

Nonparametric Identification of Dynamic Models with Unobserved State Variables*

Yingyao Hu
Johns Hopkins University

Matthew Shum
California Institute of Technology

September 2011

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}, \dots, W_{t-3} . Second, in the

*The authors can be reached at yhu@jhu.edu and mshum@caltech.edu. We thank Xiaohong Chen, Jeremy Fox, Han Hong, Ariel Pakes, and Susanne Schennach for their suggestions. Seminar participants at Berkeley-Haas, BU, Clark, UC-Davis, Harvard, LSE, MIT, NYU, Penn, Penn State, Toulouse, UCL, UCLA, USC, the Cowles 2008 Summer Conference at Yale, the 2008 ERID Conference at Duke, the 2008 Greater New York Econometrics Colloquium at Princeton, the “Econometrics of Industrial Organization” workshop at Toulouse, and SITE 2009 provided useful comments.

stationary case, where $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ is the same across all t , only four observations W_{t+1}, \dots, W_{t-2} , for some t , are required for identification.

In most applications, W_t consists of two components $W_t = (Y_t, M_t)$, where Y_t denotes the agent's action in period t , and M_t denotes the period- t observed state variable(s). X_t^* are time-varying unobserved state variables (USV for short), which are observed by agents and affect their choice of Y_t , but unobserved by the econometrician. The economic importance of models with unobserved state variables has been recognized since the earliest papers on the structural estimation of dynamic optimization models. Two examples are:

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

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

These two early papers demonstrated that dynamic optimization problems with an unobserved process partly determining the state variables are indeed empirically tractable. Their authors (cf. Miller (1984, section V); Pakes and Simpson (1989)) also provided some discussion of the restrictions implied on the data by their models, thus highlighting how identification has been a concern since the earliest structural empirical applications of dynamic models with 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

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

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

Once the CCP’s and the law of motion for the state variables are recovered, it is straightforward to use them as inputs in a CCP-based approach for estimating dynamic discrete-choice models. This approach was pioneered in Hotz and Miller (1993) and Hotz, Miller, Sanders, and Smith (1994).¹ A general criticism of these methods is that they cannot accommodate unobserved state variables. In response, Aguirregabiria and Mira (2007), Buchinsky, Hahn, and Hotz (2004), and Houde and Imai (2006), among others, recently developed CCP-based estimation methodologies allowing for agent-specific unobserved heterogeneity, which is the special case where the latent X_t^* is time-invariant. 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 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, Chen, and Hu (2009) for nonclassical measurement models (where the measurement error is not assumed to be independent of the latent “true” variable) can be applied to Markovian dynamic models, and we use those results to establish nonparametric identification.

Our results extend nonparametric identification to classes of models not covered in the

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

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

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, Kitamura, and Salanie (2008, hereafter HKS) exploit exclusion restrictions to identify Markov regime-switching models with a discrete and latent state variable. While our identification arguments are quite distinct from those in HKS, 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, Heckman, and Schennach (2006) apply the result of Hu and Schennach (2008) to show nonparametric identification of a nonlinear factor model consisting of $(W_t, W_t', W_t'', X_t^*)$, where the observed processes $\{W_t\}_{t=1}^T$, $\{W_t'\}_{t=1}^T$, and $\{W_t''\}_{t=1}^T$ constitute noisy measurements of the latent process $\{X_t^*\}_{t=1}^T$, contaminated with random disturbances. In contrast, we consider a setting where (W_t, X_t^*) jointly evolves as a dynamic Markov process. We use observations of W_t in different periods t to identify the conditional density of $(W_t, X_t^* | W_{t-1}, X_{t-1}^*)$. 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,

³That 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 below.

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

the appendix contains the proofs for several lemmas and corollaries.

2 Overview of assumptions

We assume that for each agent i , $\{(W_T, X_T^*), \dots, (W_t, X_t^*), \dots, (W_1, X_1^*)\}_i$ is an independent random draw from a bounded continuous distribution $f_{(W_T, X_T^*), \dots, (W_t, X_t^*), \dots, (W_1, X_1^*)}$. The researcher observes a panel dataset consisting of an i.i.d. random sample of $\{W_T, W_{T-1}, \dots, W_1\}_i$, with $T \geq 5$, for many agents i . We first consider identification in the nonstationary case, where the Markov law of motion $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ varies across periods. 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 $\Omega_{<t-1} \equiv \{W_{t-2}, \dots, W_1, X_{t-2}^*, \dots, X_1^*\}$, the history up to (but not including) $t-1$.
(ii) Limited feedback: $f_{W_t | W_{t-1}, X_t^*, X_{t-1}^*} = f_{W_t | W_{t-1}, X_t^*}$.

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 $g_t : \mathbb{R}^d \rightarrow \mathbb{R}$ 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, Chen, and Hu (2009, 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, Chen, and Hu (2009).

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

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

⁶Most 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, Imai, and Keane (2003), Crawford and Shum (2005), Das, Roberts, and Tybout (2007), Xu (2007), and Hendel and Nevo (2006).

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

variables (X, Y, Z, X^*) satisfy

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

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, Chen, and Hu (2009) 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

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

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) don't directly apply to this framework, Carroll, Chen, and Hu (2009) 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

$$\begin{aligned} f_{W_{t+1}, W_t, W_{t-1}, W_{t-2}} &= \int f_{W_{t+1}|W_t, X_t^*} f_{W_t|W_{t-1}, X_t^*} f_{X_t^*, W_{t-1}, W_{t-2}} dx_t^* \\ &= \int f_{W_{t+1}|W_t, X_t^*} f_{W_t, W_{t-1}, X_t^*} f_{W_{t-2}|X_t^*, W_{t-1}} dx_t^*. \end{aligned} \quad (7)$$

Comparing equations 6 and 7, we may use the identification strategy in Carroll, Chen, and Hu (2009) with $(W_{t+1}, W_t, W_{t-1}, W_{t-2}, X_t^*)$ 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, Chen, and Hu (2009).⁸

⁸Shiu 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.

Denote the supports of V_t and W_t as \mathcal{V}_t and \mathcal{W}_t , respectively.⁹ The linear operator $L_{V_{t-2}, \bar{w}_{t-1}, \bar{w}_t, V_{t+1}}$ 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 defined as¹¹

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

Similarly, we define the diagonal (or multiplication) operator

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

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 exists 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 $(\bar{w}_t, \bar{w}_{t-1}) \in \mathcal{N}^r$, $L_{V_{t-2}, \bar{w}_{t-1}, \bar{w}_t, V_{t+1}}$ 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

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) needs three periods of data in the comparable setting. Their assumptions are tailored to be suitable for panel data models. Our framework are more suitable for IO models, where the conditional independence assumptions are based on empirical IO models.

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

¹⁰For $1 \leq p < \infty$, $\mathcal{L}^p(\mathcal{X})$ is the space of measurable real functions $h(\cdot)$ integrable in the L^p -norm, ie. $\int_{\mathcal{X}} |h(x)|^p d\mu(x) < \infty$, where μ is a measure on a σ -field in \mathcal{X} . One may also consider other classes of functions, such as bounded functions in \mathcal{L}^1 , in the definition of an operator.

¹¹Analogously, 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 $h \in \mathcal{L}^p(\mathcal{X}_t^*)$, and $w_t \in \mathcal{W}_t$ as $(L_{V_{t+1} | w_t, X_t^*} h)(v_{t+1}) = \int f_{V_{t+1} | w_t, X_t^*}(v_{t+1} | w_t, x_t^*) h(x_t^*) dx_t^*$.

¹²Specifically, 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.

¹³A neighborhood of $w \in \mathbb{R}^k$ is defined as $\{\bar{w} \in \mathbb{R}^k : \|\bar{w} - w\|_E < r\}$ for some $r > 0$, where $\|\cdot\|_E$ is the Euclidean metric.

V_{t-2} and V_{t+1} as noisy “measurements” of the latent X_t^* , Assumption 2(i,ii) imposes the same restrictions between the measurements and the latent variable as Hu and Schennach (2008, Assumption 3) and Carroll, Chen, and Hu (2009, Assumption 2.4). Compared with these two papers, Assumption 2(iii) is an extra assumption we need because, in our dynamic setting, there is a second latent variable, X_{t-1}^* , in the Markov law of motion $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$. Below, we show that Assumption 2(ii) implies that pre-multiplication by the inverse operator $L_{V_{t+1} | w_t, X_t^*}^{-1}$ is valid, while 2(i,iii) imply that post-multiplication by, respectively, $L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}}^{-1}$ and $L_{V_t, w_{t-1}, V_{t-2}}^{-1}$ 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 $L_{V_{t-2}, \bar{w}_{t-1}, \bar{w}_t, V_{t+1}}$ 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 $(\bar{w}_{t-1}, \bar{w}_t)$,

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

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

The variable 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 $\bar{x}_t^* \neq \tilde{x}_t^* \in \mathcal{X}_t^*$, 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 $(\bar{w}_t, \bar{w}_{t-1}) \in \mathcal{N}^r$ with $\bar{w}_t \neq w_t$, $\bar{w}_{t-1} \neq w_{t-1}$:*

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

¹⁵There 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.

- (i) $0 < k(w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, x_t^*) < C < \infty$ for any $x_t^* \in \mathcal{X}_t^*$ and some constant C ;
(ii) $k(w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, \tilde{x}_t^*) \neq k(w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, x_t^*)$, where

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

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. (51) below shows, the $k(\dots)$ 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, 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.

¹⁶In 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. (51) below implies that the CCP term cancels out in the expression of eigenvalues in the spectral decomposition, so that Assumption 3 imposes restrictions only on the second term $f_{M_t|Y_{t-1}, M_{t-1}, X_t^*}$. See additional discussion in Example 2 below.

¹⁷To be clear, the monotonicity assumption here is a model restriction, and not without loss of generality; if it were false, our identification argument would not recover the correct CCP’s and laws of motion for the underlying model. See Matzkin (2003) and Hu and Schennach (2008) for similar uses of monotonicity restrictions in the context of nonparametric identification problems.

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 $t \in \{3, \dots, T-1\}$, Assumption 1 implies that, for any $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$,

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

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

$$L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*} = L_{V_{t+1} | w_t, X_t^*}^{-1} L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} L_{V_t, w_{t-1}, V_{t-2}}^{-1} L_{V_t | w_{t-1}, X_{t-1}^*}. \quad (13)$$

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 $t \in \{3, \dots, T-1\}$, Assumptions 1, 2, 3, 4 imply that the density $f_{V_{t+1}, W_t, W_{t-1}, V_{t-2}}$ uniquely 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 the Assumptions 1, 2, 3, and 4, the density $f_{W_{t+1}, W_t, W_{t-1}, W_{t-2}, W_{t-3}}$ for any $t \in \{4, \dots, T-1\}$ uniquely determines the density $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$.

¹⁸Recall 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.

3.1 Initial conditions

Some CCP-based estimation methodologies for dynamic optimization models (eg. Hotz, Miller, Sanders, and Smith (1994), Bajari, Benkard, and Levin (2007)) require simulation of the Markov process $(W_t, X_t^*, W_{t+1}, X_{t+1}^*, W_{t+2}, X_{t+2}^*, \dots)$ starting from some initial values W_{t-1}, X_{t-1}^* . When there are unobserved state variables, this raises difficulties because X_{t-1}^* is not observed. However, it turns out that, as a by-product of the main identification results, we are also able to identify the marginal densities f_{W_{t-1}, X_{t-1}^*} . For any given initial value of the observed variables w_{t-1} , knowledge of f_{W_{t-1}, X_{t-1}^*} allows us to draw an initial value of X_{t-1}^* consistent with w_{t-1} .

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

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

Stationarity is usually maintained in infinite-horizon dynamic programming models. Given the foregoing discussion, we present the next corollary without proof.

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

In the stationary case, initial conditions are still a concern. The following corollary, analogous to Corollary 1 for the non-stationary case, postulates the identification of the

marginal density f_{W_t, X_t^*} , for periods $t \in \{1, \dots, T-3\}$. For any of these periods, f_{W_t, X_t^*} can be used as a sampling density for the initial conditions.¹⁹

Corollary 3. (Identification of initial conditions, stationary case): *Under assumptions 1, 2, 3, 4, and 5, the observed density $f_{W_{t+1}, W_t, W_{t-1}, W_{t-2}}$ for any $t \in \{3, \dots, T-1\}$ 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 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.²⁰

4.1 Example 1: A discrete model

As a first example, let (W_t, X_t^*) denote a bivariate discrete first-order Markov process where W_t and X_t^* are both binary scalars: $\forall t, \text{supp}X_t^* = \text{supp}W_t \equiv \{0, 1\}$. This is the simplest example of the models considered in our framework. One example of such a model is a binary version of Abbring, Chiappori, and Zavadil's (2008) "dynamic moral hazard" model of auto insurance. In that model, W_t is a binary indicator of claims occurrence, and X_t^* is a binary effort indicator, with $X_t^* = 1$ denoting higher effort. In this model, moral hazard in driving behavior and experience rating in insurance pricing imply that the laws of motion

¹⁹Even 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.

²⁰A third example, based on Rust (1987), is in Appendix C.

for both W_t and X_t^* should exhibit state dependence:

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

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.²¹

Relative to the continuous case presented beforehand, some simplifications obtain in this finite-dimensional example. Notationally, the linear operators in the previous section reduce to matrices, with the L operators in the main proof corresponding to 2×2 square matrices, and the D operators to 2×2 diagonal matrices. 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 $(R_1 = i, R_2 = r_2, R_3 = j)$, for $i, j \in \{0, 1\}$.

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

$$\begin{aligned} L_{W_{t+1}, w_t, w_{t-1}, W_{t-2}} &= L_{W_{t+1}|w_t, X_t^*} D_{w_t|w_{t-1}, X_t^*} L_{X_t^*, w_{t-1}, W_{t-2}} \\ &= L_{W_{t+1}|w_t, X_t^*} D_{w_t|w_{t-1}, X_t^*} L_{X_t^*|w_{t-1}, X_{t-1}^*} L_{X_{t-1}^*, w_{t-1}, W_{t-2}} \\ &= L_{W_{t+1}|w_t, X_{t+1}^*} L_{X_{t+1}^*|w_t, X_t^*} D_{w_t|w_{t-1}, X_t^*} L_{X_t^*|w_{t-1}, X_{t-1}^*} L_{X_{t-1}^*, w_{t-1}, W_{t-2}} D_{w_{t-1}, W_{t-2}} \end{aligned} \quad (15)$$

Assumption 2(i) requires that the square matrix $L_{W_{t-2}, w_{t-1}, w_t, W_{t+1}} = L_{W_{t+1}, w_t, w_{t-1}, W_{t-2}}^T$ is invertible, which implies that $L_{W_{t+1}, w_t, w_{t-1}, W_{t-2}}$ is also invertible. This matrix is observed in the data, so that we can verify its invertibility directly.

The invertibility of $L_{W_{t+1}, w_t, w_{t-1}, W_{t-2}}$ requires the matrices on the RHS of Eq. (16) to be invertible. Below we discuss the implication of the invertibility of each matrix. By

²¹In section 2 of the supplemental material in Appendix C, we provide a more detailed discussion.

definition, we have

$$L_{W_{t+1}|w_t, X_{t+1}^*} = \left[f_{W_{t+1}|W_t, X_{t+1}^*} (i-1|w_t, j-1) \right]_{i,j} = \begin{bmatrix} 1 - p(w_t, 0) & 1 - p(w_t, 1) \\ p(w_t, 0) & p(w_t, 1) \end{bmatrix}. \quad (17)$$

The invertibility of $L_{W_{t+1}|w_t, X_{t+1}^*}$ requires $\det \left(L_{W_{t+1}|w_t, X_{t+1}^*} \right) = p(w_t, 1) - p(w_t, 0) \neq 0$. That means the effort X_{t+1}^* at period $t+1$ has impact on the claim occurrence W_{t+1} at the same period given the claim w_t at the previous period. Similarly, the invertibility of $L_{X_{t+1}^*|w_t, X_t^*}$ implies $q(0, w_t) \neq q(1, w_t)$. Therefore, given the claim w_t at period t , the effort X_t^* at period t is correlated with the effort level X_{t+1}^* at the next period. The diagonal matrix $D_{w_t|w_{t-1}, X_t^*}$ is defined as

$$\begin{aligned} D_{w_t|w_{t-1}, X_t^*} &= \text{diag} \left[f_{W_t|W_{t-1}, X_t^*} (w_t|w_{t-1}, 0), f_{W_t|W_{t-1}, X_t^*} (w_t|w_{t-1}, 1) \right] \\ &= \begin{bmatrix} [p(w_{t-1}, 0)]^{w_t} [1 - p(w_{t-1}, 0)]^{1-w_t} & 0 \\ 0 & [p(w_{t-1}, 1)]^{w_t} [1 - p(w_{t-1}, 1)]^{1-w_t} \end{bmatrix}. \end{aligned} \quad (18)$$

The invertibility of $D_{w_t|w_{t-1}, X_t^*}$ simply requires $0 < p(w_{t-1}, x_t^*) < 1$. That implies regardless of the effort level, the risk still exists in the sense that the claim may still occur with a positive probability but not with probability one. The implication of the invertibility of $L_{X_t^*|w_{t-1}, X_{t-1}^*}$ is the same as that of $L_{X_{t+1}^*|w_t, X_t^*}$. The invertibility of $L_{X_{t-1}^*|w_{t-1}, W_{t-2}}$ imposes restrictions on the initial conditions. We have

$$\begin{aligned} L_{X_{t-1}^*|w_{t-1}, W_{t-2}} &= \left[f_{X_{t-1}^*|W_{t-1}, W_{t-2}} (i-1|w_{t-1}, j-1) \right]_{i,j} \\ &= \begin{bmatrix} 1 - f_{X_{t-1}^*|W_{t-1}, W_{t-2}} (1|w_{t-1}, 0) & 1 - f_{X_{t-1}^*|W_{t-1}, W_{t-2}} (1|w_{t-1}, 1) \\ f_{X_{t-1}^*|W_{t-1}, W_{t-2}} (1|w_{t-1}, 0) & f_{X_{t-1}^*|W_{t-1}, W_{t-2}} (1|w_{t-1}, 1) \end{bmatrix} \end{aligned} \quad (19)$$

The invertibility of $L_{X_{t-1}^*|w_{t-1}, W_{t-2}}$ simply requires $f_{X_{t-1}^*|W_{t-1}, W_{t-2}} (1|w_{t-1}, 0) \neq f_{X_{t-1}^*|W_{t-1}, W_{t-2}} (1|w_{t-1}, 1)$. That means, as part of the initial conditions, given the claim w_t at period t , different claim occurrence at period $t-2$ implies a different level of effort X_{t-1}^* at period $t-1$. The last matrix is

$$\begin{aligned} D_{w_{t-1}, W_{t-2}} &= \text{diag} \left[f_{W_{t-1}, W_{t-2}} (w_{t-1}, 0), f_{W_{t-1}, W_{t-2}} (w_{t-1}, 1) \right] \\ &= \begin{bmatrix} f_{W_{t-1}, W_{t-2}} (w_{t-1}, 0) & 0 \\ 0 & f_{W_{t-1}, W_{t-2}} (w_{t-1}, 1) \end{bmatrix}. \end{aligned} \quad (20)$$

The invertibility of $D_{w_{t-1}, W_{t-2}}$ requires $f_{W_{t-1}, W_{t-2}} (w_{t-1}, w_{t-2}) > 0$, which means all possi-

ble cases are observed in the sample at the initial periods.

Moreover, by Eq. (15), the invertibility of $L_{W_{t+1}, w_t, w_{t-1}, W_{t-2}}$ also implies the invertibility of $L_{W_{t+1}|w_t, X_t^*}$, $L_{X_t^*|w_{t-1}, X_{t-1}^*}$, and $L_{X_{t-1}^*, w_{t-1}, W_{t-2}}$, 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 implied by 2(i). Assumption 2(iii) is also implied by 2(i). Specifically, $L_{W_{t-2}, w_{t-1}, W_t} = L_{W_t, w_{t-1}, W_{t-2}}^T$ with

$$L_{W_t, w_{t-1}, W_{t-2}} = L_{W_t|w_{t-1}, X_{t-1}^*} L_{X_{t-1}^*, w_{t-1}, W_{t-2}}. \quad (21)$$

By Assumption 2(i), $L_{W_t, w_{t-1}, w_{t-2}, W_{t-3}}$ is invertible, which implies that $L_{W_t|w_{t-1}, X_{t-1}^*}$ is invertible. The invertibility of $L_{X_{t-1}^*, w_{t-1}, W_{t-2}}$ was shown above to be implied by that of $L_{W_{t+1}, w_t, w_{t-1}, W_{t-2}}$. Thus, the matrices $L_{W_t, w_{t-1}, W_{t-2}}$ and $L_{W_{t-2}, w_{t-1}, W_t}$ are both invertible. Therefore, in discrete models, Assumption 2 collapses to condition 2(i), which is testable from the observed data.

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

In this discrete case, Assumption 4 is an ‘‘ordering’’ assumption on the columns of the $L_{W_{t+1}|w_t, X_t^*}$ matrix, which are the eigenvectors of \mathbf{AB} . This is because, for a matrix diagonalization $T = SDS^{-1}$, where D is diagonal, and T and S are square matrices, any permutation of the eigenvalues (the diagonal elements in D) and their corresponding eigenvectors (the columns in S) results in the same diagonal representation of T .

If the goal is only to identify $f_{W_t, X_t^*|W_{t-1}, X_{t-1}^*}$ for a single period t , then we could dispense with Assumption 4 altogether, and pick two arbitrary orderings in recovering $L_{W_{t+1}|w_t, X_t^*}$ and $L_{W_t|w_{t-1}, X_{t-1}^*}$. By doing this, we cannot pin down the exact value of X_t^* or X_{t-1}^* , but the recovered density of $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ is still consistent with the two arbitrary orderings for X_t^* and X_{t-1}^* , in the sense that the implied transition matrix $X_t^*|w_{t-1}, X_{t-1}^*$ for every $w_{t-1} \in \mathcal{W}_{t-1}$ is consistent with the true, but unknown ordering of X_t^* and X_{t-1}^* .²²

But this will not suffice if we wish to recover the transition density $f_{W_t, X_t^*|W_{t-1}, X_{t-1}^*}$ in two periods $t = t_1, t_2$, with $t_1 \neq t_2$. If we want to compare values of X_t^* across these two periods, then we must invoke Assumption 4 to pin down values of X_t^* which are consistent

²²We thank Thierry Magnac for this insight.

across the two periods. For this example, one reasonable monotonicity restriction is

$$\text{for } w_t = \{0, 1\} : \mathbb{E}[W_{t+1}|w_t, X_t^* = 1] < \mathbb{E}[W_{t+1}|w_t, X_t^* = 0] \quad (22)$$

The restriction (22) 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).²³

To see how this restriction orders the eigenvectors, and pins down the value of X_t^* , note that $\mathbb{E}[W_{t+1}|w_t, X_t^*] = f(W_{t+1} = 1|w_t, X_t^*)$, which is the second component of each eigenvector. Therefore, the monotonicity restriction (22) 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 model have been usefully applied in empirical work in IO (Ryan (2006)), productivity (Xu (2007), Collard-Wexler (2006)), and international trade (Dunne, Klimer, Roberts, and Xu (2006)).²⁴ In this model, $W_t = (Y_t, M_t)$, where Y_t is a firm’s R&D in year t , and M_t is the product’s installed base. The unobserved state variable X_t^* is the firm’s product quality, 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:

$$X_t^* = 0.8X_{t-1}^* + 0.2 \exp(\psi(Y_{t-1})) \nu_t. \quad (23)$$

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. (23) implies $f_{X_t^*|Y_{t-1}, M_{t-1}, X_{t-1}^*} = f_{X_t^*|Y_{t-1}, X_{t-1}^*}$.

Installed base evolves as:

$$M_{t+1} = M_t[1 + \exp(\eta_{t+1} + X_{t+1}^*)] \quad (24)$$

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

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

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. (24) 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:

$$\begin{aligned} Y_t &= Y^*(M_t, X_t^*, \gamma_t) \\ &= \operatorname{argmax}_{0 \leq y \leq \bar{I}} \left[\underbrace{\Pi(M_t, X_t^*)}_{\text{profits}} - \underbrace{\gamma_t}_{\text{shock}} \cdot \underbrace{Y_t^2}_{\text{R\&D cost}} + \beta \mathbb{E} \underbrace{V(M_{t+1}, X_{t+1}^*, \gamma_{t+1})}_{\text{value fxn}} \right] \end{aligned} \quad (25)$$

\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. (25) 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

$$f_{W_3|W_2, X_3^*} = f_{Y_3|M_3, X_3^*} \cdot f_{M_3|M_2, X_3^*}. \quad (26)$$

The first term in Eq. (26) is the density of R&D Y_3 . Because the first term is not a function of M_2 , Eq. (51) 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:

$$\begin{aligned} k(w_3, \bar{w}_3, w_2, \bar{w}_2, x_3^*) &= \frac{f_{W_3|W_2, X_3^*}(w_3|w_2, x_3^*) f_{W_3|W_2, X_3^*}(\bar{w}_3|\bar{w}_2, x_3^*)}{f_{W_3|W_2, X_3^*}(\bar{w}_3|w_2, x_3^*) f_{W_3|W_2, X_3^*}(w_3|\bar{w}_2, x_3^*)} \\ &= \frac{f_{M_3|M_2, X_3^*}(m_3|m_2, x_3^*) f_{M_3|M_2, X_3^*}(\bar{m}_3|\bar{m}_2, x_3^*)}{f_{M_3|M_2, X_3^*}(\bar{m}_3|m_2, x_3^*) f_{M_3|M_2, X_3^*}(m_3|\bar{m}_2, x_3^*)}. \end{aligned} \quad (27)$$

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

$$\begin{aligned} f_{M_3|M_2, X_3^*}(m_3|m_2, x_3^*) &= \frac{1}{(m_3 - m_2)} \exp \left[\log \left(\frac{m_3 - m_2}{m_2} \right) - x_3^* - e^{\log \left(\frac{m_3 - m_2}{m_2} \right) - x_3^*} \right] \\ &= \frac{e^{-x_3^*}}{m_2} \exp \left(-e^{-x_3^*} \left[\frac{m_3 - m_2}{m_2} \right] \right). \end{aligned} \quad (28)$$

Plugging this into Eq. (27), we obtain an expression for the eigenvalues

$$k(w_3, \bar{w}_3, w_2, \bar{w}_2, x_3^*) = \exp \left(-e^{-x_3^*} \left[\frac{-(\bar{m}_3 - m_3)(\bar{m}_2 - m_2)}{m_2 \bar{m}_2} \right] \right). \quad (29)$$

For given m_3 , we can pick a finite and nonzero m_2 ,²⁵ and set $(\bar{m}_3, \bar{m}_2) = (m_3 - \Delta, m_2 + \Delta)$, with Δ nonzero and small. At these values, the eigenvalues in Eq. (29) simplify to $\exp \left(-e^{-x_3^*} \left[\frac{\Delta^2}{m_2(m_2 + \Delta)} \right] \right)$ so that, for fixed m_3 , and $x_3^* \in \mathbb{R}$, $0 < k(w_3, \bar{w}_3, w_2, \bar{w}_2, x_3^*) < 1$, which satisfies Assumption 3(i). Moreover, the eigenvalues in Eq. (29) are monotonic in x_3^* for any given $(w_3, \bar{w}_3, w_2, \bar{w}_2)$, 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. (23) and (24)) 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

$$\begin{aligned} L_{M_4, w_3, w_2, M_1} &= L_{M_4 | w_3, X_3^*} D_{w_3 | w_2, X_3^*} L_{X_3^*, w_2, M_1} \\ &= L_{M_4 | w_3, X_3^*} D_{w_3 | w_2, X_3^*} L_{X_3^* | w_2, X_2^*} L_{X_2^*, w_2, M_1} \end{aligned} \quad (30)$$

$$L_{M_3, w_2, M_1} = L_{M_3 | w_2, X_2^*} L_{X_2^*, w_2, M_1}. \quad (31)$$

²⁵In 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 (\bar{m}_3, \bar{m}_2) as needed.

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 $L_{X_2^*, w_2, M_1}$.²⁶ 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 (m_3, m_2, x_3^*) 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. (24) whereby, for every (y_3, m_3) , M_4 is a convolution of X_4^* , ie. $\log [M_4 - M_3] - \log M_3 = X_4^* + \eta_4$. We have

$$\begin{aligned}
g(m_4) &\equiv \left(L_{M_4|w_3, X_4^*} h \right) (m_4) & (32) \\
&= \int_{-\infty}^{\infty} f_{M_4|w_3, X_4^*} (m_4|w_3, x_4^*) h(x_4^*) dx_4^* \\
&= \int_{-\infty}^{\infty} \frac{1}{m_4 - m_3} f_{\eta_4} \left(\log \left(\frac{m_4 - m_3}{m_3} \right) - x_4^* \right) h(x_4^*) dx_4^* \\
&= \frac{1}{m_4 - m_3} \int_{-\infty}^{\infty} f_{\eta_4} (\varphi_4 - x_4^*) h(x_4^*) dx_4^*, & \left[\varphi_4 \equiv \log \left(\frac{m_4 - m_3}{m_3} \right) \right] \\
&\equiv \frac{1}{m_4 - m_3} \times \left(L_{\varphi_4|X_4^*} h \right) (\varphi_4)
\end{aligned}$$

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 $f_{\eta_4} (\varphi_4 - x_4^*)$. 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,²⁷ 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 $h(x_4^*) = 0$ 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^* , ie. $X_3^* = 0.8X_2^* + 0.2 \exp(\psi(Y_2)) \nu_3$ (cf. Eq. (23)). 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

²⁶By 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. (24), 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.

²⁷The characteristic function for η_4 is $\phi_{\eta_4}(\tau) = \Gamma(1 + i\tau)$, which is nonzero for any $\tau \in \mathbb{R}$.

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 exogenously²⁸ – that is, $\psi(\cdot) = 0$ in Eq. (23) – 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. (23), 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; however, for a version of the Rust bus engine model (in the Supplemental appendix), stronger restrictions on (Y_1, X_1^*, M_1) – the joint distribution of initial states and decision variables – are required. At the same time, we also reiterate that these are sufficient conditions, and may not be necessary for the general results. In the appendix, we provide and discuss a *necessary* condition for operators to satisfy completeness, which allows for very general and flexible classes of joint densities.

²⁸A large class of investment models (eg. Olley and Pakes (1996), Levinsohn and Petrin (2003)) assume that the unobserved variable X_t^* (denoting productivity) evolves exogenously.

5 Concluding remarks

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

For a general k -th order Markov process ($k < \infty$), it can be shown that the $3k+2$ observations $W_{t+k}, \dots, W_{t-2k-1}$ can identify the Markov law of motion $f_{W_t, X_t^* | W_{t-1}, \dots, W_{t-k}, X_{t-1}^*, \dots, X_{t-k}^*}$, 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. Akerberg, Benkard, Berry, and Pakes (2007, Section 2.4.3) discuss the difficulties with multivariate unobserved state variables in the context of dynamic investment models.

Finally, this paper has focused on identification, but not estimation. In ongoing work, we are using our identification results to guide the estimation of dynamic models with unobserved state variables. This would complement recent papers on the estimation of parametric dynamic models with unobserved state variables, using non-CCP-based approaches.²⁹

6 APPENDIX A: Proofs

Proof: (Lemma 1)

²⁹Imai, Jain, and Ching (2009) and Norets (2009) consider Bayesian estimation, and Fernandez-Villaverde and Rubio-Ramirez (2007) consider efficient simulation estimation based on particle filtering.

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

$$\begin{aligned}
& \int \int f_{W_{t+1}, W_t, X_t^*, X_{t-1}^*, W_{t-1}, W_{t-2}} dx_t^* dx_{t-1}^* & (33) \\
&= \int \int f_{W_{t+1}|W_t, W_{t-1}, W_{t-2}, X_t^*, X_{t-1}^*} f_{W_t, X_t^*|W_{t-1}, W_{t-2}, X_{t-1}^*} f_{X_{t-1}^*, W_{t-1}, W_{t-2}} dx_t^* dx_{t-1}^* \\
&= \int \int f_{W_{t+1}|W_t, X_t^*} f_{W_t, X_t^*|W_{t-1}, X_{t-1}^*} f_{X_{t-1}^*, W_{t-1}, W_{t-2}} dx_t^* dx_{t-1}^* \\
&= \int \int f_{W_{t+1}|W_t, X_t^*} f_{W_t|W_{t-1}, X_t^*, X_{t-1}^*} f_{X_t^*|W_{t-1}, X_{t-1}^*} f_{X_{t-1}^*, W_{t-1}, W_{t-2}} dx_t^* dx_{t-1}^* \\
&= \int \int f_{W_{t+1}|W_t, X_t^*} f_{W_t|W_{t-1}, X_t^*, X_{t-1}^*} f_{X_t^*|W_{t-1}, W_{t-2}, X_{t-1}^*} f_{X_{t-1}^*, W_{t-1}, W_{t-2}} dx_t^* dx_{t-1}^* \\
&= \int \int f_{W_{t+1}|W_t, X_t^*} f_{W_t|W_{t-1}, X_t^*, X_{t-1}^*} f_{X_t^*, X_{t-1}^*, W_{t-1}, W_{t-2}} dx_t^* dx_{t-1}^*.
\end{aligned}$$

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

$$\begin{aligned}
f_{W_{t+1}, W_t, W_{t-1}, W_{t-2}} &= \int f_{W_{t+1}|W_t, X_t^*} f_{W_t|W_{t-1}, X_t^*} \left(\int f_{X_t^*, X_{t-1}^*, W_{t-1}, W_{t-2}} dx_{t-1}^* \right) dx_t^* \\
&= \int f_{W_{t+1}|W_t, X_t^*} f_{W_t|W_{t-1}, X_t^*} f_{X_t^*, W_{t-1}, W_{t-2}} dx_t^*. & (34)
\end{aligned}$$

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

$$L_{W_{t+1}, w_t, w_{t-1}, W_{t-2}} = L_{W_{t+1}|w_t, X_t^*} D_{w_t|w_{t-1}, X_t^*} L_{X_t^*, w_{t-1}, W_{t-2}}. \quad (35)$$

For the variable(s) $V_t \subseteq W_t$, for all periods t , introduced in Assumption 2, Eq. (35) implies that the joint density of $\{V_{t+1}, W_t, W_{t-1}, V_{t-2}\}$ is expressed in operator notation as $L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} = L_{V_{t+1}|w_t, X_t^*} D_{w_t|w_{t-1}, X_t^*} L_{X_t^*, w_{t-1}, V_{t-2}}$, as postulated by Lemma 1. *Q.E.D.*

Proof: (Lemma 2)

Assumption 1 implies the following two equalities:

$$\begin{aligned}
f_{V_{t+1}, W_t, W_{t-1}, V_{t-2}} &= \int f_{V_{t+1}|W_t, X_t^*} f_{W_t, X_t^*, W_{t-1}, V_{t-2}} dx_t^* \\
f_{W_t, X_t^*, W_{t-1}, V_{t-2}} &= \int f_{W_t, X_t^*|W_{t-1}, X_{t-1}^*} f_{X_{t-1}^*, W_{t-1}, V_{t-2}} dx_{t-1}^*. & (36)
\end{aligned}$$

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

$$\begin{aligned} L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} &= L_{V_{t+1} | w_t, X_t^*} L_{w_t, X_t^* | w_{t-1}, V_{t-2}} \\ L_{w_t, X_t^* | w_{t-1}, V_{t-2}} &= L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*} L_{X_{t-1}^* | w_{t-1}, V_{t-2}}. \end{aligned} \quad (37)$$

Substituting the second line into the first, we get

$$\begin{aligned} L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} &= L_{V_{t+1} | w_t, X_t^*} L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*} L_{X_{t-1}^* | w_{t-1}, V_{t-2}} \\ \Leftrightarrow L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*} L_{X_{t-1}^* | w_{t-1}, V_{t-2}} &= L_{V_{t+1} | w_t, X_t^*}^{-1} L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}}. \end{aligned} \quad (38)$$

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

$$f_{V_t, w_{t-1}, V_{t-2}} = \int f_{V_t | w_{t-1}, X_{t-1}^*} f_{X_{t-1}^* | w_{t-1}, V_{t-2}} dx_{t-1}^* \quad (39)$$

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

$$L_{V_t, w_{t-1}, V_{t-2}} = L_{V_t | w_{t-1}, X_{t-1}^*} L_{X_{t-1}^* | w_{t-1}, V_{t-2}} \Rightarrow L_{X_{t-1}^* | w_{t-1}, V_{t-2}} = L_{V_t | w_{t-1}, X_{t-1}^*}^{-1} L_{V_t, w_{t-1}, V_{t-2}} \quad (40)$$

where the right-hand side applies Assumption 2(ii). Hence, substituting the above into Eq. (38), we obtain the desired representation

$$\begin{aligned} L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*} L_{V_t | w_{t-1}, X_{t-1}^*}^{-1} L_{V_t, w_{t-1}, V_{t-2}} &= L_{V_{t+1} | w_t, X_t^*}^{-1} L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} \\ \Rightarrow L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*} L_{V_t | w_{t-1}, X_{t-1}^*}^{-1} &= L_{V_{t+1} | w_t, X_t^*}^{-1} L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} L_{V_t, w_{t-1}, V_{t-2}}^{-1} \\ \Rightarrow L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*} &= L_{V_{t+1} | w_t, X_t^*}^{-1} L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} L_{V_t, w_{t-1}, V_{t-2}}^{-1} L_{V_t | w_{t-1}, X_{t-1}^*}. \end{aligned} \quad (41)$$

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

Proof: (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 $(\bar{w}_t, \bar{w}_{t-1})$ within the neighborhood \mathcal{N}^r to satisfy Assumption 3. Because $(\bar{w}_t, \bar{w}_{t-1}) \in \mathcal{N}^r$, also $(\bar{w}_t, w_{t-1}), (w_t, \bar{w}_{t-1}) \in \mathcal{N}^r$. 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}) , (\bar{w}_t, w_{t-1}) , (w_t, \bar{w}_{t-1}) , $(\bar{w}_t, \bar{w}_{t-1})$, each pair of equations will share one operator in common. Specifically:

$$\text{for } (w_t, w_{t-1}) : L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} = L_{V_{t+1} | w_t, X_t^*} D_{w_t | w_{t-1}, X_t^*} L_{X_t^*, w_{t-1}, V_{t-2}}, \quad (42)$$

$$\text{for } (\bar{w}_t, w_{t-1}) : L_{V_{t+1}, \bar{w}_t, w_{t-1}, V_{t-2}} = L_{V_{t+1} | \bar{w}_t, X_t^*} D_{\bar{w}_t | w_{t-1}, X_t^*} L_{X_t^*, w_{t-1}, V_{t-2}}, \quad (43)$$

$$\text{for } (w_t, \bar{w}_{t-1}) : L_{V_{t+1}, w_t, \bar{w}_{t-1}, V_{t-2}} = L_{V_{t+1} | w_t, X_t^*} D_{w_t | \bar{w}_{t-1}, X_t^*} L_{X_t^*, \bar{w}_{t-1}, V_{t-2}}, \quad (44)$$

$$\text{for } (\bar{w}_t, \bar{w}_{t-1}) : L_{V_{t+1}, \bar{w}_t, \bar{w}_{t-1}, V_{t-2}} = L_{V_{t+1} | \bar{w}_t, X_t^*} D_{\bar{w}_t | \bar{w}_{t-1}, X_t^*} L_{X_t^*, \bar{w}_{t-1}, V_{t-2}}. \quad (45)$$

Assumption 2(ii) implies that $L_{V_{t+1} | \bar{w}_t, X_t^*}$ is invertible. Moreover, Assumption 3(i) implies $f_{W_t | W_{t-1}, X_t^*}(\bar{w}_t | w_{t-1}, x_t^*) > 0$ for all x_t^* so that $D_{\bar{w}_t | w_{t-1}, X_t^*}$ is invertible. We can then solve for $L_{X_t^*, w_{t-1}, V_{t-2}}$ from Eq. (43) as

$$D_{\bar{w}_t | w_{t-1}, X_t^*}^{-1} L_{V_{t+1} | \bar{w}_t, X_t^*}^{-1} L_{V_{t+1}, \bar{w}_t, w_{t-1}, V_{t-2}} = L_{X_t^*, w_{t-1}, V_{t-2}}. \quad (46)$$

Plugging in this expression to Eq. (42) leads to

$$L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} = L_{V_{t+1} | w_t, X_t^*} D_{w_t | w_{t-1}, X_t^*} D_{\bar{w}_t | w_{t-1}, X_t^*}^{-1} L_{V_{t+1} | \bar{w}_t, X_t^*}^{-1} L_{V_{t+1}, \bar{w}_t, w_{t-1}, V_{t-2}}. \quad (47)$$

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

$$\begin{aligned} \mathbf{A} &\equiv L_{V_{t+1}, w_t, w_{t-1}, V_{t-2}} L_{V_{t+1}, \bar{w}_t, w_{t-1}, V_{t-2}}^{-1} \\ &= L_{V_{t+1} | w_t, X_t^*} D_{w_t | w_{t-1}, X_t^*} D_{\bar{w}_t | w_{t-1}, X_t^*}^{-1} L_{V_{t+1} | \bar{w}_t, X_t^*}^{-1}. \end{aligned} \quad (48)$$

Similar manipulations of Eqs. (44) and Eq. (45) lead to

$$\begin{aligned} \mathbf{B} &\equiv L_{V_{t+1}, \bar{w}_t, \bar{w}_{t-1}, V_{t-2}} L_{V_{t+1}, w_t, \bar{w}_{t-1}, V_{t-2}}^{-1} \\ &= L_{V_{t+1} | \bar{w}_t, X_t^*} D_{\bar{w}_t | \bar{w}_{t-1}, X_t^*} D_{w_t | \bar{w}_{t-1}, X_t^*}^{-1} L_{V_{t+1} | w_t, X_t^*}^{-1}. \end{aligned} \quad (49)$$

Assumption 2(i) guarantees that, for any w_t , $(\bar{w}_t, w_{t-1}, \bar{w}_{t-1})$ exist so that (9) and (10) are

valid operations. Finally, we postmultiply Eq. (48) by Eq. (49) to obtain

$$\begin{aligned}
\mathbf{AB} &= L_{V_{t+1}|w_t, X_t^*} D_{w_t|w_{t-1}, X_t^*} D_{\bar{w}_t|w_{t-1}, X_t^*}^{-1} \left(L_{V_{t+1}|\bar{w}_t, X_t^*}^{-1} L_{V_{t+1}|\bar{w}_t, X_t^*} \right) \times \\
&\quad \times D_{\bar{w}_t|\bar{w}_{t-1}, X_t^*} D_{w_t|\bar{w}_{t-1}, X_t^*}^{-1} L_{V_{t+1}|w_t, X_t^*}^{-1} \\
&= L_{V_{t+1}|w_t, X_t^*} \left(D_{w_t|w_{t-1}, X_t^*} D_{\bar{w}_t|w_{t-1}, X_t^*}^{-1} D_{\bar{w}_t|\bar{w}_{t-1}, X_t^*} D_{w_t|\bar{w}_{t-1}, X_t^*}^{-1} \right) L_{V_{t+1}|w_t, X_t^*}^{-1} \\
&\equiv L_{V_{t+1}|w_t, X_t^*} D_{w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, X_t^*} L_{V_{t+1}|w_t, X_t^*}^{-1}, \quad \text{where} \tag{50}
\end{aligned}$$

$$\begin{aligned}
(D_{w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, X_t^*} h)(x_t^*) &= \left(D_{w_t|w_{t-1}, X_t^*} D_{\bar{w}_t|w_{t-1}, X_t^*}^{-1} D_{\bar{w}_t|\bar{w}_{t-1}, X_t^*} D_{w_t|\bar{w}_{t-1}, X_t^*}^{-1} \right) (x_t^*) \\
&= \frac{f_{W_t|W_{t-1}, X_t^*}(w_t|w_{t-1}, x_t^*) f_{W_t|W_{t-1}, X_t^*}(\bar{w}_t|\bar{w}_{t-1}, x_t^*)}{f_{W_t|W_{t-1}, X_t^*}(\bar{w}_t|w_{t-1}, x_t^*) f_{W_t|W_{t-1}, X_t^*}(w_t|\bar{w}_{t-1}, x_t^*)} h(x_t^*) \tag{51} \\
&\equiv k(w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, x_t^*) h(x_t^*).
\end{aligned}$$

This equation implies that the observed operator \mathbf{AB} on the left hand side of Eq. (50) has an inherent eigenvalue-eigenfunction decomposition, with the eigenvalues corresponding to the function $k(w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, x_t^*)$ and the eigenfunctions corresponding to the density $f_{V_{t+1}|W_t, X_t^*}(\cdot|w_t, x_t^*)$. The decomposition in Eq. (50) is similar to the decomposition in Hu and Schennach (2008) or Carroll, Chen, and Hu (2009).

Assumption 3 ensures that this decomposition is unique. Specifically, Eq. (50) implies that the operator \mathbf{AB} on the LHS has the same spectrum as the diagonal operator $D_{w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, X_t^*}$. Assumption 3(i) guarantees that the spectrum of the diagonal operator $D_{w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, X_t^*}$ is bounded. Since an operator is bounded by the largest element of its spectrum, Assumption 3(i) also implies that the operator \mathbf{AB} 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. (50) itself does not imply that the eigenvalues $k(w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, x_t^*)$ 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 \bar{w}_t , w_{t-1} , and \bar{w}_{t-1} such that the eigenvalues are distinct across all x_t^* .^{30,31}

Second, the eigenfunctions $f_{V_{t+1}|W_t, X_t^*}(\cdot|w_t, x_t^*)$ in the spectral decomposition (50) 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. (50). *Q.E.D.*

Proof: (Corollary 1)

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

$$\begin{aligned} f_{V_t, W_{t-1}=w_{t-1}} &= L_{V_t|w_{t-1}, X_{t-1}^*} f_{W_{t-1}=w_{t-1}, X_{t-1}^*} \\ \Leftrightarrow f_{W_{t-1}=w_{t-1}, X_{t-1}^*} &= L_{V_t|w_{t-1}, X_{t-1}^*}^{-1} f_{V_t, W_{t-1}=w_{t-1}} \end{aligned} \quad (52)$$

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

Proof: (Corollary 3)

Under stationarity, the operator $L_{V_{t-1}|w_{t-2}, X_{t-2}^*}$ is the same as $L_{V_{t+1}|w_t, X_t^*}$, which is identified from the observed density $f_{V_{t+1}, W_t, W_{t-1}, V_{t-2}}$ (by Lemma 3). Because $f_{V_{t-1}, W_{t-2}} = \int f_{V_{t-1}|W_{t-2}, X_{t-2}^*} f_{W_{t-2}, X_{t-2}^*} dx_{t-2}^*$, the same argument as in the proof of Corollary 1 then implies that f_{W_{t-2}, X_{t-2}^*} is identified from the observed density $f_{V_{t-1}, W_{t-2}}$. *Q.E.D.*

³⁰Specifically, the operators \mathbf{AB} corresponding to different values of $(\bar{w}_t, w_{t-1}, \bar{w}_{t-1})$ 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 $f_{V_{t+1}|W_t, X_t^*}(\cdot|w_t, \tilde{x}_t^*)$, one can always find values of $(\bar{w}_t, w_{t-1}, \bar{w}_{t-1})$ such that the two different eigenfunctions correspond to two different eigenvalues, i.e., $k(w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, x_t^*) \neq k(w_t, \bar{w}_t, w_{t-1}, \bar{w}_{t-1}, \tilde{x}_t^*)$.

³¹When w_t (resp. w_{t-1}) is close to \bar{w}_t (resp. \bar{w}_{t-1}), Eq. (51) 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 .

7 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

$$\begin{aligned} (L_{X|Z}h)(x) &= \int f_\epsilon(x - z) h(z) dz \\ (L_{X|Z}^*h)(z) &= \int f_\epsilon(x - z) h(x) dx. \end{aligned} \tag{53}$$

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: (Lemma 4)

We have

$$\begin{aligned} g(x) &\equiv (L_{X|Z}h)(x) \\ &= \int f_\epsilon(x - z) h(z) dz. \end{aligned} \tag{54}$$

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

$$\phi_g(t) = \phi_\epsilon(t)\phi_h(t). \tag{55}$$

Therefore, $\phi_g = 0$ implies $\phi_h = 0$ if $\phi_\epsilon(t) \neq 0$ 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

$$\begin{aligned} \varphi(z) &\equiv (L_{X|Z}^*\psi)(z) \\ &= \int f_\epsilon(x - z) \psi(x) dx \\ &\equiv \int \kappa(z - x) \psi(x) dx \end{aligned} \tag{56}$$

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

$$\begin{aligned}\phi_\varphi(t) &= \phi_\kappa(t)\phi_\psi(t) \\ &= \phi_\epsilon(-t)\phi_\psi(t).\end{aligned}\tag{57}$$

Again, $\phi_\varphi = 0$ implies $\phi_\psi = 0$ because $\phi_\epsilon(t) \neq 0$ for any $t \in \mathbb{R}$. Thus, $L_{X|Z}^*$ is injective. *Q.E.D.*

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

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

$$\begin{aligned}f_{X_2^*, w_2, M_1} &= \int \int f_{X_2^*, y_2, m_2, X_1^*, Y_1, M_1} dy_1 dx_1^* \\ &= f_{y_2|m_2, X_2^*} f_{m_2|X_2^*, M_1} \int \int f_{X_2^*|Y_1, X_1^*} f_{Y_1|X_1^*, M_1} f_{X_1^*, M_1} dy_1 dx_1^* \\ &= f_{y_2|m_2, X_2^*} f_{m_2|X_2^*, M_1} \int f_{X_2^*|X_1^*} \left(\int f_{Y_1|X_1^*, M_1} dy_1 \right) f_{X_1^*, M_1} dx_1^* \\ &= f_{y_2|m_2, X_2^*} f_{m_2|X_2^*, M_1} \left(\int f_{X_2^*|X_1^*} f_{X_1^*} dx_1^* \right) f_{M_1} \\ &= f_{y_2|m_2, X_2^*} f_{X_2^*} f_{m_2|X_2^*, M_1} f_{M_1}\end{aligned}\tag{58}$$

In the third line, we have utilized the restriction that $\psi(\cdot) = 0$ in Eq. (23) 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

$$L_{X_2^*, w_2, M_1} = D_{y_2|m_2, X_2^*} D_{X_2^*} L_{m_2|X_2^*, M_1} D_{M_1}.\tag{59}$$

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:

$$\begin{aligned}
g(x_2^*) &\equiv \left(L_{m_2|X_2^*,M_1}h\right)(x_2^*) & (60) \\
&= \int_0^{m_2} f_{m_2|X_2^*,M_1}(m_2|x_2^*,m_1)h(m_1)dm_1 \\
&= \int_0^{m_2} \frac{1}{m_2-m_1}f_{\eta_2}\left(\log\left(\frac{m_2-m_1}{m_1}\right)-x_2^*\right)h(m_1)dm_1 \\
&= \int_0^{m_2} \frac{1}{m_2-m_1}\left(\frac{-m_2}{(m_2-m_1)m_1}\right)^{-1}f_{\eta_2}\left(\log\left(\frac{m_2-m_1}{m_1}\right)-x_2^*\right)h(m_1)d\log\left(\frac{m_2-m_1}{m_1}\right) \\
&= \int_{m_2}^0 \frac{m_1}{m_2}f_{\eta_2}\left(\log\left(\frac{m_2-m_1}{m_1}\right)-x_2^*\right)h(m_1)d\log\left(\frac{m_2-m_1}{m_1}\right) \\
&= \int_{-\infty}^{\infty} f_{\eta_2}(\varphi_2-x_2^*)h\left(\frac{m_2}{e^{\varphi_2}+1}\right)\frac{1}{e^{\varphi_2}+1}d\varphi_2, \quad \left[\varphi_2 \equiv \log\left(\frac{m_2-m_1}{m_1}\right)\right] \\
&\equiv \int_{-\infty}^{\infty} f_{\eta_2}(\varphi_2-x_2^*)\tilde{h}(\varphi_2)d\varphi_2, \quad \left[\tilde{h}(\varphi_2) \equiv h\left(\frac{m_2}{e^{\varphi_2}+1}\right)\frac{1}{e^{\varphi_2}+1}\right] \\
&= \left(L_{\varphi_2|X_2^*}^*\tilde{h}\right)(x_2^*),
\end{aligned}$$

where the operator $L_{\varphi_2|X_2^*}^*$ is defined analogously to Eq. (53). As shown above, $g(x_2^*) = 0$ for any $x_2^* \in \mathbb{R}$ implies that $\left(L_{\varphi_2|X_2^*}^*\tilde{h}\right)(x_2^*) = 0$ for any $x_2^* \in \mathbb{R}$, where the kernel of $L_{\varphi_2|X_2^*}^*$ has a convolution form $f_{\eta_2}(\varphi_2-x_2^*)$. 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, $\left(L_{\varphi_2|X_2^*}^*\tilde{h}\right)(x_2^*) = 0$ for any $x_2^* \in \mathbb{R}$ implies $\tilde{h}(\varphi_2) = 0$ for any $\varphi_2 \in \mathbb{R}$. Next, because $\tilde{h}(\varphi_2) = h\left(\frac{m_2}{e^{\varphi_2}+1}\right)\frac{1}{e^{\varphi_2}+1}$ and $\frac{1}{e^{\varphi_2}+1}$ is nonzero, $\tilde{h}(\varphi_2) = 0$ for any $\varphi_2 \in \mathbb{R}$ implies $h\left(\frac{m_2}{e^{\varphi_2}+1}\right) = 0$ for any $\varphi_2 \in \mathbb{R}$. Given $\varphi_2 \equiv \log\left(\frac{m_2-m_1}{m_1}\right)$, we have $h(m_1) = 0$ for any $m_1 \in (0, m_2)$. Altogether, then, $g(x_2^*) = 0$ for any $x_2^* \in \mathbb{R}$ implies $h(m_1) = 0$ for any $m_1 \in (0, m_2)$, thus demonstrating the injectivity of the operator $L_{m_2|X_2^*,M_1}$, as claimed. *Q.E.D.*

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) :

$$\begin{aligned}
f_{M_1,w_2,w_3,M_4} &= \int f_{M_4|w_3,X_3^*}f_{w_3|w_2,X_3^*}f_{X_3^*,w_2,M_1}dx_3^*. & (61) \\
&= \int \left(\int f_{M_4|w_3,X_4^*}f_{X_4^*|w_3,X_3^*}dx_4^*\right)f_{w_3|w_2,X_3^*}\left(\int f_{X_3^*|w_2,X_2^*}f_{X_2^*,w_2,M_1}dx_2^*\right)dx_3^* \\
&= \int \left(\int f_{M_4|w_3,X_4^*}f_{X_4^*|w_3,X_3^*}dx_4^*\right)f_{w_3|w_2,X_3^*}\left(\int f_{X_3^*|w_2,X_2^*}f_{y_2|w_2,X_2^*}f_{X_2^*}f_{m_2|X_2^*,M_1}f_{M_1}dx_2^*\right)dx_3^*.
\end{aligned}$$

Therefore, the equivalent operator equation is

$$\begin{aligned} L_{M_1, w_2, w_3, M_4} &= L_{M_1, y_2, m_2, y_3, m_3, M_4} \\ &= D_{M_1} L_{m_2|X_2^*, M_1}^* D_{X_2^*} D_{y_2|m_2, X_2^*} L_{X_3^*|w_2, X_2^*}^* D_{w_3|w_2, X_3^*} L_{X_4^*|w_3, X_3^*}^* L_{M_4|w_3, X_4^*}^* \end{aligned} \quad (62)$$

In the above, the L^* operators are defined analogously to Eq. (53), 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 above.

Finally, all the densities corresponding to the diagonal operators in Eq.(62) 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

$$\begin{aligned} f_{M_1, w_2, M_3} &= \int f_{M_3|w_2, X_2^*} f_{X_2^*, w_2, M_1} dx_2^* \\ &= \int \left(\int f_{M_3|w_2, X_3^*} f_{X_3^*|w_2, X_2^*} dx_3^* \right) f_{X_2^*, w_2, M_1} dx_2^* \\ &= \int \left(\int f_{M_3|w_2, X_3^*} f_{X_3^*|w_2, X_2^*} dx_3^* \right) f_{y_2|m_2, X_2^*} f_{X_2^*} f_{m_2|X_2^*, M_1} f_{M_1} dx_2^*. \end{aligned} \quad (63)$$

Therefore, the equivalent operator equation is

$$L_{M_1, w_2, M_3} = D_{M_1} L_{m_2|X_2^*, M_1}^* D_{X_2^*} D_{y_2|m_2, X_2^*} L_{X_3^*|w_2, X_2^*}^* L_{M_3|w_2, X_3^*}^*. \quad (64)$$

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. (62), and we argued above that these were injective. Thus, L_{M_1, w_2, M_3} is injective. *Q.E.D.*

8 APPENDIX C: Supplemental materials

8.1 Additional example: dynamic discrete-choice model based on Rust

(1987)

In addition to the two examples presented in the main paper, we present here a discussion of our assumptions in the context of a third example: Rust's (1987) bus-engine replacement

model, augmented to allow for time-varying serially-correlated unobserved state variables. In this model, $W_t = (Y_t, M_t)$, where Y_t is the indicator that the bus engine was replaced in week t , and M_t is the mileage since the last engine replacement.

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

$$u_t = \begin{cases} u_0(S_t) + \epsilon_{0t} & \text{if } Y_t = 0 \\ u_1(S_t) + \epsilon_{1t} & \text{if } Y_t = 1 \end{cases} \quad (65)$$

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

The choice-specific utility functions are:

$$u_0(S_t) = -c(M_t); \quad u_1(S_t) = -RC. \quad (66)$$

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

$$M_{t+1} = \begin{cases} M_t[1 + \exp(\eta_{t+1} + X_{t+1}^*)] & \text{if } Y_t = 0 \\ \exp(\eta_{t+1} + X_{t+1}^*) & \text{if } Y_t = 1 \end{cases} \quad (67)$$

where $\eta_{t+1} \in \mathbb{R}$ follows an extreme value distribution with density $f_{\eta_{t+1}}(\eta) = \exp(\eta - e^\eta)$. This law of motion implies $f_{M_{t+1}|Y_t, M_t, X_t^*, X_{t+1}^*} = f_{M_{t+1}|Y_t, M_t, X_{t+1}^*}$. Hence, X_t^* affects the evolution of mileage, but not the agent's utilities. Furthermore, following Rust's assumptions, previous mileage M_{t-1} has no direct effect on current mileage M_t when the engine was replaced in the previous period ($Y_{t-1} = 1$).

X_t^* , the time-varying unobserved state variable, denotes the general physical condition (wear and tear) of the bus, which is unobserved by the econometrician but observed by the bus mechanics, and affects their decisions about replacing the bus engine. It evolves as an AR(1) process:

$$X_t^* = 0.8X_{t-1}^* + 0.2\nu_t. \quad (68)$$

$\nu_t \in \mathbb{R}$ is a standard normal shock, distributed independently over t . This law of motion

implies $f_{X_t^*|Y_{t-1}, M_{t-1}, X_{t-1}^*} = f_{X_t^*|X_{t-1}^*}$. We also assume that $X_1^* \in \mathbb{R}$. Hence, $X_t^*|X_{t-1}^*$ is distributed with density determined by $f_{\nu_t}(\cdot)$.

In this stationary dynamic optimization model, the conditional choice probabilities take the multinomial logit form (for $Y_t = 0, 1$): $P(Y_t|S_t) = \exp(V_{Y_t}(S_t)) / \left[\sum_{y=0}^1 \exp(V_y(S_t)) \right]$ where $V_y(S_t)$ is the choice-specific value function in period t , defined recursively by $V_y(S_t) = u_y(S_t) + \beta E \left[\log \left\{ \sum_{y'=0}^1 \exp(V_{y'}(S_{t+1})) \right\} | Y_t = y, S_t \right]$.

The arguments we use here are very similar to those we used to verify the assumptions in Example 2 of the main paper. To avoid confusion, we repeat most of the details here. We verify the assumptions out of order, leaving 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

$$f_{W_3|W_2, X_3^*} = f_{Y_3|M_3, X_3^*} \cdot f_{M_3|Y_2, M_2, X_3^*}. \quad (69)$$

Assumption 3(i) requires that, for any w_3 , there exists $(w_2, \bar{w}_3, \bar{w}_2)$ such that the eigenvalues $k(w_3, \bar{w}_3, w_2, \bar{w}_2, x_3^*)$ are bounded between 0 and a constant C . The first term is the CCP $f_{Y_3|M_3, X_3^*}$, which is a logit probability. Because the per-period utilities, net of the ϵ 's, are bounded away from $-\infty$ and $+\infty$, the logit choice probabilities are also bounded from zero. Moreover, the CCP's are not a function of W_2 , so that Eq. (12) in the main text implies that the eigenvalues $k(w_3, \bar{w}_3, w_2, \bar{w}_2, x_3^*)$ in the spectral decomposition will not be a function of the CCP's.

The second term is the mileage law of motion $f_{M_3|Y_2, M_2, X_3^*}$ which, by assumption, is

$$\begin{aligned} & f_{M_3|Y_2, M_2, X_3^*}(m_3|y_2, m_2, x_3^*) \\ &= \frac{1}{m_3 - (1 - y_2)m_2} f_{\eta_3}(\ln(m_3 - (1 - y_2)m_2) - (1 - y_2) \ln m_2 - x_3^*) \end{aligned} \quad (70)$$

and f_{η_3} denotes the density of the extreme value distribution. For any $w_3 = (y_3, m_3)$, we consider

$$\begin{aligned} w_2 &= (y_2, m_2) = (0, m_2), \\ \bar{w}_2 &= (\bar{y}_2, \bar{m}_2) = (0, m_2 + \Delta), \\ \bar{w}_3 &= (\bar{y}_3, \bar{m}_3) = (y_3, m_3 - \Delta), \end{aligned} \quad (71)$$

which implies that the bus engine was not replaced in period 2. We may show that

$$k(w_3, \bar{w}_3, w_2, \bar{w}_2, x_3^*) = \exp\left(-e^{-x_3^*} \left[\frac{\Delta^2}{m_2 \bar{m}_2}\right]\right), \quad (72)$$

which is bounded between zero and one. Therefore, Assumption 3(i) holds. Furthermore, the equation above implies that the eigenvalue $k(w_3, \bar{w}_3, w_2, \bar{w}_2, x_3^*)$ is monotonic in x_3^* , which implies Assumption 3(ii).

Assumption 4 presumes a known functional G such that $G[f_{M_4|Y_3, M_3, X_3^*}(\cdot|y_3, m_3, x_3^*)]$ is monotonic in x_3^* , where we use $V_t = M_t$, for all periods t . Eqs. (68) and (67) imply that

$$M_4 = (1 - Y_3)M_3 + (M_3)^{1-Y_3} \exp(\eta_4 + 0.2\nu_4) \cdot \exp(0.8X_3^*). \quad (73)$$

Let C_{med} denote the median of the random variable $\exp(\eta_4 + 0.2\nu_4)$. Then

$$\text{med} \left[f_{M_4|Y_3, M_3, X_3^*}(\cdot|y_3, m_3, x_3^*) \right] = (1 - y_3)m_3 + (m_3)^{1-y_3} C_{med} \cdot \exp(0.8x_3^*) \quad (74)$$

which is monotonic in x_3^* . Hence, we can pin down $x_3^* = \text{med} \left[f_{M_4|Y_3, M_3, X_3^*}(\cdot|y_3, m_3, x_3^*) \right]$.

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

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. Applying Claim 2 from Example 2 of the main paper, it suffices to show 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} . Assumption 1, along with the assumptions on the laws of motion, implies that

$$\begin{aligned} L_{M_4, w_3, w_2, M_1} &= L_{M_4|w_3, X_3^*} D_{w_3|w_2, X_3^*} L_{X_3^*, w_2, M_1} \\ &= L_{M_4|w_3, X_3^*} D_{w_3|w_2, X_3^*} L_{X_3^*|X_2^*} L_{X_2^*, w_2, M_1} \end{aligned} \quad (75)$$

$$L_{M_3, w_2, M_1} = L_{M_3|w_2, X_2^*} L_{X_2^*, w_2, M_1}. \quad (76)$$

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

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 $L_{M_4|w_3, X_4^*}$, $D_{w_3|w_2, X_3^*}$, $L_{X_3^*|X_2^*}$ and $L_{X_2^*, w_2, M_1}$.

(i) The diagonal operator $D_{w_3|w_2, X_3^*}$ has corresponding density 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 have shown that both $f_{y_3|m_3, X_3^*}$ and $f_{m_3|m_2, X_3^*}$ are nonzero, for all values of (y_3, m_3, m_2, x_3^*) in the support. Therefore, the operator $D_{w_3|w_2, X_3^*}$ is injective.

(ii) For $L_{M_4|w_3, X_4^*}$, we use Eq. (67) whereby, for every (y_3, m_3) , M_4 is a convolution of

X_4^* , ie. $\log [M_4 - (1 - y_3)m_3] - (1 - y_3) \log m_3 = X_4^* + \eta_4$. As is well-known, as long as the characteristic function of η_4 has no real zeros, which is satisfied by the assumed extreme value distribution, the corresponding operator is injective.

(iii) Similarly, X_3^* is a convolution of X_2^* , ie. $X_3^* = 0.8X_2^* + 0.2\nu_3$ (cf. Eq. (68)). Hence, $L_{X_3^*|w_2, X_2^*}$ is injective if the characteristic function of ν_3 has no real zeros, which is satisfied by the assumed normal distribution.

(iv) For the operator $L_{X_2^*, w_2, M_1}$, 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 $f_{X_1^*, M_1}$. However, as in Example 2 of the main paper, it turns out that some stochastic assumptions on the initial conditions (Y_1, M_1, X_1^*) ensure injectivity of $L_{X_2^*, w_2, M_1}$.

Claim 3: If (i) $f_{X_1^*, M_1} = f_{X_1^*} f_{M_1}$; and (ii) $Y_1 = 0$ with probability one, and is exogenous, then $L_{X_2^*, w_2, M_1}$ is injective.

Proof: The kernel of the operator $L_{X_2^*, w_2, M_1}$ is

$$\begin{aligned}
f_{X_2^*, w_2, M_1} &= \int \int f_{X_2^*, y_2, m_2, X_1^*, Y_1, M_1} dy_1 dx_1^* & (77) \\
&= f_{y_2|m_2, X_2^*} \int \int f_{m_2|X_2^*, M_1, Y_1} f_{X_2^*|X_1^*} f_{Y_1|M_1, X_1^*} f_{X_1^*, M_1} dy_1 dx_1^* \\
&= f_{y_2|m_2, X_2^*} f_{m_2|X_2^*, M_1, Y_1=0} \int f_{X_2^*|X_1^*} f_{X_1^*, M_1} dx_1^* \\
&= f_{y_2|m_2, X_2^*} f_{m_2|X_2^*, M_1, Y_1=0} \left[\int f_{X_2^*|X_1^*} f_{X_1^*} dx_1^* \right] f_{M_1} \\
&= f_{y_2|m_2, X_2^*} f_{X_2^*} f_{m_2|X_2^*, M_1, Y_1=0} f_{M_1}
\end{aligned}$$

In the third line, we have utilized condition (ii), which implies $f_{Y_1|M_1, X_1^*}(0|m_1, x_1^*) = 1$, i.e., no engine is changed in the initial period of data. The fourth line applies the independence condition (i). The equivalent operator equation is

$$L_{X_2^*, w_2, M_1} = D_{y_2|m_2, X_2^*} D_{X_2^*} L_{m_2|X_2^*, M_1, Y_1=0} D_{M_1}. \quad (78)$$

The injectivity of $L_{X_2^*, w_2, M_1}$ then relies on that of the operator $L_{m_2|X_2^*, M_1, Y_1=0}$ which, as Eq. (67) shows, is based on a convolution form. Using an argument identical to that used in the proof of Claim 1 in Appendix B of the main paper, we can show that $L_{m_2|X_2^*, M_1, Y_1=0}$ is injective, which implies that $L_{X_2^*, w_2, M_1}$ is also injective. ■

Therefore, for this example, 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} . By applying Claim 2 in Example 2 of the main paper, we also obtain the injectivity of L_{M_1, w_2, w_3, M_4} and L_{M_1, w_2, M_3} , as required by Assumption 2(i) and 2(iii).

Without the assumption that the initial value Y_1 is exogenous, and that $Y_1 = 0$ with probability one (so that the engine is changed with zero probability in the initial period of data), the choice probability $f_{Y_1|M_1, X_1^*}$ would also appear on the RHS of the preceding equations, and additional assumptions regarding this probability would be required to ensure injectivity. However, because these choice probabilities are endogenously determined, it is awkward to impose assumptions directly on it.

8.2 Additional comparison with Kasahara-Shimotsu (2009)

Here, we provide some additional details on the results in Kasahara and Shimotsu (2009) (KS), and show that KS's identification results are not applicable to the dynamic models with time-varying unobservables considered in our paper.

We start by summarizing KS's main results. Throughout, we state KS's results using our notation in this paper. Since KS assume that the unobserved heterogeneity X^* is time-invariant, we attach no t subscript to it.³² Using the notation in our paper, the second equality of KS's Eq. (3) is:

$$\begin{aligned} & f_{Y_1, M_1, \dots, Y_T, M_T} \\ &= \sum_{X^*} f_{X^*} f_{M_1, Y_1 | X^*} \prod_{t=2}^T f_{M_t | M_{t-1}, Y_{t-1}, \dots, M_1, Y_1, X^*} f_{Y_t | M_t, M_{t-1}, Y_{t-1}, X^*}. \end{aligned} \quad (79)$$

In their baseline model (ie. their Assumption 1), they assume that the unobserved heterogeneity X^* does not affect the law of motion for the observed state variable M_t , and that Y_t is independent of (M_{t-1}, Y_{t-1}) conditional on M_t and X^* . This leads to

$$\frac{f_{Y_1, M_1, \dots, Y_T, M_T}}{\prod_{t=2}^T f_{M_t | M_{t-1}, Y_{t-1}}} = \sum_{X^*} f_{X^*} f_{M_1, Y_1 | X^*} \prod_{t=2}^T f_{Y_t | M_t, X^*}, \quad [\text{Eq. (9) in KS (2009)}] \quad (80)$$

³²The correspondence between KS's notation and ours is as follows:

$$\overbrace{\left[\begin{array}{c} a_t, x_t, s_t, m, \pi^m, Q^m(s_t | s_{t-1}) \\ P_t^m(a_t | x_t, x_{t-1}, a_{t-1}) \end{array} \right]}^{\text{KS (2009)}} \Leftrightarrow \overbrace{\left[\begin{array}{c} Y_t, M_t, W_t, X^*, f_{X^*}, f_{W_t | W_{t-1}, X^*} \\ f_{Y_t | M_t, M_{t-1}, Y_{t-1}, X^*} \end{array} \right]}^{\text{our notation}}.$$

which is Eq. (9) in KS. Notice that the LHS of the above is observed, and they demonstrate (in their Proposition 1) that the unknown densities on the RHS are identified from the observed quantity on the LHS for $T \geq 3$.

In section 3.2 of their paper, they consider a first-order Markovian model where the observed variables W_t can depend on W_{t-1} and X^* . They show that, by using $T \geq 6$ periods of data W_1, \dots, W_T , and fixing the values in the odd periods $w_1, w_3, w_5, \dots, w_{T-1}$, one obtains

$$f_{w_1, W_2, w_3, W_4, \dots, w_{T-1}, W_T} = \sum_{X^*} f_{w_1, X^*} \left(\prod_{t=2,4,\dots}^{T-2} f_{w_{t+1}, W_t | w_{t-1}, X^*} \right) f_{W_T | w_{T-1}, X^*}, \quad [\text{Eq. (27) in KS (2009)}] \quad (81)$$

which is Eq. (27) in KS. As they note, Eq. (27) has the same ‘‘independent marginals’’ form as Eq. (9), so that their identification scheme also applies to first-order Markov process with time-invariant X^* for $T \geq 6$. This is their Proposition 6.

However, this scheme no longer works in the case where the latent variable X_t^* varies over time, even if X_t^* is discrete. To see this, we consider a joint first-order Markov process $\{W_t, X_t^*\}$ where both W_t and X_t^* vary over time, as in Example 1 in the main text of our paper. Analogously to Eq. (27) in KS, we may have

$$f_{w_1, W_2, w_3, W_4, \dots, w_{T-1}, W_T} = \sum_{X_{T-1}^*} \dots \sum_{X_5^*} \sum_{X_3^*} \sum_{X_1^*} f_{w_1, X_1^*} \left(\prod_{t=2,4,\dots}^{T-2} f_{w_{t+1}, X_{t+1}^*, W_t | w_{t-1}, X_{t-1}^*} \right) f_{W_T | w_{T-1}, X_{T-1}^*}. \quad (82)$$

Obviously, this takes a very different form than Eq. (27) above, because the components on the RHS involve values of the latent variable X_t^* in different periods. Hence, KS’s identification scheme does not apply here. Notice that using more periods of data only exacerbates the problem; the more periods of data one uses, the more latent variables X_t^* appear when X_t^* is time-varying.

In conclusion, the identification strategy in KS does not apply to models where X_t^* is time-varying, even if X_t^* is discrete. An important innovation of the present paper is that we provide nonparametric identification for dynamic models with time-varying unobserved variables.

8.3 Miscellaneous remarks

8.3.1 Remarks on dynamic investment models

For Example 2 in the main paper, we considered a general investment model in the framework of Doraszelski and Pakes (2007). There is a recent and growing empirical literature based on these types of dynamic models, including Collard-Wexler (2006), Ryan (2006), and Dunne, Klimer, Roberts, and Xu (2006). Pakes (2008, section 3) and Akerberg, Benkard, Berry, and Pakes (2007) discuss additional examples.

On the other hand, the productivity literature has by and large been based on the “pure” investment model, typified by Olley and Pakes (1996) (OP). This model differs in an important way from the types of models considered in our paper. Namely, in OP, capital stock (corresponding to the M variable in Example 2) evolves deterministically, conditional on the previous period’s capital (M_{t-1}) and investment (Y_{t-1}). This feature violates two of our maintained assumptions (# 2,3), which require that M_t depend on X_t^* even conditional on (Y_{t-1}, M_{t-1}) . For this reason, in Example 2 in the main paper, we do not consider the “pure” investment model as in OP, but rather a generalized investment model in which M_t does not evolve deterministically.

8.3.2 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. Consider two scenarios involving an operator equation

$$L_{R_1, r_2, R_4} = L_{R_1 | r_2, R_3} L_{r_2, R_3, R_4}. \quad (83)$$

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

$$L_{R_1 | r_2, R_3}^{-1} L_{R_1, r_2, R_4} = L_{r_2, R_3, R_4}. \quad (84)$$

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 equation (83). We would need the operator L_{r_2, R_3, R_4} to be invertible as follows:

$$L_{R_1, r_2, R_4} L_{r_2, R_3, R_4}^{-1} = L_{R_1 | r_2, R_3}. \quad (85)$$

As proved in Lemma 1 in Hu and Schennach (2008), the sufficient condition for obtaining Eq. (85) from Eq. (83) is that the operator L_{R_4, R_3, r_2} is one-to-one.³³ (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 $L_{V_{t+1}, \bar{w}_t, w_{t-1}, V_{t-2}}^{-1}$ and $L_{V_{t+1}, w_t, \bar{w}_{t-1}, V_{t-2}}^{-1}$ in, respectively, Eqs. (9) and (10). Similarly, Assumption 2(iii) guarantees the validity of post-multiplication by $L_{V_t, w_{t-1}, V_{t-2}}^{-1}$, 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).

8.3.3 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*

$$\frac{\partial}{\partial r_3} f_{R_1|R_3}(r_1|r_3) \neq 0 \text{ almost surely for } \forall r_1 \in \mathcal{S}_1, \forall r_3 \in \mathcal{S}_3. \quad (86)$$

Proof. (Lemma 5) Suppose Eq. (86) fails, so there exists an interval $\mathcal{S}_3 \equiv [\underline{r}, \bar{r}]$ such that, for $\forall r_3 \in \mathcal{S}_3$ and $\forall r_1 \in \mathcal{R}_1$, $\frac{\partial}{\partial r_3} f_{R_1|R_3}(r_1|r_3) = 0$. Define $h_0(r_3) = I_{\mathcal{S}_3}(r_3)g(r_3)$. Then

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

³³A similar assumption is also used in Carroll, Chen, and Hu (2009).

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

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

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

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.

References

- ABBRING, J., P. CHIAPPORI, AND T. ZAVADIL (2008): “Better Safe than Sorry? Ex Ante and Ex Post Moral Hazard in Dynamic Insurance Data,” Tilburg Univeristy, working paper.
- ACKERBERG, D., L. BENKARD, S. BERRY, AND A. PAKES (2007): “Econometric Tools for Analyzing Market Outcomes,” in *Handbook of Econometrics, Vol. 6A*, ed. by J. Heckman, and E. Leamer. North-Holland.
- AGUIREGABIRIA, V., AND P. MIRA (2002): “Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models,” *Econometrica*, 70, 1519–1543.
- (2007): “Sequential Estimation of Dynamic Discrete Games,” *Econometrica*, 75, 1–53.
- ARCIDIACONO, P., AND R. MILLER (2006): “CCP Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity,” Manuscript, Duke University.
- ARELLANO, M. (2003): *Panel Data Econometrics*. Oxford University Press.
- ARELLANO, M., AND B. HONORÉ (2000): “Panel Data Models: Some Recent Developments,” in *Handbook of Econometrics, Vol. 5*. North-Holland.
- AUBIN, J.-P. (2000): *Applied Functional Analysis*. Wiley-Interscience.
- BAJARI, P., L. BENKARD, AND J. LEVIN (2007): “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, 75, 1331–1370.
- BAJARI, P., V. CHERNOZHUKOV, H. HONG, AND D. NEKIPELOV (2007): “Nonparametric and Semiparametric Analysis of a Dynamic Game Model,” Manuscript, University of Minnesota.
- BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): “Nonparametric IV Estimation of Shape-Invariant Engel Curves,” *Econometrica*, 75, 1613–1669.
- BOUISSOU, M., J. J. LAFFONT, AND Q. VUONG (1986): “Tests of Noncausality under Markov Assumptions for Qualitative Panel Data,” *Econometrica*, 54, 395–414.

- BUCHINSKY, M., J. HAHN, AND J. HOTZ (2004): “Estimating Dynamic Discrete Choice Models with Heterogeneous Agents: a Cluster Analysis Approach,” Working Paper, UCLA.
- CARROLL, R., X. CHEN, AND Y. HU (2009): “Identification and estimation of nonlinear models using two samples with nonclassical measurement errors,” *Journal of Nonparametric Statistics*, forthcoming.
- CHEN, X., H. HONG, AND D. NEKIPELOV (2007): “Measurement Error Models,” manuscript, Yale University.
- COLLARD-WEXLER, A. (2006): “Demand Fluctuations and Plant Turnover in the Ready-to-Mix Concrete Industry,” manuscript, New York University.
- CRAWFORD, G., AND M. SHUM (2005): “Uncertainty and Learning in Pharmaceutical Demand,” *Econometrica*, 73, 1137–1174.
- CUNHA, F., J. HECKMAN, AND S. SCHENNACH (2006): “Estimating the Technology of Cognitive and Noncognitive Skill Formation,” manuscript, University of Chicago.
- DAS, S., M. ROBERTS, AND J. TYBOUT (2007): “Market Entry Costs, Producer Heterogeneity, and Export Dynamics,” *Econometrica*, 75, 837–874.
- D’HAULTFOEUILLE, X. (2009): “On the Completeness Condition in Nonparametric Instrumental Problems,” *Econometric Theory*, forthcoming.
- DORASZELSKI, U., AND A. PAKES (2007): “A Framework for Dynamic Analysis in IO,” in *Handbook of Industrial Organization, Vol. 3*, ed. by M. Armstrong, and R. Porter, chap. 30. North-Holland.
- DUNFORD, N., AND J. SCHWARTZ (1971): *Linear Operators*, vol. 3. Wiley.
- DUNNE, T., S. KLIMER, M. ROBERTS, AND D. XU (2006): “Entry and Exit in Geographic Markets,” manuscript, Penn State University.
- ERDEM, T., S. IMAI, AND M. KEANE (2003): “Brand and Quantity Choice Dynamics under Price Uncertainty,” *Quantitative Marketing and Economics*, 1, 5–64.
- ERICSON, R., AND A. PAKES (1995): “Markov-Perfect Industry Dynamics: A Framework for Empirical Work,” *Review of Economic Studies*, 62, 53–82.
- FERNANDEZ-VILLAYERDE, J., AND J. RUBIO-RAMIREZ (2007): “Estimating Macroeconomic Models: A Likelihood Approach,” *Review of Economic Studies*, 74, 1059–1087.
- HENDEL, I., AND A. NEVO (2006): “Measuring the Implications of Sales and Consumer Stockpiling Behavior,” *Econometrica*, 74, 1637–1674.
- HENRY, M., Y. KITAMURA, AND B. SALANIE (2008): “Nonparametric identification of mixtures with exclusion restrictions,” University of Montreal, working paper.
- HONG, H., AND M. SHUM (2009): “Pairwise-Difference Estimation of a Dynamic Optimization Model,” *Review of Economic Studies*, 77, 273–304.
- HOTZ, J., AND R. MILLER (1993): “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *Review of Economic Studies*, 60, 497–529.
- HOTZ, J., R. MILLER, S. SANDERS, AND J. SMITH (1994): “A Simulation Estimator for Dynamic Models of Discrete Choice,” *Review of Economic Studies*, 61, 265–289.

- HOUDE, J.-F., AND S. IMAI (2006): “Identification and 2-step Estimation of DDC models with Unobserved Heterogeneity,” Working Paper, Queen’s University.
- HU, Y. (2008): “Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: a General Solution,” *Journal of Econometrics*, 144, 27–61.
- HU, Y., AND S. SCHENNACH (2008): “Instrumental variable treatment of nonclassical measurement error models,” *Econometrica*, 76, 195–216.
- HU, Y., AND J. SHIU (2009): “Identification and estimation of nonlinear dynamic panel data models with unobserved covariates,” Working paper 557, Johns Hopkins University.
- HU, Y., AND M. SHUM (2009): “Nonparametric Identification of Dynamic Models with Unobserved State Variables, Supplemental Material,” available at <http://www.hss.caltech.edu/~mshum/papers/onlineapp.pdf>.
- IMAI, S., N. JAIN, AND A. CHING (2009): “Bayesian Estimation of Dynamic Discrete Choice Models,” *Econometrica*, 77, 1865–1899.
- KASAHARA, H., AND K. SHIMOTSU (2009): “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choice,” *Econometrica*, 77, 135–175.
- LEVINSOHN, J., AND A. PETRIN (2003): “Estimating Production Functions Using Inputs to Control for Unobservables,” *Review of Economic Studies*, 70, 317–342.
- LI, T. (2002): “Robust and consistent estimation of nonlinear errors-in-variables models,” *Journal of Econometrics*, 110, 1–26.
- MAGNAC, T., AND D. THESMAR (2002): “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, 70, 801–816.
- MATZKIN, R. (2003): “Nonparametric Estimation of Nonadditive Random Functions,” *Econometrica*, 71, 1339–1376.
- MILLER, R. (1984): “Job Matching and Occupational Choice,” *Journal of Political Economy*, 92, 1086–1120.
- NEWHEY, W., AND J. POWELL (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., AND A. PAKES (1996): “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, 64, 1263–1297.
- PAKES, A. (1986): “Patents as Options: Some Estimates of the Value of Holding European Patent Stocks,” *Econometrica*, 54(4), 755–84.
- (2008): “Theory and Empirical Work in Imperfectly Competitive Markets,” Fisher-Schultz Lecture at 2005 Econometric Society World Congress (London, England).
- PAKES, A., M. OSTROVSKY, AND S. BERRY (2007): “Simple Estimators for the Parameters of Discrete Dynamic Games (with Entry Exit Examples),” *RAND Journal of Economics*, 38, 373–399.
- PAKES, A., AND M. SIMPSON (1989): “Patent Renewal Data,” *Brookings Papers on Economic Activity: Microeconomics*, pp. 331–403.

- PESENDORFER, M., AND P. SCHMIDT-DENGLER (2008): “Asymptotic Least Squares Estimators for Dynamic Games,” *Review of Economic Studies*, 75, 901–928.
- RUST, J. (1987): “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55, 999–1033.
- (1994): “Structural Estimation of Markov Decision Processes,” in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle, and D. McFadden, pp. 3082–146. North Holland.
- RYAN, S. (2006): “The Costs of Environmental Regulation in a Concentrated Industry,” manuscript, MIT.
- SCHENNACH, S. (2004): “Estimation of nonlinear models with measurement error,” *Econometrica*, 72, 33–76.
- (2007): “Instrumental variable estimation of nonlinear errors-in-variables models,” *Econometrica*, 75, 201–23.
- XU, D. (2007): “A Structural Empirical Model of R&D, Firm Heterogeneity, and Industry Evolution,” Manuscript, New York University.