X_2^*} f_{m_2|X_2^*,M_1} \iint f_{X_2^*|Y_1,X_1^*} \\ &\quad \times f_{Y_1|X_1^*,M_1} f_{X_1^*,M_1} dy_1 dx_1^* \\ &= f_{y_2|m_2,X_2^*} f_{m_2|X_2^*,M_1} \int f_{X_2^*|X_1^*} \left(\int f_{Y_1|X_1^*,M_1} dy_1 \right) \\ &\quad \times f_{X_1^*,M_1} dx_1^* \\ &= f_{y_2|m_2,X_2^*} f_{m_2|X_2^*,M_1} \left(\int f_{X_2^*|X_1^*} f_{X_1^*} dx_1^* \right) f_{M_1} \\ &= f_{y_2|m_2,X_2^*} f_{X_2^*} f_{m_2|X_2^*,M_1} f_{M_1}. \end{aligned} \tag{52}$$

In the third line, we have utilized the restriction that $\psi(\cdot) = 0$ in Eq. (17) so that the density of $f_{Y_1|X_1^*,M_1}$ can be integrated out. The fourth line applies the independence of (X_1^*, M_1) so that $f_{X_1^*,M_1} = f_{X_1^*} f_{M_1}$. The corresponding operator equation is

$$L_{X_2^*,w_2,M_1} = D_{y_2|m_2,X_2^*} D_{X_2^*} L_{m_2|X_2^*,M_1} D_{M_1}. \tag{53}$$

Given that all the densities in the diagonal operators are nonzero and bounded, it remains to show the injectivity of $L_{m_2|X_2^*,M_1}$. For a fixed m_2, we have:

$$\begin{aligned} g\left(x_2^*\right) &\equiv \left(L_{m_2|X_2^*,M_1} h\right)\left(x_2^*\right) \\ &= \int_0^{m_2} f_{m_2|X_2^*,M_1}\left(m_2|x_2^*, m_1\right) h(m_1) dm_1 \\ &= \int_0^{m_2} \frac{1}{m_2 - m_1} f_{\eta_2}\left(\log\left(\frac{m_2 - m_1}{m_1}\right) - x_2^*\right) h(m_1) dm_1 \\ &= \int_0^{m_2} \frac{1}{m_2 - m_1} \left(\frac{-m_2}{(m_2 - m_1) m_1}\right)^{-1} \\ &\quad \times f_{\eta_2}\left(\log\left(\frac{m_2 - m_1}{m_1}\right) - x_2^*\right) h(m_1) d\log\left(\frac{m_2 - m_1}{m_1}\right) \\ &= \int_{m_2}^{0} \frac{m_1}{m_2} f_{\eta_2}\left(\log\left(\frac{m_2 - m_1}{m_1}\right) - x_2^*\right) \\ &\quad \times h(m_1) d\log\left(\frac{m_2 - m_1}{m_1}\right) \\ &= \int_{-\infty}^{\infty} f_{\eta_2}\left(\varphi_2 - x_2^*\right) h\left(\frac{m_2}{e^{\varphi_2} + 1}\right) \frac{1}{e^{\varphi_2} + 1} d\varphi_2, \\ &\quad \left[\varphi_2 \equiv \log\left(\frac{m_2 - m_1}{m_1}\right)\right] \\ &\equiv \int_{-\infty}^{\infty} f_{\eta_2}\left(\varphi_2 - x_2^*\right) \widetilde{h}\left(\varphi_2\right) d\varphi_2, \\ &\quad \left[\widetilde{h}\left(\varphi_2\right) \equiv h\left(\frac{m_2}{e^{\varphi_2} + 1}\right) \frac{1}{e^{\varphi_2} + 1}\right] \\ &= \left(L^*_{\varphi_2|X_2^*} \widetilde{h}\right)\left(x_2^*\right), \end{aligned} \tag{54}$$

where the operator $L^*_{\varphi_2|X_2^*}$ is defined analogously to Eq. (47). As shown above, $g\left(x_2^*\right) = 0$ for any $x_2^* \in \mathbb{R}$ implies that $(L^*_{\varphi_2|X_2^*}\widetilde{h})$ $(x_2^*) = 0$ for any $x_2^* \in \mathbb{R}$, where the kernel of $L^*_{\varphi_2|X_2^*}$ has a convolution form $f_{\eta_2}\left(\varphi_2 - x_2^*\right)$. Since the characteristic function of η_2 has no zeros on the real line, we can apply Lemma 4 to obtain the injectivity of $L^*_{\varphi_2|X_2^*}$. Accordingly, $\left(L^*_{\varphi_2|X_2^*}\widetilde{h}\right)\left(x_2^*\right) = 0$ for any $x_2^* \in \mathbb{R}$ implies $\widetilde{h}\left(\varphi_2\right) = 0$ for any $\varphi_2 \in \mathbb{R}$. Next, because $\widetilde{h}(\varphi_2) = h(\frac{m_2}{e^{\varphi_2}+1})\frac{1}{e^{\varphi_2}+1}$ and $\frac{1}{e^{\varphi_2}+1}$ is nonzero, $\widetilde{h}(\varphi_2) = 0$ for any $\varphi_2 \in \mathbb{R}$ implies $h(\frac{m_2}{e^{\varphi_2}+1}) = 0$ for any $\varphi_2 \in \mathbb{R}$. Given $\varphi_2 \equiv \log\left(\frac{m_2-m_1}{m_1}\right)$, we have $h\left(m_1\right) = 0$ for any $m_1 \in (0, m_2)$. Altogether, then, $g\left(x_2^*\right) = 0$ for any $x_2^* \in \mathbb{R}$ implies $h\left(m_1\right) = 0$ for any $m_1 \in (0, m_2)$, thus demonstrating the injectivity of the operator $L_{m_2|X_2^*,M_1}$, as claimed. □

Proof of Claim 2. First, we show the injectivity of L_{M_1,w_2,w_3,M_4}. For fixed (w_2, w_3):

$$\begin{aligned} &f_{M_1,w_2,w_3,M_4} \\ &= \int f_{M_4|w_3,X_3^*} f_{w_3|w_2,X_3^*} f_{X_3^*,w_2,M_1} dx_3^*. \\ &= \int \left(\int f_{M_4|w_3,X_4^*} f_{X_4^*|w_3,X_3^*} dx_4^* \right) f_{w_3|w_2,X_3^*} \\ &\quad \times \left(\int f_{X_3^*|w_2,X_2^*} f_{X_2^*,w_2,M_1} dx_2^* \right) dx_3^* \\ &= \int \left(\int f_{M_4|w_3,X_4^*} f_{X_4^*|w_3,X_3^*} dx_4^* \right) f_{w_3|w_2,X_3^*} \\ &\quad \times \left(\int f_{X_3^*|w_2,X_2^*} f_{y_2|m_2,X_2^*} f_{X_2^*} f_{m_2|X_2^*,M_1} f_{M_1} dx_2^* \right) dx_3^*. \end{aligned} \tag{55}$$

Therefore, the equivalent operator equation is

$$\begin{aligned} L_{M_1,w_2,w_3,M_4} &= L_{M_1,y_2,m_2,y_3,m_3,M_4} \\ &= D_{M_1} L^*_{m_2|X_2^*,M_1} D_{X_2^*} D_{y_2|m_2,X_2^*} L^*_{X_3^*|w_2,X_2^*} \\ &\quad \times D_{w_3|w_2,X_3^*} L^*_{X_4^*|w_3,X_3^*} L^*_{M_4|w_3,X_4^*}. \end{aligned} \tag{56}$$

In the above, the L^* operators are defined analogously to Eq. (47), and all the L^* operators are based on convolution kernels. Earlier, in the main text and Claim 1, we showed that the operators $L_{m_2|X_2^*,M_1}$, $L_{X_3^*|w_2,X_2^*}$, $L_{X_4^*|w_3,X_3^*}$, and $L_{M_4|w_3,X_4^*}$ are injective; hence, by applying Lemma 4, we also obtain the injectivity of $L^*_{m_2|X_2^*,M_1}$, $L^*_{X_3^*|w_2,X_2^*}$, $L^*_{X_4^*|w_3,X_3^*}$, and $L^*_{M_4|w_3,X_4^*}$ using an argument similar to that used in the proof of Claim 1.

Finally, all the densities corresponding to the diagonal operators in Eq. (56) are nonzero and bounded, implying that these operators are injective. Hence, L_{M_1,w_2,w_3,M_4} is also injective.

Second, for L_{M_1,w_2,M_3}, we have

$$\begin{aligned} f_{M_1,w_2,M_3} &= \int f_{M_3|w_2,X_2^*} f_{X_2^*,w_2,M_1} dx_2^* \\ &= \int \left(\int f_{M_3|w_2,X_3^*} f_{X_3^*|w_2,X_2^*} dx_3^* \right) f_{X_2^*,w_2,M_1} dx_2^* \\ &= \int \left(\int f_{M_3|w_2,X_3^*} f_{X_3^*|w_2,X_2^*} dx_3^* \right) \\ &\quad \times f_{y_2|m_2,X_2^*} f_{X_2^*} f_{m_2|X_2^*,M_1} f_{M_1} dx_2^*. \end{aligned} \tag{57}$$

Therefore, the equivalent operator equation is

$$L_{M_1,w_2,M_3} = D_{M_1} L^*_{m_2|X_2^*,M_1} D_{X_2^*} D_{y_2|m_2,X_2^*} L^*_{X_3^*|w_2,X_2^*} L^*_{M_3|w_2,X_3^*}. \tag{58}$$

By stationarity, the injectivity of $L^*_{M_3|w_2,X_3^*}$ is implied by that of $L^*_{M_4|w_3,X_4^*}$. All the other operators on the RHS also appeared in Eq. (56), and we argued above that these were injective. Thus, L_{M_1,w_2,M_3} is injective. □

Appendix C. Miscellaneous remarks

C.1. Further discussion on Assumption 2

In this section, we discuss how Assumption 2 is used to ensure the validity of two different ways for taking operator inverses.

Please cite this article in press as: Hu, Y., Shum, M., Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics (2012), doi:10.1016/j.jeconom.2012.05.023

Consider two scenarios involving an operator equation

$$L_{R_1,r_2,R_4} = L_{R_1|r_2,R_3} L_{r_2,R_3,R_4}. \tag{59}$$

In the first scenario, suppose we want to solve for L_{r_2,R_3,R_4} given L_{R_1,r_2,R_4} and $L_{R_1|r_2,R_3}$. The assumption that $L_{R_1|r_2,R_3}$ is one-to-one guarantees that we may have

$$L^{-1}_{R_1|r_2,R_3} L_{R_1,r_2,R_4} = L_{r_2,R_3,R_4}. \tag{60}$$

As an example, Assumption 2(ii) guarantees that pre-multiplication by the inverse operator $L_{V_{t+1}|w_t,X_t^*}$ is valid, which is used in the equation following Eq. (9).

In the second scenario, suppose we need to solve for $L_{R_1|r_2,R_3}$ given L_{R_1,r_2,R_4} and L_{r_2,R_3,R_4} in Eq. (59). We would need the operator L_{r_2,R_3,R_4} to be invertible as follows:

$$L_{R_1,r_2,R_4} L^{-1}_{r_2,R_3,R_4} = L_{R_1|r_2,R_3}. \tag{61}$$

As proved in Lemma 1 in Hu and Schennach (2008), the sufficient condition for obtaining Eq. (61) from Eq. (59) is that the operator L_{R_4,R_3,r_2} is one-to-one.[31] (Notice that the operator L_{R_4,R_3,r_2} is from $L^p(\mathcal{R}_3)$ to $L^p(\mathcal{R}_4)$.)

Assumption 2(i) is an example of this. It is used to justify the post-multiplication by $L^{-1}_{V_{t+1},\bar{w}_t,w_{t-1},V_{t-2}}$ and $L^{-1}_{V_{t+1},w_t,\bar{w}_{t-1},V_{t-2}}$ in, respectively, Eqs. (9) and (10). Similarly, Assumption 2(iii) guarantees the validity of post-multiplication by $L^{-1}_{V_t,w_{t-1},V_{t-2}}$, which is done in the second line in Eq. (29). Throughout this paper, we only post-multiply by the inverses of $L_{V_{t+1},w_t,w_{t-1},V_{t-2}}$ and $L_{V_t,w_{t-1},V_{t-2}}$; all other cases of inverses involve pre-multiplication. For a more technical discussion, see Aubin (2000, Sections 4.5–4.6).

C.2. Necessary conditions for completeness

As the discussion of the dynamic investment model has illustrated, the functional forms of the operators for which we can verify completeness are restrictive. But in those examples we have focused on providing sufficient conditions; those conditions, while restrictive, may be far from necessary.

To show this, in this section we provide a necessary condition for completeness.

Lemma 5 (*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*

$$\frac{\partial}{\partial r_3} f_{R_1|R_3}(r_1|r_3) \neq 0 \quad \text{almost surely for } \forall r_1 \in \mathcal{S}_1,\ \forall r_3 \in \mathcal{S}_3. \tag{62}$$

Proof of Lemma 5. Suppose Eq. (62) fails, so there exists an interval $\mathcal{S}_3 \equiv [\underline{r}, \overline{r}]$ 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

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

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 $G(\overline{r}) - G(\underline{r}) = \int_{\underline{r}}^{\overline{r}} g(r) dr = 0$, but $g(r) \neq 0$ for any r in a nontrivial subset of $[\underline{r}, \overline{r}]$. We have $L_{R_1|R_3} h_0 = 0$, but $h_0 \neq 0$. Therefore, Eq. (10) fails, and $L_{R_1|R_3}$ is not one-to-one. □

Intuitively, the necessary condition here ensures that there is enough variation in the distribution of R_1 for different values of R_3. Note that this allows for great flexibility in specifying the functional forms of the operators, and may even allow for some "semiparametric" specifications, in which, for instance, the laws of motions for the state variable are specified as flexible polynomials.

[31] A similar assumption is also used in Carroll et al. (2010).

References

Abbring, J., Chiappori, P., Zavadil, T., 2008. Better safe than sorry? ex ante and ex post moral hazard in dynamic insurance data. Tilburg Univeristy. Working Paper.
Ackerberg, D., Benkard, L., Berry, S., Pakes, A., 2007. Econometric tools for analyzing market outcomes. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics Vol. 6A. North-Holland.
Aguirregabiria, V., Mira, P., 2002. Swapping the nested fixed point algorithm: a class of estimators for discrete Markov decision models. Econometrica 70, 1519–1543.
Aguirregabiria, V., Mira, P., 2007. Sequential estimation of dynamic discrete games. Econometrica 75, 1–53.
Arcidiacono, P., Miller, R., 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., Honoré, B., 2000. Panel data models: some recent developments. In: Handbook of Econometrics, Vol. 5. North-Holland.
Aubin, J.-P., 2000. Applied Functional Analysis. Wiley-Interscience.
Bajari, P., Benkard, L., Levin, J., 2007a. Estimating dynamic models of imperfect competition. Econometrica 75, 1331–1370.
Bajari, P., Chernozhukov, V., Hong, H., Nekipelov, D., 2007b. Nonparametric and semiparametric analysis of a dynamic game model. Manuscript. University of Minnesota.
Blundell, R., Chen, X., Kristensen, D., 2007. Nonparametric IV estimation of shape-invariant engel curves. Econometrica 75, 1613–1669.
Bouissou, M., Laffont, J.J., Vuong, Q., 1986. Tests of noncausality under Markov assumptions for qualitative panel data. Econometrica 54, 395–414.
Buchinsky, M., Hahn, J., Hotz, J., 2004. Estimating dynamic discrete choice models with heterogeneous agents: a cluster analysis approach. Working Paper. UCLA.
Carroll, R., Chen, X., Hu, Y., 2010. Identification and estimation of nonlinear models using two samples with nonclassical measurement errors. Journal of Nonparametric Statistics 22, 379–399.
Chen, X., Hong, H., Nekipelov, D., 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., Shum, M., 2005. Uncertainty and learning in pharmaceutical demand. Econometrica 73, 1137–1174.
Cunha, F., Heckman, J., Schennach, S., 2006. Estimating the technology of cognitive and noncognitive skill formation. Manuscript. University of Chicago.
Das, S., Roberts, M., Tybout, J., 2007. Market entry costs, producer heterogeneity, and export dynamics. Econometrica 75, 837–874.
d'Haultfoeuille, X., 2011. On the completeness condition in nonparametric instrumental problems. Econometric Theory 27, 460–471.
Doraszelski, U., Pakes, A., 2007. A framework for dynamic analysis in IO. In: Armstrong, M., Porter, R. (Eds.), Handbook of Industrial Organization, Vol. 3. North-Holland (Chapter 30).
Dunford, N., Schwartz, J., 1971. Linear Operators, Vol. 3. Wiley.
Dunne, T., Klimer, S., Roberts, M., Xu, D., 2006. Entry and exit in geographic markets. Manuscript. Penn State University.
Erdem, T., Imai, S., Keane, M., 2003. Brand and quantity choice dynamics under price uncertainty. Quantitative Marketing and Economics 1, 5–64.
Ericson, R., Pakes, A., 1995. Markov-perfect industry dynamics: a framework for empirical work. Review of Economic Studies 62, 53–82.
Fernandez-Villaverde, J., Rubio-Ramirez, J., 2007. Estimating macroeconomic models: a likelihood approach. Review of Economic Studies 74, 1059–1087.
Hendel, I., Nevo, A., 2006. Measuring the implications of sales and consumer stockpiling behavior. Econometrica 74, 1637–1674.
Henry, M., Kitamura, Y., Salanie, B., 2008. Nonparametric identification of mixtures with exclusion restrictions. Working Paper. University of Montreal.
Hong, H., Shum, M., 2009. Pairwise-difference estimation of a dynamic optimization model. Review of Economic Studies 77, 273–304.
Hotz, J., Miller, R., 1993. Conditional choice probabilties and the estimation of dynamic models. Review of Economic Studies 60, 497–529.

Please cite this article in press as: Hu, Y., Shum, M., Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics (2012), doi:10.1016/j.jeconom.2012.05.023

Hotz, J., Miller, R., Sanders, S., Smith, J., 1994. A simulation estimator for dynamic models of discrete choice. Review of Economic Studies 61, 265–289.

Houde, J.-F., Imai, S., 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., Schennach, S., 2008. Instrumental variable treatment of nonclassical measurement error models. Econometrica 76, 195–216.

Hu, Y., Shum, M., 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., Jain, N., Ching, A., 2009. Bayesian estimation of dynamic discrete choice models. Econometrica 77, 1865–1899.

Kasahara, H., Shimotsu, K., 2009. Nonparametric identification of finite mixture models of dynamic discrete choice. Econometrica 77, 135–175.

Levinsohn, J., Petrin, A., 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., Thesmar, D., 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., Powell, J., 2003. Instrumental variable estimation of nonparametric models. Econometrica 71, 1565–1578.

Norets, A., 2009. Inference in dynamic discrete choice models with serially correlated unobserved state variables. Econometrica 77, 1665–1682.

Olley, S., Pakes, A., 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), 755–784.

Pakes, A., Ostrovsky, M., Berry, S., 2007. Simple estimators for the parameters of discrete dynamic games (with entry exit examples). Rand Journal of Economics 38, 373–399.

Pakes, A., Simpson, M., 1989. Patent renewal data. In: Brookings Papers on Economic Activity: Microeconomics. pp. 331–403.

Pesendorfer, M., Schmidt-Dengler, P., 2008. Asymptotic least squares estimators for dynamic games. Review of Economic Studies 75, 901–928.

Rust, J., 1994. Structural estimation of Markov decision processes. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. 4. North Holland, pp. 3082–3146.

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.

Schennach, S., 2007. Instrumental variable estimation of nonlinear errors-in-variables models. Econometrica 75, 201–223.

Shiu, J., Hu, Y., Identification and Estimation of nonlinear dynamic panel data models with unobserved covariates", working paper 557, Johns Hopkins University, 2010.

Xu, D., 2007. A structural empirical model of R&D, firm heterogeneity, and industry evolution. Manuscript. New York University.

Please cite this article in press as: Hu, Y., Shum, M., Nonparametric identification of dynamic models with unobserved state variables. Journal of Econometrics (2012), doi:10.1016/j.jeconom.2012.05.023