Nonparametric Identification of Dynamic Models with Unobserved State Variables*

Yingyao Hu and Matthew Shum
Dept. of Economics, Johns Hopkins University

April 16, 2008
Preliminary draft:
Comments welcome

Abstract

We consider the identification of a Markov process $\{W_t, X_t^*\}$ for $t = 1, 2, ..., T$ when only $\{W_t\}$ for $t = 1, 2, ..., T$ is observed. In structural dynamic models, W_t denotes the sequence of choice variables and observed state variables of an optimizing agent, while X_t^* denotes the sequence of serially correlated unobserved state variables. The Markov setting allows the distribution of the unobserved state variable X_t^* to depend on W_{t-1} and X_{t-1}^*. We show that the Markov transition density $f_{W_t, X_t^*|W_{t-1}, X_{t-1}^*}$ is identified from the observation of five periods of data $W_{t+1}, W_t, W_{t-1}, W_{t-2}, W_{t-3}$ under reasonable assumptions. 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 $\{W_t, X_t^*\}$ for $t = 1, 2, ..., T$ when only $\{W_t\}$ for $t = 1, 2, ..., T$ is observed. In structural dynamic models, W_t denotes the sequence of choice variables and observed state variables of an optimizing agent. X_t^* denotes the sequence of serially correlated unobserved state variables, which are observed by the agent, but not by the econometrician.

*The authors can be reached at yhu@jhu.edu and mshum@jhu.edu. We thank Jeremy Fox, Han Hong, and Susanne Schennach for their suggestions. Seminar participants at LSE, NYU, Penn, and Toulouse provided useful comments.

We show that the Markov transition density $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is identified from the observation of five periods of data $W_{t+1}, W_t, W_{t-1}, W_{t-2}, W_{t-3}$ under reasonable assumptions. In most applications, W_t consists of two elements $W_t = (Y_t, M_t)$, where Y_t denotes the agent's action in period t, and M_t denotes the period-t observed state variable. X_t^* are persistent unobserved state variables (USV for short), which are observed by agents and affect their choice of Y_t, but are unobserved by the econometrician. In turn, the realization of the USV X_t^* can also be affected by Y_{t-1} or M_{t-1}, in addition to X_{t-1}^*. We begin by giving two motivating examples of well-known Markovian dynamic discrete-choice models which have been estimated in the existing literature.

Example 1: Rust (1987) In Rust's bus engine replacement model, Y_t is an indicator for whether Harold Zurcher (the bus depot manager) decides to replace the bus engine in week t. M_t is the accumulated mileage of the bus since the last engine replacement, in week t. Although Rust's original paper had no persistent unobserved state variable X_t^*, it is reasonable to extend the model to allow for them. For example, X_t^* could be Harold Zurcher's health, or weather or road conditions during week t.[1] ■

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

The main result in this paper concerns the identification of the Markov transition density $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$. Once this is known, it can be factorized into conditional and marginal distributions of economic interest. For Markov dynamic choice models (such as the two examples given above; also see Rust (1994)), $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ can be factored into

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

The first term denotes the conditional choice probability for the agent's optimal choice

[1] See Norets (2006), who likewise considers an example of the Rust (1987) model extended to accommodate persistent unobserved state variables.

in period t. In Markov dynamic decision models, the optimal policy function typically depends just on the current state variables (M_t, X_t^*), but not past values.[2] The second term is the Markovian transition probabilities for the state variables (M_t, X_t^*). This setting accommodates quite general feedback in the unobserved state variable process from previous values W_{t-1}, X_{t-1}^* to X_t^*.

Once the CCP's and the state transitions 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), and subsequent methodological developments in this vein include Aguirregabiria and Mira (2002), Pesendorfer and Schmidt-Dengler (2003), Bajari, Benkard, and Levin (2007), Aguirregabiria and Mira (2007), Pakes, Ostrovsky, and Berry (2007), and Hong and Shum (2007).[3] Alternatively, it is possible to use our identification results for the CCP's and state transition densities as a "first-step" in an argument for identification of the per-period utility functions, in the spirit of Magnac and Thesmar (2002) and Bajari, Chernozhukov, Hong, and Nekipelov (2007), who considered the case of dynamic discrete-choice models without serially correlated unobserved state variables.

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

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

[2]See Rust (1994, section 2) for a discussion of optimal policy functions in Markovian dynamic decision models.

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

While these papers have focused on estimation, our focus is on identification. Kasahara and Shimotsu (2007) consider the nonparametric identification of dynamic models when the latent variable X_t^* is time-invariant and discrete. In section 3.2 of their paper, Kasahara and Shimotsu prove the nonparametric identification of the Markov kernel $W_{t+1}|W_t, X^*$ in this setting, using six periods of data. In this paper, we build upon these results to the case where X_t^* is continuous, and can vary over time and evolve depending on (W_{t-1}, X_{t-1}^*).

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

The paper is organized as follows. Section 2 contains our main identification result, which we prove for the case where X_t^* is continuous. We discuss the implications of the identification assumptions in the context of Rust's (1987) bus engine replacement model in Section 3. Section 4 discusses the nonparametric identification of DDC models given the results in section 2. We conclude in Section 5. The appendix includes the proof of the theorem, remarks, and a special case where the unobserved state variable X_t^* is discrete.

2 Nonparametric identification with unobserved state variables

Consider a dynamic process $\left\{\left(W_{t+1}, X_{t+1}^*\right), \left(W_t, X_t^*\right), ..., \left(W_1, X_1^*\right)\right\}_i$ for agent i. We assume that for each agent i, $\left\{\left(W_{t+1}, X_{t+1}^*\right), \left(W_t, X_t^*\right), ..., \left(W_1, X_1^*\right)\right\}_i$ is an independent random draw from the identical distribution $f_{W_{t+1},W_t,...,W_1,X_{t+1}^*,X_t^*,...,X_1^*}$. The variable W_t describes the observed behavior and status of agent i in period t. The variable X_t^* stands for the unobserved state variable at period t. The researcher observes an i.i.d. random sample of the observed component of the process for five periods $\{W_{t+1}, W_t, W_{t-1}, W_{t-2}, W_{t-3}\}_i$, across many agents i.

Let $\mathcal{W}_t \subseteq \mathbb{R}^d$ be the support of W_t and $\mathcal{X}_t^* \subseteq \mathbb{R}$ be the support of X_t^*. Define $\Omega_{<t} = \left\{W_{t-1}, ..., W_1, X_{t-1}^*, ..., X_1^*\right\}$. We assume the dynamic process satisfies:

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

$$f_{W_t,X_t^*|W_{t-1},X_{t-1}^*,\Omega_{<t-1}} = f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}; \tag{2}$$

(ii) Limited feedback:

$$f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} = f_{W_t|W_{t-1},X_t^*}. \tag{3}$$

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

$$\begin{aligned} f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} &= f_{Y_t,M_t|Y_{t-1},M_{t-1},X_t^*,X_{t-1}^*} \\ &= f_{Y_t|M_t,Y_{t-1},M_{t-1},X_t^*,X_{t-1}^*} \cdot f_{M_t|Y_{t-1},M_{t-1},X_t^*,X_{t-1}^*} \\ &= f_{Y_t|M_t,X_t^*,Y_{t-1},M_{t-1}} \cdot f_{M_t|Y_{t-1},M_{t-1},X_t^*}. \end{aligned} \tag{4}$$

In the bottom line of the above display, the limited feedback assumption eliminates X_{t-1}^* as a conditioning variable in both terms. In most applications of Markov dynamic choice models, the first term (corresponding to the CCP) can be further simplified to $f_{Y_t|M_t,X_t^*}$, because the Markovian transition probabilities for the state variables (M_t, X_t^*) imply that the optimal policy function depends just on the current state variables, but not past values. Hence, the above display shows that Assumption 1 imposes weaker restrictions on the first term than typical Markov dynamic optimization models. Moreover, if we move outside the class of Markov dynamic optimization models, Eq. (4) 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.[4]

In the second term of the above display, the limited feedback condition rules out direct

[4]These may include linear or nonlinear panel data models with lagged dependent variables, and serially correlated errors, cf. Arellano and Honore (2000). Arellano (2003, chs. 7–8) considers linear panel models with lagged dependent variables and persistent unobservables, which is also related to our framework.

feedback from last period's unobserved state variable X^*_{t-1} to the current observed state variable X^*_t. However, it allows indirect effects via X^*_{t-1}'s influence on Y_{t-1} or M_{t-1}. Indeed, most empirical applications of dynamic optimization models with unobserved state variables satisfy the Markov and limited feedback conditions above. Examples of models in the industrial organization setting satisfying these conditions include Pakes (1986), Ackerberg (2003), Erdem, Imai, and Keane (2003), Crawford and Shum (2005), Das, Roberts, and Tybout (2007), Xu (2007), and Hendel and Nevo (2007). Finally, note that when X^*_t is time invariant, so that $X^*_t = X^*_{t-1}$, the limited feedback assumption is trivial.

Our goal is to identify the Markov transition density

$$f_{W_t, X^*_t | W_{t-1}, X^*_{t-1}}.$$

Since W_{t+1} is usually a vector and X^*_t is a scalar, we first reduce the dimensionality of W_{t+1} by defining

$$V_{t+1} \equiv g(W_{t+1})$$

where the function $g : \mathbb{R}^d \to \mathbb{R}$ is known. (When W_{t+1} is a scalar, we may just let $g(w) = w$.) The restrictions imposed later on the function g guarantee that the scalar random variable V_{t+1} still contains enough information to identify X^*_t. Similarly, we reduce the dimensionality of W_{t-2} by defining

$$Z_{t-2} \equiv q(W_{t-2}),$$

with a known function $q : \mathbb{R}^d \to \mathbb{R}$. We introduce the function q only for the reason of avoiding technical complications. As we discuss later, we can avoid this, at the expense of introducing additional notation involving the generalized inverse of an operator.

The identification argument consists of four steps. The discussion in this section omits the derivation of some equations. A complete proof, including all derivations, is given in the Appendix.

Step 1: Identification of $\mathbf{f_{V_{t+1}|W_t, X^*_t}}$. The most substantial step of the argument is the first step, which demonstrates the identification of $f_{V_{t+1}|W_t, X^*_t}$. Consider the joint density of $\{V_{t+1}, W_t, W_{t-1}, Z_{t-2}\}$, which is observed in the data. Assumption 1 implies, for any

$(v, w_t, w_{t-1}, z) \in g(\mathcal{W}_{t+1}) \times \mathcal{W}_t \times \mathcal{W}_{t-1} \times q\left(\mathcal{W}_{t-2}\right)$

$$
\begin{aligned}
& f_{V_{t+1},W_t|W_{t-1},Z_{t-2}}\left(v, w_t|w_{t-1}, z\right) \\
= & \int f_{V_{t+1}|W_t,X_t^*}\left(v|w_t, x_t^*\right) f_{W_t|W_{t-1},X_t^*}\left(w_t|w_{t-1}, x_t^*\right) f_{X_t^*|W_{t-1},Z_{t-2}}\left(x_t^*|w_{t-1}, z\right) dx_t^*.
\end{aligned} \tag{5}
$$

Let $\mathcal{L}^p\left(\mathcal{X}\right)$, $1 \leq p < \infty$ stand for the space of function $h\left(\cdot\right)$ with $\int_{\mathcal{X}} |h(x)|^p dx < \infty$. For any $1 \leq p \leq \infty$, we define the integral operator $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} : \mathcal{L}^p\left(q\left(\mathcal{W}_{t-2}\right)\right) \rightarrow \mathcal{L}^p\left(g(\mathcal{W}_{t+1})\right)$ for any given $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$ and any $h \in \mathcal{L}^p\left(q\left(\mathcal{W}_{t-2}\right)\right)$,

$$
\left(L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} h\right)(v) = \int f_{V_{t+1},W_t|W_{t-1},Z_{t-2}}\left(v, w_t|w_{t-1}, z\right) h(z) dz.
$$

Notice that we treat (w_t, w_{t-1}) as fixed and $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$ is a mapping from $\mathcal{L}^p\left(q\left(\mathcal{W}_{t-2}\right)\right)$ to $\mathcal{L}^p\left(g(\mathcal{W}_{t+1})\right)$.

For any given $w_t \in \mathcal{W}_t$, we also define the operator corresponding to the unobserved density $f_{V_{t+1}|W_t,X_t^*}$, i.e., $L_{V_{t+1}|w_t,X_t^*} : \mathcal{L}^p\left(\mathcal{X}_t^*\right) \rightarrow \mathcal{L}^p\left(g(\mathcal{W}_{t+1})\right)$, as follows:

$$
\left(L_{V_{t+1}|w_t,X_t^*} h\right)(v) = \int f_{V_{t+1}|W_t,X_t^*}(v|w_t, x_t^*) h(x_t^*) dx_t^*.
$$

Given this notation, we can rewrite Eq. (5) as

$$
L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}}. \tag{6}
$$

The last two operators on the right-hand side of the above are defined in Eqs. (29) and (34) in the Appendix. It can be shown that the identification of an operator, e.g, $L_{V_{t+1}|w_t,X_t^*}$, is equivalent to that of its corresponding density, e.g., $f_{V_{t+1}|W_t,X_t^*}$. Define for any given $w_t \in \mathcal{W}_t$

$$
\mathcal{A}\left(w_t\right) = \left\{\widetilde{w}_{t-1} \in \mathcal{W}_{t-1} : L_{V_{t+1},w_t|\widetilde{w}_{t-1},Z_{t-2}} \text{ is one-to-one}\right\}.
$$

A one-to-one, or injective, operator can be inverted. Identification of $L_{V_{t+1}|w_t,X_t^*}$ from the observed $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$ requires

Assumption 2 *for any* $w_t \in \mathcal{W}_t$,
(i) $L_{V_{t+1}|w_t,X_t^*}$ *is one-to-one ;*
(ii) $\Pr\left\{\mathcal{A}\left(w_t\right)\right\} > 0$.

Assumption 2(i) implies that the function g, which reduces the dimension of W_t to the scalar $V_{t+1} = g(W_{t+1})$, still contains enough information on X_t^*. A sufficient condition for Assumption 2(i) is that $L_{V_{t+1}|w_t,X_t^*}h = 0$ implies $h = 0$.[5] Notice that Assumption 2(ii) is imposed on the observables.

Remark: The one-to-one assumptions on $L_{V_{t+1}|w_t,X_t^*}$ and $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$ rule out cases where X_t^* has a continuous support, but W_{t+1} has only discrete components. 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 the one-to-one assumption. ■

Remark: If we wish to avoid introducing the function q, then the corresponding operator $L_{V_{t+1},w_t|w_{t-1},W_{t-2}}$ may possibly be surjective. In this case, Assumption 2(ii) may be replaced by the condition that

$$\Pr\left\{\widetilde{w}_{t-1} : L_{V_{t+1},w_t|\widetilde{w}_{t-1},W_{t-2}} L^*_{V_{t+1},w_t|\widetilde{w}_{t-1},W_{t-2}} \text{ is one-to-one}\right\} > 0.$$

where L^* denotes an adjoint operator.[6] We would then need to use the generalized inverse of $L_{V_{t+1},w_t|w_{t-1},W_{t-2}}$ instead of the inverse of $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$. By using $Z_{t-2} = q(W_{t-2})$ and reducing the dimensionality of W_{t-2} to that of X_t^*, we avoid the technical complications of stating assumptions in terms of inner products or adjoint operators. ■

We then assume, in Assumption 3 below, that for any given $w_t \in \mathcal{W}_t$ there exists $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1} \times \mathcal{W}_{t-1}$ such that $\overline{w}_t \neq w_t$, $\overline{w}_{t-1} \neq w_{t-1}$, and that $L_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}}$, $L_{V_{t+1},w_t|\overline{w}_{t-1},Z_{t-2}}$ and $L_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},Z_{t-2}}$ are all one-to-one and invertible mappings. For each of the four pairs of points $(w_t,\ w_{t-1})$, $(w_t,\ \bar{w}_{t-1})$,$(\bar{w}_t,\ w_{t-1})$, and $(\bar{w}_t,\ \bar{w}_{t-1})$, an equation analogous to (6) holds. By manipulating these equations, we obtain

$$\begin{aligned} & L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}} \left(L_{V_{t+1},w_t|\overline{w}_{t-1},Z_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},Z_{t-2}} \right)^{-1} \\ \equiv\ & L_{V_{t+1}|w_t,X_t^*} D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|w_t,X_t^*} \end{aligned} \tag{7}$$

[5] A detailed discussion on one-to-one operators can be found in Carrasco, Florens, and Renault (2005) and Hu and Schennach (2008).

[6] Let $L^*_{x,z} : \mathcal{L}^2(\mathcal{Z}) \to \mathcal{L}^2(\mathcal{X})$ denotes the adjoint operator of operator $L_{x,z} : \mathcal{L}^2(\mathcal{X}) \to \mathcal{L}^2(\mathcal{Z})$ such that $\langle L_{x,z}\varphi, \phi\rangle_{\mathcal{Z}} = \langle \varphi, L^*_{x,z}\phi\rangle_{\mathcal{X}}$, where the inner product is defined as $\langle\varphi,\phi\rangle = \int \varphi(t)\phi(t)\,dt$.

where $D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} : \mathcal{L}^p\left(\mathcal{X}_t^*\right) \to \mathcal{L}^p\left(\mathcal{X}_t^*\right)$

$$\left(D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}g\right)\left(x_t^*\right) = k\left(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*\right)g(x_t^*),$$

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

The operator $D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}$ is a "diagonal" or multiplication operator with a given $(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1})$. This equation implies that the observed operator on the left hand side, which is a mapping from $\mathcal{L}^p\left(g(\mathcal{W}_{t+1})\right) \to \mathcal{L}^p\left(g(\mathcal{W}_{t+1})\right)$, has an eigenvalue-eigenfunction decomposition.[7] The eigenvalues correspond to the elements in $D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}$. The eigenfunctions correspond to $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$, which are normalized because, as densities, they integrate to one: $\int f_{V_{t+1}|W_t,X_t^*}dx_{t+1} = 1$.[8]

The identification of $f_{V_{t+1}|W_t,X_t^*}$ then relies on the uniqueness of the eigenvalue-eigenfunction decomposition in Eq. (7). The next assumption ensures this uniqueness. Define a set $\mathcal{B}\left(w_t\right)$ for a given w_t such that any $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1}) \in \mathcal{B}\left(w_t\right)$ satisfies the following conditions:

1. $\overline{w}_t \in \mathcal{W}_t$; $w_{t-1} \in \mathcal{A}\left(\overline{w}_t\right)$; $\overline{w}_{t-1} \in \mathcal{A}\left(w_t\right) \cap \mathcal{A}\left(\overline{w}_t\right)$; $\overline{w}_t \neq w_t$; and $\overline{w}_{t-1} \neq w_{t-1}$;
2. $k\left(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*\right) < \infty$ for all $x_t^* \in \mathcal{X}_t^*$.

Essentially, for a given $w_t \in \mathcal{W}_t$, the set $\mathcal{B}(w_t)$ contains triples of points $(\bar{w}_t, w_{t-1}, \bar{w}_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1} \times \mathcal{W}_{t-1}$ such that $\overline{w}_t \neq w_t$, $\overline{w}_{t-1} \neq w_{t-1}$, and that $L_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}}$, $L_{V_{t+1},w_t|\overline{w}_{t-1},Z_{t-2}}$ and $L_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},Z_{t-2}}$ are all one-to-one mappings.[9] Furthermore, at these points, the eigenvalues $k\left(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*\right)$ are bounded away from $+\infty$. The boundedness of the eigenvalues allows us to use the results on the spectral decomposition of bounded linear operators in Dunford and Schwartz (1971).

A sufficient condition for $k\left(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*\right) < \infty$ for all $x_t^* \in \mathcal{X}_t^*$ is that, for all $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$, there exist functions $L(w_t, w_{t-1})$ and $U(w_t, w_{t-1})$ such that for all

[7] A similar decomposition is also exploited by Carroll, Chen, and Hu (2008) to identify a general nonlinear model using two samples, when both samples contain nonclassical measurement errors.

[8] Notice that the eigenfunction in $L_{V_{t+1}|w_t,X_t^*}$ does not depend on $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1})$, while the eigenvalue in $D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}$ may be different for a different $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1})$. This suggests that $L_{V_{t+1}|w_t,X_t^*}$ is overidentified, considering multiple values of $(\bar{w}_t, w_{t-1}, \bar{w}_{t-1})$.

[9] Notice that we do not require $w_{t-1} \in \mathcal{A}(w_t)$, so that $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$ need not be one-to-one, because it is not inverted in Eq. (7).

$x_t^* \in \mathcal{X}_t^*$

$$0 < L(w_t, w_{t-1}) \leq f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*) \leq U(w_t, w_{t-1}) < \infty. \tag{8}$$

Formally, the assumption for uniqueness of the decomposition in Eq. 7 is:

Assumption 3 *for any given $w_t \in \mathcal{W}_t$,*
(i) $\Pr\{\mathcal{B}(w_t)\} > 0$;
(ii) for any $\widehat{x}_t^* \neq \widetilde{x}_t^* \in \mathcal{X}_t^*$, there exists $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1}) \in \mathcal{B}(w_t)$ such that

$$k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, \widehat{x}_t^*) \neq k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, \widetilde{x}_t^*).$$

Part (i) of this assumption guarantees that for any given $w_t \in \mathcal{W}_t$, there exists more than one $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1}) \in \mathcal{B}(w_t)$ such that $\overline{w}_t \neq w_t$, $\overline{w}_{t-1} \neq w_{t-1}$, and that $L_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}}$, $L_{V_{t+1},w_t|\overline{w}_{t-1},Z_{t-2}}$ and $L_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},Z_{t-2}}$ are all one-to-one. This validates taking inverses of the operators in Eq. 7.

Part (ii) implies that all the eigenvalues are finite and distinctive for some $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1})$ in Eq. 7. Notice that $\ln k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*)$ can be treated as a second order difference of $\ln f_{W_t|W_{t-1},X_t^*}$ with respect to w_t and w_{t-1}. Therefore, a sufficient condition for part (ii) is that for any $x_t^* \in \mathcal{X}_t^*$ and $w_t \in \mathcal{W}_t$, there exists $w_{t-1} \in \mathcal{W}_{t-1}$ such that

$$\frac{\partial^3}{\partial w_t \partial w_{t-1} \partial x_t^*} \ln f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*) \neq 0. \tag{9}$$

In the case where w_t and w_{t-1} are vector-valued, but contain a continuous component m_t (resp. m_{t-1}), we can modify Eq. (9) appropriately to

$$\frac{\partial^3}{\partial m_t \partial m_{t-1} \partial x_t^*} \ln f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*) \neq 0. \tag{10}$$

Remark: Given the forgoing discussion, Assumptions 2 and 3 may be replaced by the following sufficient conditions:

1. For any $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{W}_{t-1}$, $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$ and $L_{V_{t+1}|w_t,X_t^*}$ are one-to-one ;
2. For any $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{W}_{t-1}$, there exist functions $L(w_t, w_{t-1})$ and $U(w_t, w_{t-1})$ such that the density $f_{W_t|W_{t-1},X_t^*}$ satisfies Eq. (8) for all $x_t^* \in \mathcal{X}_t^*$;

3. For any $w_t \in \mathcal{W}_t$ and $x_t^* \in \mathcal{X}_t^*$, there exists $w_{t-1} \in \mathcal{W}_{t-1}$ such that the density $f_{W_t|W_{t-1},X_t^*}$ satisfies Eq. (9) or Eq. (10). ■

Remark: Since condition 2 in the definition of $\mathcal{B}(w_t)$ must be satisfied for all $w_t \in \mathcal{W}_t$, it will be violated if $f_{W_t|W_{t-1},X_t^*}$ is identically zero for all X_t^*, and all W_{t-1}. However, in practice, most empirical applications of dynamic models avoid this possibility by including i.i.d. shocks which smooth out the CCP's and state transitions in order to avoid zeros, which are inconvenient from a computational point of view. In section 3, we present examples of $f_{W_t|W_{t-1},X_t^*}$ which satisfy Assumption 3. ■.

Without further assumptions, the decomposition in Eq. (7) holds for any permutation of the x_t^* argument. Hence, an eigenfunction $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ for a given w_t is only identified up to the value of the index x_t^*. The following assumption allows us to "pin down" the value of x_t^* corresponding to each eigenfunction:

Assumption 4 *for any given* $w_t \in \mathcal{W}_t$,
(i) There exist 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^*;
(ii) Without loss of generality, we normalize x_t^* *as* $x_t^* = G\left[f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)\right]$.

This assumption pins down the value of x_t^* identified from each eigenfunction $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$, by "ordering" the eigenfunctions using the G functional. Since the value of x_t^* is unobserved, there is no difference between x_t^* and its monotone transformation. Assumption 4 normalizes x_t^* in a way that depends on w_t, which accommodates the fact that X_t^* may be correlated with W_t.

Given Assumption 4, then, the density $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ and operator $L_{V_{t+1}|w_t,X_t^*}$ are nonparametrically identified, for all $w_t \in \mathcal{W}_t$, from the eigenvalue-eigenfunction decomposition in Eq. (7).

Remark: The functional G may map the density f to its mean, mode, median or other quantile, for example, $G[f] = \int x f(x)dx$ or $G[f] = \inf\left\{\widetilde{x}^* : \int_{-\infty}^{\widetilde{x}^*} f(x)dx \geq \tau\right\}$. Moreover, the functional G may depend on w_t. When G corresponds to a quantile, Matzkin (2003) suggests that for a fixed w_t one may have $V_{t+1} = h_{w_t}(X_t^*, \varepsilon)$, where ε is independent of X_t^* and has a standard uniform distribution. The function h_{w_t} can be interpreted as the inverse of the cdf $F_{V_{t+1}|W_t=w_t,X_t^*}$. That implies the τ-th quantile of $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ is $h_{w_t}(x_t^*, \tau)$. Assumption 4 then requires that $h_{w_t}(x_t^*, \tau)$ is monotonic in x_t^* for a known τ. We may then normalize x_t^* as $x_t^* = h_{w_t}(x_t^*, \tau)$ without loss of generality. ■

Step 2: Identification of $\mathbf{f_{W_{t+1}|W_t,X_t^*}}$. In order to identify the density $f_{W_{t+1}|W_t,X_t^*}$, we define the following operators $L_{W_{t+1},w_t|w_{t-1},Z_{t-2}} : \mathcal{L}^p\left(q(\mathcal{W}_{t-2})\right) \to \mathcal{L}^p\left(\mathcal{W}_{t+1}\right)$ and $L_{W_{t+1}|w_t,X_t^*} : \mathcal{L}^p\left(\mathcal{X}_t^*\right) \to \mathcal{L}^p\left(\mathcal{W}_{t+1}\right)$ as

$$
\begin{aligned}
\left(L_{W_{t+1},w_t|w_{t-1},Z_{t-2}}h\right)(x) &= \int f_{W_{t+1},W_t|W_{t-1},Z_{t-2}}(x,w_t|w_{t-1},z)h(z)dz, \\
\left(L_{W_{t+1}|w_t,X_t^*}h\right)(x) &= \int f_{W_{t+1}|W_t,X_t^*}(x|w_t,x_t^*)h(x_t^*)dx_t^*.
\end{aligned}
$$

For any $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{W}_{t-1}$,

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

Hence, by manipulating these equations, we derive that, for any $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{A}(w_t)$,

$$
L_{W_{t+1}|w_t,X_t^*} = L_{W_{t+1},w_t|w_{t-1},Z_{t-2}}L^{-1}_{V_{t+1},w_t|w_{t-1},Z_{t-2}}L_{V_{t+1}|w_t,X_t^*}. \tag{11}
$$

for any $w_t \in \mathcal{W}_t$. The first two elements on the right-hand side are observed in the data, while the last element is identified from Step 1. Hence, $L_{W_{t+1}|w_t,X_t^*}$ and the density $f_{W_{t+1}|W_t,X_t^*}$ are identified.

REMARK: In the time-invariant case where $X_t^* = X^*$, $\forall t$, the conditional density $f_{W_{t+1}|W_t,X^*}$ is the main object of interest, and is enough to permit CCP-based estimation of dynamic discrete-choice models. However, when X_t^* varies over time, knowing $f_{W_{t+1}|W_t,X^*}$ is not enough to permit CCP-based estimation. ■

Step 3: Identification of $\mathbf{f_{W_t,X_t^*,W_{t-1},Z_{t-2}}}$. From inspection of Eq. (5), we see that the observed joint density of $V_{t+1}, W_t, W_{t-1}, W_{t-2}$ can be expressed as

$$
f_{V_{t+1},W_t=w_t,W_{t-1},W_{t-2}} = L_{V_{t+1}|w_t,X_t^*}f_{W_t=w_t,X_t^*,W_{t-1},W_{t-2}}.
$$

for any given $w_t \in \mathcal{W}_t$. With $L_{V_{t+1}|w_t,X_t^*}$ identified in the first step, and its invertibility from Assumption 2, the density $f_{W_t,X_t^*,W_{t-1},W_{t-2}}$ is identified as

$$
f_{W_t=w_t,X_t^*,W_{t-1},W_{t-2}} = L^{-1}_{V_{t+1}|w_t,X_t^*}f_{V_{t+1},W_t=w_t,W_{t-1},W_{t-2}}.
$$

Given the known mapping from W_{t-2} to Z_{t-2}, the identification of $f_{W_t,X_t^*,W_{t-1},W_{t-2}}$ implies that of $f_{W_t,X_t^*,W_{t-1},Z_{t-2}}$. Moreover, because the density of W_{t-1}, Z_{t-2} is identified from the data, the conditional density $f_{W_t,X_t^*|W_{t-1},Z_{t-2}}$ is also identified.

Step 4: Identification of $\mathbf{f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}}$. Finally, we show that the density of interest $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is identified. Assumption 1(i) implies

$$f_{W_t,X_t^*|W_{t-1},Z_{t-2}} = \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^*. \tag{12}$$

The density $f_{W_t,X_t^*|W_{t-1},Z_{t-2}}$ on the left hand side of Eq. 12 is identified from Step 3.

Thus far, we have only used the four observations $\{W_{t+1}, W_t, W_{t-1}, W_{t-2}\}$. In order to identify the density $f_{X_{t-1}^*|W_{t-1},Z_{t-2}}$ on the right hand side of Eq. (12), we use one more period of the data W_{t-3}. Replacing t by $t-1$ in the previous three steps implies that the density of $\{W_t, W_{t-1}, W_{t-2}, W_{t-3}\}$ identifies $f_{W_{t-1},X_{t-1}^*|W_{t-2},Z_{t-3}}$ for $Z_{t-3} = q(W_{t-3})$. In turn, we can identify the density $f_{X_{t-1}^*|W_{t-1},W_{t-2}}$ from $f_{W_{t-1},X_{t-1}^*|W_{t-2},Z_{t-3}}$ as

$$f_{X_{t-1}^*,W_{t-1},W_{t-2}} = \int f_{W_{t-1},X_{t-1}^*|W_{t-2},Z_{t-3}} f_{W_{t-2},Z_{t-3}} dz_{t-3}.$$

Given the known mapping q from W_{t-2} to Z_{t-2}, we can identify $f_{X_{t-1}^*|W_{t-1},Z_{t-2}}$.

Now that the densities $f_{W_t,X_t^*|W_{t-1},Z_{t-2}}$ and $f_{X_{t-1}^*|W_{t-1},Z_{t-2}}$ in Eq. (12) have been identified, the density of interest $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ may be identified under the following assumption. Define the two operators

$$\begin{aligned} L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} &: \mathcal{L}^p\left(\mathcal{X}_{t-1}^*\right) \to \mathcal{L}^p\left(\mathcal{X}_t^*\right), \\ \left(L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} h\right)(x_t^*) &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}(w_t, x_t^*|w_{t-1}, x_{t-1}^*) h(x_{t-1}^*) dx_{t-1}^*. \end{aligned}$$

$$\begin{aligned} L_{X_{t-1}^*|w_{t-1},Z_{t-2}} &: \mathcal{L}^p\left(q(\mathcal{W}_{t-2})\right) \to \mathcal{L}^p\left(\mathcal{X}_{t-1}^*\right), \\ \left(L_{X_{t-1}^*|w_{t-1},Z_{t-2}} h\right)(x_{t-1}^*) &= \int f_{X_{t-1}^*|w_{t-1},Z_{t-2}}(x_{t-1}^*|w_{t-1}, z) h(z) dz. \end{aligned}$$

We assume

Assumption 5 *for any $w_{t-1} \in \mathcal{W}_{t-1}$, $L_{X_{t-1}^*|w_{t-1},Z_{t-2}}$ is one-to-one.*

Under Assumption 5, the operator $L_{X_{t-1}^*|w_{t-1},Z_{t-2}}$ is invertible and, as shown fully in the proof in the Appendix, the desired operator $L_{w_t,X_t^*|w_{t-1},X_{t-1}^*}$, for every $w_t \in \mathcal{W}_t$ and $w_{t-1} \in$

$\mathcal{W}_{t-1}$, is identified as

$$L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} = \left(L^{-1}_{V_{t+1}|w_t,X_t^*} L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}\right) L^{-1}_{X_{t-1}^*|w_{t-1},Z_{t-2}}.$$

We summarize the main identification results in the following theorem:

Theorem 1 *Under the Assumptions 1, 2, 3, 4, and 5, the density* $f_{W_{t+1},W_t,W_{t-1},W_{t-2},W_{t-3}}$ *uniquely determines the density* $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$.

Proof. See appendix. ■

In the case where X_t^* is discrete, the whole identification procedure still holds, which is presented in detail in the appendix. This result implies that the whole dynamic process $\{W_t, X_t^*\}$ is identified even if we only observe $\{W_t\}$. Moreover, the density f_{W_{t-1},X_{t-1}^*} is identified by marginalizing the identified density $f_{W_{t-1},X_{t-1}^*,W_{t-2},Z_{t-3}}$ over the last two arguments, implying that the unconditional density $f_{W_t,X_t^*,W_{t-1},X_{t-1}^*}$ is also identified from $f_{W_{t+1},W_t,W_{t-1},W_{t-2},W_{t-3}}$.

3 Comments on Assumptions in Specific Example: Rust's (1987) Engine Replacement Model

Because some of the assumptions that we made for our identification argument are quite abstract, in this section we discuss these assumptions in the context of a version of Rust's (1987) bus-engine replacement model, augmented to allow for persistent unobserved state variables. As we remarked before, 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.

Because the stylized model we consider here is fully parametric, it may be identified without needing our identification results. However, what we focus on here is not the identifiability of this model, but rather whether data generated from this model would allow us to nonparametrically identify the Markov kernel $W_t, X_t^*|W_{t-1}, X_{t-1}^*$.

We introduce two specifications of the model, which differ in how the unobserved state variable X_t^* enters. In both specifications, we assume that X_t^* evolves as a first-order Markov process, which can depend on past realizations of Y_t and M_t. For technical reasons (as will be clear below), we will restrict X_t^* to have a bounded support: for $[L, U]$ such that $-\infty < L < U < +\infty$,

$$X_t^* = \begin{cases} 0.5X_{t-1}^* + 0.3\psi\left(M_{t-1}\right) + 0.2\nu_t & \text{if } Y_{t-1} = 0 \\ 0.8X_{t-1}^* + 0.2\nu_t & \text{if } Y_{t-1} = 1 \end{cases} \tag{13}$$

with

$$\psi\left(M_{t-1}\right) = L + (U - L)\frac{\exp\left(M_{t-1}\right) - 1}{\exp\left(M_{t-1}\right) + 1}.$$

ν_t is a truncated standard normal shock over the interval $[L, U]$, distributed independently over weeks t, and the $\psi(\cdot)$ function maps mileage $M_{t-1} \in [0, +\infty)$ into $[L, U]$. We also assume that the support of the initial value X_0^* is $[L, U]$, which guarantees that the support of X_t^* is $[L, U]$ for all t. Hence, $X_t^* | X_{t-1}^*, Y_{t-1}, M_{t-1}$ is distributed with density determined by $f_{\nu_t}(\cdot)$. Furthermore, we assume that the characteristic function of ν_t satisfies that $\phi_{\nu_t}(s) \neq 0$ for any real s, which simply requires $L + U \neq 0$. This restriction on ϕ_{ν_t} guarantees that the operator corresponding to the density $f_{X_t^*|X_{t-1}^*, Y_{t-1}, M_{t-1}}$ is one-to-one.

Let $S_t \equiv (M_t, X_t^*)$ denote the persistent state variables in this model. Following Rust (1987), we assume that the single-period utility from each choice is additive in a function of the state variables S_t, and a choice-specific non-persistent preference shock:

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

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.

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

$$\begin{aligned} u_0(S_t) &= -c(M_t) + X_t^* \\ u_1(S_t) &= -RC. \end{aligned} \tag{14}$$

In the above, $c(M_t)$ denotes the maintenance cost function, which is increasing in mileage M_t, and $0 < RC < +\infty$ denotes the cost of replacing the engine. We also assume that the maintenance cost function $c(\cdot)$ is bounded below and above:

$$c(0) = 0; \lim_{M \to +\infty} c(M) = \bar{c} < +\infty.$$

Mileage evolves as:

$$M_{t+1} = \begin{cases} M_t + \eta_{t+1} & \text{if } Y_t = 0 \\ \eta_{t+1} & \text{if } Y_t = 1 \end{cases} \tag{15}$$

where the incremental mileage $\eta_{t+1} > 0$ is a standard normal random variable, truncated to $[0, 1]$, with density

$$\tilde{\phi}(\eta) = \frac{\phi(\eta)}{\Phi(1) - \Phi(0)}. \tag{16}$$

Above, ϕ and Φ denote, respectively, the standard normal density and CDF.[10] independent across weeks, and independent of $(X_t^*, \epsilon_{0t}, \epsilon_{1t})$. ■

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

$$\begin{aligned} u_0(S_t) &= -c(M_t) \\ u_1(S_t) &= -RC. \end{aligned} \tag{17}$$

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

$$M_{t+1} = \begin{cases} M_t + \eta_{t+1} \cdot \exp(X_{t+1}^*) & \text{if } Y_t = 0 \\ \eta_{t+1} \cdot \exp(X_{t+1}^*) & \text{if } Y_t = 1. \end{cases} \tag{18}$$

Here, the incremental mileage $\eta_{t+1} \cdot \exp(X_{t+1}^*)$ is distributed as a mixture of a truncated normal and truncated lognormal distribution. ■

Finally, for the dimension-reducing mappings $g(\cdot)$ and $q(\cdot)$ introduced at the beginning of

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

Section 2, we use:

$$V_{t+1} = g(W_{t+1}) = M_{t+1}$$
$$Z_{t-2} = q(W_{t-2}) = M_{t-2}.$$

That is, the $g(\cdot)$ and $q(\cdot)$ mappings pick out the continuous component of W_t, which is just the mileage M_t.

The main difference between the two specifications is that in Specification A, the unobserved state variable X_t^* affects utilities directly (and therefore the CCP's), but not the mileage process. In Specification B, X_t^* directly affects the evolution of mileage, but not the agent's utilities. We will see that these two specifications differ in how well they satisfy the assumptions of the identification proof.

Given the assumptions so far, the conditional choice probabilities take the multinomial logit form (for $Y_t = 0, 1$):

$$P(Y_t|S_t) = \frac{\exp\left(V_{Y_t}(S_t)\right)}{\sum_{y=0}^{1} \exp\left(V_y(S_t)\right)}$$

where $V_y(S_t)$ is the choice-specific value function in period t, defined recursively by

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

Assumption 1 has already been discussed in much detail thus far, and it is satisfied for both specifications. We now comment on each remaining assumption in turn.

Assumption 2 contains two "injectivity" (or one-to-one) assumptions, and we consider both in some detail. The first requirement is that: for all $w_t \in \mathcal{W}_t$, there exists w_{t-1} such that $L_{M_{t+1},w_t|w_{t-1},M_{t-2}}$ is one-to-one. (Note that we have substituted M_{t+1} for $g(W_{t+1})$, and M_{t-2} for $q(W_{t-2})$.)

Consider Specification A, and consider w_t such that $Y_t = 1$ (so that the engine is replaced in period t). In this case, $M_{t+1}|Y_t = 1$ is truncated normally distributed on $[0, 1]$, and does not depend stochastically on either w_{t-1} or M_{t-2}. Hence, the one-to-one assumption fails.

Now consider Specification B, using the same w_t such that $Y_t = 1$. Because X_t^* directly enters the mileage process, the distribution of M_{t+1} depends on X^*_{t+1}. Similarly, M_{t-2} is a mixture of a truncated lognormal with a truncated normal random variable, and this distribution depends on X^*_{t-2}. Since (X^*_{t+1}, X^*_{t-2}) are correlated, conditional on w_{t-1} (which does not include X^*_{t-1}), the one-to-one assumption should be satisfied.

The second requirement in Assumption 2 requires that, for all w_t, the mapping $L_{M_{t+1}|w_t,X^*_t}$ is one-to-one. As before, consider a value w_t such that $Y_t = 1$. In Specification A, $M_{t+1}|w_t, X^*_t$ is distributed according to a standard normal distribution truncated to $[0, 1]$, regardless of the value of X^*_t. Hence, the one-to-one requirement fails. For Specification B, however, M_{t+1} is distributed according to a mixture distribution which depends on X^*_{t+1}. Given the serial correlation between X^*_{t+1} and X^*_t, the one-to-one assumption should be satisfied.

Assumption 3 concerns the behavior of $f_{W_t|W_{t-1},X^*_{t-1}}$, at fixed values of w_t, w_{t-1} but holding for all values of X^*_t. First, we consider the sufficient condition (8), given right before Assumption 3, that for given (w_t, w_{t-1}), the density $f_{W_t|W_{t-1},X^*_t}$ must be bounded strictly between 0 and $+\infty$. We note that

$$f_{W_t|W_{t-1},X^*_t} = f_{Y_t|M_t,X^*_t} \cdot f_{M_t|Y_{t-1},M_{t-1},X^*_t}. \tag{19}$$

The mileage transition $f_{M_t|Y_{t-1},M_{t-1},X^*_t}$ is a uniform distribution, so it is bounded away from zero and $+\infty$. Moreover, as noted above, the CCP $f_{Y_t|M_t,X^*_t}$ is a logit probability. Because the per-period utilities (under both specification A and B), net of the ϵ's, are bounded away from $-\infty$ and $+\infty$, the logit choice probabilities are also bounded away from zero.

The bounded support assumption on the observed state variable M_t is crucial here. However, in practice, these assumptions on M_t imply very little loss in generality, because typically in estimating these models, one can take the upper and lower bounds on M_t from the observed data.

Second, we consider Eq. (10), which is a sufficient condition which ensures that the eigenvalues in the decomposition (7) are distinctive. Because of the factorization in Eq. (19), and the fact that the choice probabilities are bounded away from zero, a sufficient condition for Eq. (10) is that

$$\frac{\partial^3}{\partial m_t \partial m_{t-1} \partial x^*_t} \ln f_{M_t|Y_{t-1},M_{t-1},X^*_t}(m_t|y_{t-1}, m_{t-1}, x^*_t) \neq 0 \tag{20}$$

for all m_t, x^*_t, and some $w_{t-1} = (y_{t-1}, m_{t-1})$.

For any value of m_t, pick any m_{t-1} such that $y_{t-1} = 0$ (ie., the bus engine was not replaced in period $t-1$). Under Specification B, the density of $M_t|Y_{t-1}, M_{t-1}, X_t^*$ for this pair of (m_t, m_{t-1}), is distributed with density

$$\frac{1}{\exp(x_t^*)} \cdot \tilde{\phi}\left(\frac{m_t - m_{t-1}}{\exp(x_t^*)}\right) \tag{21}$$

on the range $m_t \in [m_{t-1}, m_{t-1} + \exp(x_t^*)]$, where $\tilde{\phi}(\cdot)$ denotes the truncated standard density from Eq. (16). Eq. (20) is satisfied for this density, thus ensuring the distinctiveness of the eigenvalues for Specification B.

On the other hand, for specification A, the sufficient condition cannot be satisfied, because the conditional distribution $M_t|Y_{t-1}, M_{t-1}, X_t^*$ is never a function of x_t^*. Hence, the distinctiveness of the eigenvalues is not assured for this specification.

Assumption 4 presumes a known functional G such that $G\left[f_{M_{t+1}|Y_t,M_t,X_t^*}(\cdot|y_t, m_t, x_t^*)\right]$ is monotonic in x_t^*. Let the functional G map a density f to its median, i.e., $G[f] = \inf\left\{\widetilde{x}^* : \int_{-\infty}^{\widetilde{x}^*} f(x)dx \geq 0.5\right\}$. Eqs. (13) and (18) imply that

$$M_{t+1} = \begin{cases} M_t + \eta_{t+1} \cdot \exp(0.2\nu_{t+1}) \cdot \exp(0.3\psi(M_t)) \cdot \exp(0.5X_t^*) & \text{if } Y_t = 0 \\ \eta_{t+1} \cdot \exp(0.2\nu_{t+1})) \cdot \exp(0.8X_t^*) & \text{if } Y_t = 1. \end{cases} \tag{22}$$

Let constant C_{med} stand for the median of the random variable $\eta_{t+1} \cdot \exp(0.2\nu_{t+1})$, which is a product of a truncated normal and a truncated lognormal random variable. Given the distribution of η_{t+1} and ν_{t+1} and the value of (y_t, m_t), we have

$$G\left[f_{M_{t+1}|Y_t,M_t,X_t^*}(\cdot|y_t, m_t, x_t^*)\right] = \begin{cases} m_t + C_{med} \cdot \exp(0.3\psi(m_t)) \cdot \exp(0.5x_t^*) & \text{if } y_t = 0 \\ C_{med} \cdot \exp(0.8x_t^*) & \text{if } y_t = 1, \end{cases}$$

which is monotonic in x_t^*. The normalization just requires redefining x_t^* according the the equation above in the whole identification procedure.

Assumption 5 requires that, for any w_t, the conditional density $L_{X_{t-1}^*|w_{t-1},M_{t-2}}$ is one-to-one. From inspection of the transition density for the latent process X_t^* in Eq. (13), we see that X_{t-1}^* depends on M_{t-2} if $Y_{t-2} = 0$, but not is $Y_{t-2} = 1$. Generally, the conditional distribution of $X_{t-1}^*|w_{t-1}, M_{t-2}$ will include both observations with $Y_{t-2} = 1$ and $Y_{t-2} = 0$.

Therefore, so long as $P(Y_{t-2} = 0|w_{t-1}, M_{t-2}) > 0$, then the one-to-one assumption should hold.

4 Using the Markov Kernel $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ to Identify DDC models

The identification of the Markov kernel $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ is only the first step in establishing nonparametric identification of the underlying dynamic model. However, once $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ can be identified, nonparametric identification of the remaining parts of the models – particularly, the per-period utility functions – can proceed by straightforward application of the identification results in Magnac and Thesmar (2002) and Bajari, Chernozhukov, Hong, and Nekipelov (2007), which were developed for dynamic models without persistent latent variables X_t^*. In this section, we use the identification arguments in Bajari, Chernozhukov, Hong, and Nekipelov (2007) to show nonparametric identification of the per-period utility functions once the nonparametric identification of $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ has been established.

We make the following assumptions, which are standard in this literature (except for the inclusion of X_t^* as the unobserved state variable):

1. Agents are optimizing in an infinite-horizon, stationary setting. Hence, $W_t, X_t^*|W_{t-1}X_{t-1}^*$ is identical for all periods t. Therefore, in the rest of this section, we use primes $'$'s to denote next-period values.

2. Actions Y are chosen from the set $\mathcal{Y} = \{0, 1, \ldots, K\}$.

3. The state variables are $S \equiv (M, X^*)$.

4. The per-period utility from taking action $y \in \mathcal{Y}$ in period t is:

   $$u_y(S_t) + \epsilon_{y,t}, \ \forall y \in \mathcal{Y}.$$

   The $\epsilon_{y,t}$'s are utility shocks which are independent of S_t, and distributed i.i.d with known distribution $F(\epsilon)$ across periods t and actions y. Let $\vec{\epsilon}_t \equiv (\epsilon_{0,1}, \epsilon_{1,t}, \ldots, \epsilon_{K,t})$.

5. From the data, the CCP's

$$p_y(S) \equiv \text{Prob}(Y = 1|S),$$

and the Markov transition kernel for S, denoted $p(S'|Y,S)$, are identified. Nonparametric identification of these two elements was the main result demonstrated in Section 2 of this paper.

6. $u_0(S)$, the per-period utility from $Y = 0$, is normalized to zero, for all S.

7. β, the discount factor, is known.[11]

Following the arguments in Magnac and Thesmar (2002) and Bajari, Chernozhukov, Hong, and Nekipelov (2007), we will show the nonparametric identification of $u_y(\cdot),\ y = 1,\dots,K$, the per-period utility functions for all action except $Y = 0$.

The Bellman equation for this dynamic optimization problem is

$$V(S,\vec{\epsilon}) = \max_{y\in\mathcal{Y}} \left(u_y(S) + \epsilon_y + \beta E_{S',\vec{\epsilon}'|Y,S} V(S',\vec{\epsilon}') \right)$$

where $V(S,\vec{\epsilon})$ denotes the value function. We define the choice-specific value function as

$$V_y(S) \equiv u_y(S) + \beta E_{S',\vec{\epsilon}'|Y,S} V(S',\vec{\epsilon}').$$

Given these definitions, an agent's optimal choice when the state is S is given by

$$y^*(S) = \text{argmax}_{y\in\mathcal{Y}} \left(V_y(S) + \epsilon_y \right).$$

Hotz and Miller (1993) and Magnac and Thesmar (2002) show that in this setting, there is a known one-to-one mapping, $q(S) : \mathbb{R}^K \to \mathbb{R}^K$, which maps the K-vector of choice probabilities $(p_1(S),\dots,p_K(S))$ to the K-vector $(\Delta_1(S),\dots,\Delta_K(S))$, where $\Delta_y(S)$ denotes the difference in choice-specific value functions

$$\Delta_y(S) \equiv V_y(S) - V_0(S).$$

Let the i-th element of $q(p_1(S),\dots,p_K(S))$, denoted $q_i(S)$, be equal to $\Delta_i(S)$. The known mapping q derives just from $F(\epsilon)$, the known distribution of the utility shocks.

[11] Magnac and Thesmar (2002) discuss the possibility of identifying β via exclusion restrictions, but we do not pursue that here.

Hence, since the choice probabilities can be identified from the data, and the mapping q is known, the value function differences $\Delta_1(S), \ldots, \Delta_K(S)$ is also known.

Next, we note that the choice-specific value function also satisfies a Bellman-like equation:

$$\begin{aligned} V_y(S) &= u_y(S) + \beta E_{S'|S,Y} \left[E_{\vec{\epsilon}'} \max_{y' \in \mathcal{Y}} (V_{y'}(S') + \epsilon'_y) \right] \\ &= u_y(S) + \beta E_{S'|S,Y} \left[H(\Delta_1(S'), \ldots, \Delta_K(S')) + V_0(S') \right] \end{aligned} \tag{23}$$

where $H(\cdots)$ denotes McFadden's "social surplus" function, for random utility models (cf. Rust (1994, pp. 3104ff)). Like the q mapping, H is a known function, which depends just on $F(\epsilon)$, the known distribution of the utility shocks.

Using the assumption that $u_0(S) = 0, \ \forall S$, the Bellman equation for $V_0(S)$ is

$$V_0(S) = \beta E_{S'|S,Y} \left[H(\Delta_1(S'), \ldots, \Delta_K(S')) + V_0(S') \right]. \tag{24}$$

In this equation, everything is known (including, importantly, the distribution of $S'|S, Y$), except the $V_0(\cdot)$ function. Hence, by iterating over Eq. (24), we can recover the $V_0(S)$ function. Once $V_0(\cdot)$ is known, the other choice-specific value functions can be recovered as

$$V_y(S) = \Delta_y(S) + V_0(S), \ \forall y \in \mathcal{Y}, \ \forall S.$$

Finally, the per-period utility functions $u_y(S)$ can be recovered from the choice-specific value functions as

$$u_y(S) = V_y(S) - \beta E_{S'|S,Y} \left[H(\Delta_1(S'), \ldots, \Delta_K(S')) + V_0(S') \right], \ \forall y \in \mathcal{Y}, \ \forall S,$$

where everything on the right-hand side is known.

Remark: For the case where $F(\epsilon)$ is the Type 1 Extreme Value distribution, the social surplus function is

$$H(\Delta_1(S), \ldots, \Delta_K(S)) = \log \left[1 + \sum_{y=1}^{K} \exp(\Delta_y(S)) \right]$$

and the mapping q is such that

$$q_y(S) = \Delta_y(S) = \log(p_y(S)) - \log(p_0(S)), \ \forall y = 1, \ldots K,$$

where $p_0(S) \equiv 1 - \sum_{y=1}^{K} p_y(S)$. ■

5 Concluding remarks

In this paper, we have considered the identification of a Markov process $\{W_t, X_t^*\}$ for $t = 1, 2, ..., T$ when only $\{W_t\}$ for $t = 1, 2, ..., T$ is observed. We showed that the Markov transition density $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ is identified from the distribution of the five observations $W_{t+1}, W_t, W_{t-1}, W_{t-2}, W_{t-3}$ under reasonable assumptions. Identification of $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ is a crucial input in methodologies for estimating dynamic models based on the "conditional-choice-probability (CCP)" approach pioneered by Hotz and Miller.

In the identification arguments, we have not invoked a stationarity assumption, which would require that the $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ be invariant across periods t. Because of this, our identification argument works in both stationary and non-stationary settings. One caveat is that, because we require the five observations $W_{t+1}, W_t, W_{t-1}, W_{t-2}, W_{t-3}$ to identify $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ for every t, we would only be able to identify $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ from period $t = 4, \ldots T - 1$.

Another assumption we made is that the unobserved state variable X_t^* is scalar-valued. We believe the proof can be extended to cases where X_t^* is a multivariate process. This may enable our identification procedure to be applied to dynamic game settings, where M_t and X_t^* may contain the set of, respectively, observed and unobserved state variables for all agents in the game.

Finally, this paper has focused on identification, but not estimation. While our identification proof is constructive, and can be mimicked directly for estimation, it is cumbersome to invert the functional operators computationally. For this reason, it may be more convenient to estimate using a semi-nonparametric sieve maximum likelihood procedure (Carroll, Chen, and Hu (2008)). In ongoing work, we are applying our identification results to estimate dynamic discrete-choice models with unobserved state variables.

A Proofs

Proof. (Theorem 1)

As in the main text, the identification proof proceeds in four steps, which are labeled here in the proof for convenience.

Step 1 Assumption 1(i) implies that

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

(For simplicity, we omit all the arguments in the density functions.) Assumption 1(ii) then implies that

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

Hence, by combining the above two displays, we obtain that

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

Eq. (25) implies an equality between corresponding operators. Let $\mathcal{L}^p(\mathcal{X})$, $1 \leq p < \infty$ stand for the space of function $h(\cdot)$ with $\int_{\mathcal{X}} |h(x)|^p dx < \infty$, and let $\mathcal{L}^\infty(\mathcal{X})$ denote the space of function $h(\cdot)$ with $\sup_{x\in\mathcal{X}} |h(x)| < \infty$. For any $1 \leq p \leq \infty$, we define operators as follows: for any function $h \in \mathcal{L}^p(\mathcal{W}_{t-m})$

$$\begin{aligned} L_{W_{t+1},w_t|w_{t-1},W_{t-2}} &: \mathcal{L}^p(\mathcal{W}_{t-2}) \rightarrow \mathcal{L}^p(\mathcal{W}_{t+1}), \\ \left(L_{W_{t+1},w_t|w_{t-1},W_{t-2}}h\right)(x) &= \int f_{W_{t+1},W_t|W_{t-1},W_{t-2}}(x, w_t|w_{t-1}, z)h(z)dz, \end{aligned} \tag{26}$$

$$\begin{aligned} L_{W_{t+1}|w_t,X_t^*} &: \mathcal{L}^p(\mathcal{X}_t^*) \rightarrow \mathcal{L}^p(\mathcal{W}_{t+1}), \\ \left(L_{W_{t+1}|w_t,X_t^*}h\right)(x) &= \int f_{W_{t+1}|W_t,X_t^*}(x|w_t, x_t^*)h(x_t^*)dx_t^*, \end{aligned} \tag{27}$$

$$\begin{aligned} L_{X_t^*|w_{t-1},W_{t-2}} &: \mathcal{L}^p(\mathcal{W}_{t-2}) \rightarrow \mathcal{L}^p(\mathcal{X}_t^*), \\ \left(L_{X_t^*|w_{t-1},W_{t-2}}h\right)(x_t^*) &= \int f_{X_t^*|W_{t-1},W_{t-2}}(x_t^*|w_{t-1}, z)h(z)dz, \end{aligned} \tag{28}$$

$$\begin{aligned} D_{w_t|w_{t-1},X_t^*} &: \mathcal{L}^p(\mathcal{X}_t^*) \rightarrow \mathcal{L}^p(\mathcal{X}_t^*), \\ \left(D_{w_t|w_{t-1},X_t^*}h\right)(x_t^*) &= f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*)h(x_t^*). \end{aligned} \tag{29}$$

The operator $D_{w_t|w_{t-1},X_t^*}$ is a "diagonal" or multiplication operator. As shown in Hu and Schennach (2008), the identification of an operator, e.g, $L_{V_{t+1}|w_t,X_t^*}$, is equivalent to that of its corresponding density, e.g., $f_{V_{t+1}|W_t,X_t^*}$. For any given $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$, we have for any function $h \in \mathcal{L}^p(\mathcal{W}_{t-2})$

$$\begin{aligned} & \left(L_{W_{t+1},w_t|w_{t-1},W_{t-2}}h\right)(x) \\ = & \int f_{W_{t+1},W_t|W_{t-1},W_{t-2}}(x, w_t|w_{t-1}, z)h(z)dz \\ = & \int f_{W_{t+1}|W_t,X_t^*}(x|w_t, x_t^*)f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*)\left(\int f_{X_t^*|W_{t-1},W_{t-2}}(x_t^*|w_{t-1}, z)h(z)dz\right)dx_t^* \\ = & \int f_{W_{t+1}|W_t,X_t^*}(x|w_t, x_t^*)f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*)\left(L_{X_t^*|w_{t-1},W_{t-2}}h\right)(x_t^*)\,dx_t^* \\ = & \int f_{W_{t+1}|W_t,X_t^*}(x|w_t, x_t^*)\left(D_{w_t|w_{t-1},X_t^*}L_{X_t^*|w_{t-1},W_{t-2}}h\right)(x_t^*)\,dx_t^* \\ = & \left(L_{W_{t+1}|w_t,X_t^*}D_{w_t|w_{t-1},X_t^*}L_{X_t^*|w_{t-1},W_{t-2}}h\right)(x). \end{aligned}$$

Therefore, Eq. (25) is equivalent to

$$L_{W_{t+1},w_t|w_{t-1},W_{t-2}} = L_{W_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},W_{t-2}}. \tag{30}$$

As discussed in the main text, we introduce the two dimension-reducing functions g ,q : $\mathbb{R}^d \to \mathbb{R}$, and

$$\begin{aligned} V_{t+1} &= g(W_{t+1}), \\ Z_{t-2} &= q\left(W_{t-2}\right). \end{aligned}$$

Analogously to Eq. (30), the joint density of $\{V_{t+1}, W_t, W_{t-1}, Z_{t-2}\}$ can be expressed in operator notation, for any $(x, w_t, w_{t-1}, z) \in g\left(\mathcal{W}_{t+1}\right) \times \mathcal{W}_t \times \mathcal{W}_{t-1} \times q(\mathcal{W}_{t-2})$, as

$$L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}}, \tag{31}$$

where

$$\begin{aligned} L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} &: \mathcal{L}^p\left(q\left(\mathcal{W}_{t-2}\right)\right) \to \mathcal{L}^p\left(g(\mathcal{W}_{t+1})\right), \\ \left(L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} h\right)(v) &= \int f_{V_{t+1},W_t|W_{t-1},Z_{t-2}}\left(v, w_t|w_{t-1}, z\right) h(z)dz, \end{aligned} \tag{32}$$

$$\begin{aligned} L_{V_{t+1}|w_t,X_t^*} &: \mathcal{L}^p\left(\mathcal{X}_t^*\right) \to \mathcal{L}^p\left(g(\mathcal{W}_{t+1})\right), \\ \left(L_{V_{t+1}|w_t,X_t^*} h\right)(v) &= \int f_{V_{t+1}|W_t,X_t^*}(v|w_t, x_t^*) h(x_t^*) dx_t^*, \end{aligned} \tag{33}$$

$$\begin{aligned} L_{X_t^*|w_{t-1},Z_{t-2}} &: \mathcal{L}^p\left(q\left(\mathcal{W}_{t-2}\right)\right) \to \mathcal{L}^p\left(\mathcal{X}_t^*\right), \\ \left(L_{X_t^*|w_{t-1},Z_{t-2}} h\right)(x_t^*) &= \int f_{X_t^*|W_{t-1},Z_{t-2}}(x_t^*|w_{t-1}, z) h(z)dz. \end{aligned} \tag{34}$$

The operator $L_{V_{t+1}|w_t,X_t^*}$ does not depend on w_{t-1} and $L_{X_t^*|w_{t-1},Z_{t-2}}$ does not depend on w_t. This important feature helps the identification of $L_{V_{t+1}|w_t,X_t^*}$ in Eq. (31).

For any $w_t \in \mathcal{W}_t$, we consider the points $(\bar{w}_t, w_{t-1}, \bar{w}_{t-1})$ satisfying

$\overline{w}_t \in \mathcal{W}_t,\ w_{t-1} \in \mathcal{A}\left(\overline{w}_t\right),\ \overline{w}_{t-1} \in \mathcal{A}\left(w_t\right) \cap \mathcal{A}\left(\overline{w}_t\right),\ \overline{w}_{t-1} \neq w_{t-1}$, and $\overline{w}_t \neq w_t$. For these points, an operator equality analogous to Eq. (30) holds:

$$\begin{aligned}
\text{for } (w_t, w_{t-1}) \quad &: \quad L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}}, &(35)\\
\text{for } (\overline{w}_t, w_{t-1}) \quad &: \quad L_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}} = L_{V_{t+1}|\overline{w}_t,X_t^*} D_{\overline{w}_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}}, &(36)\\
\text{for } (w_t, \overline{w}_{t-1}) \quad &: \quad L_{V_{t+1},w_t|\overline{w}_{t-1},Z_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} L_{X_t^*|\overline{w}_{t-1},Z_{t-2}}, &(37)\\
\text{for } (\overline{w}_t, \overline{w}_{t-1}) \quad &: \quad L_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},Z_{t-2}} = L_{V_{t+1}|\overline{w}_t,X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L_{X_t^*|\overline{w}_{t-1},Z_{t-2}}. &(38)
\end{aligned}$$

Assumptions 2 and 3 guarantee that bottom three left-hand side operators can be inverted. Postmultiplying Eq. (35) by the inverse of Eq. (36) leads to

$$\begin{aligned}
\mathbf{A} \quad &\equiv \quad L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}} \\
&= \quad L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}} \left(L_{V_{t+1}|\overline{w}_t,X_t^*} D_{\overline{w}_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}} \right)^{-1} \\
&= \quad L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*}, \qquad (39)
\end{aligned}$$

which eliminates $L_{X_t^*|w_{t-1},Z_{t-2}}$. Similarly, eliminating $L_{X_t^*|\overline{w}_{t-1},Z_{t-2}}$ in Eqs. (37) and (38) results in

$$\begin{aligned}
\mathbf{B} \quad &\equiv \quad L_{V_{t+1},w_t|\overline{w}_{t-1},Z_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},Z_{t-2}} \\
&= \quad L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} D^{-1}_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*}. \qquad (40)
\end{aligned}$$

We then eliminate $L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*}$ in Eqs. (39) and (40) to obtain

$$\begin{aligned}
\mathbf{AB}^{-1} \quad &\equiv \quad L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}} \left(L_{V_{t+1},w_t|\overline{w}_{t-1},Z_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},Z_{t-2}} \right)^{-1} \\
&= \quad L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*} \times \\
&\quad\ \times \left(L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} D^{-1}_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*} \right)^{-1} \\
&= \quad L_{V_{t+1}|w_t,X_t^*} \left(D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*} \right) L^{-1}_{V_{t+1}|w_t,X_t^*} \\
&\equiv \quad L_{V_{t+1}|w_t,X_t^*} D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|w_t,X_t^*}, \qquad (41)
\end{aligned}$$

where

$$\begin{aligned}
& \left(D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} h \right) (x_t^*) \\
= \quad & \left(D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*} h \right) (x_t^*) \\
= \quad & k\left(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, x_t^*\right) h\left(x_t^*\right),
\end{aligned}$$

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

This equation implies that the observed operator $\mathbf{AB}^{-1}$ on the left hand side of Eq. (41) has an inherent eigenvalue-eigenfunction decomposition. The eigenfunctions are $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t,x_t^*)$, which is normalized by $\int f_{V_{t+1}|W_t,X_t^*}(x|w_t,x_t^*)dx=1$.

The decomposition in Eq. (41) is similar to but more complicated than the decomposition in Hu and Schennach (2008) or Carroll, Chen, and Hu (2008). Their results imply that such a decomposition is unique under Assumptions 3 and 4. We may show the reasoning as follows.

The eigenfunctions in $L_{V_{t+1}|W_t,X_t^*}$ do not depend on $\overline{w}_t$, w_{t-1}, or $\overline{w}_{t-1}$, while the eigenvalues in $D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}$ may be different for different values of $\overline{w}_t$, w_{t-1}, or $\overline{w}_{t-1}$. Suppose that for two indices $\widehat{x}_t^*\neq\widetilde{x}_t^*$, the two eigenvalues are the same, i.e., $k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},\widehat{x}_t^*\right)=k\left(w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},\widetilde{x}_t^*\right)$ for some $(\overline{w}_t,w_{t-1},\overline{w}_{t-1})\in\mathcal{W}_t\times\mathcal{W}_{t-1}\times\mathcal{W}_{t-1}$. In this circumstance, we cannot identify the two corresponding eigenfunctions. But Assumption 3 guarantees that there exists another set of points $(\widehat{w}_t,\widetilde{w}_{t-1},\widehat{w}_{t-1})\in\mathcal{W}_t\times\mathcal{W}_{t-1}\times\mathcal{W}_{t-1}$ such that the decomposition (41) evaluated at these points yields the distinctive eigenvalues $k\left(w_t,\widehat{w}_t,\widetilde{w}_{t-1},\widehat{w}_{t-1},\widehat{x}_t^*\right)\neq k\left(w_t,\widehat{w}_t,\widetilde{w}_{t-1},\widehat{w}_{t-1},\widetilde{x}_t^*\right)$. These eigenvalues enable the determination of $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t,\widehat{x}_t^*)$ and $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t,\widetilde{x}_t^*)$, the eigenfunctions for $\widehat{x}_t^*$ and $\widetilde{x}_t^*$.

Therefore, Assumption (3) implies that the eigenfunctions $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t,x_t^*)$ are identified up to the value of x_t^* for any given $w_t\in\mathcal{W}_t$. Moreover, Assumption 4 pins down the value of x_t^* in each eigenfunction $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t,x_t^*)$. 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 decomposition in Eq. (41).

Step 2 We show the identification of the density $f_{W_{t+1}|W_t,X_t^*}$. Eqs. (25) and (31) imply for any given $w_t\in\mathcal{W}_t$, and $w_{t-1}\in\mathcal{A}(w_t)$

$$\begin{aligned}
& L_{W_{t+1},w_t|w_{t-1},Z_{t-2}}L^{-1}_{V_{t+1},w_t|w_{t-1},Z_{t-2}}L_{V_{t+1}|w_t,X_t^*}\\
= \; & L_{W_{t+1}|w_t,X_t^*}D_{w_t|w_{t-1},X_t^*}L_{X_t^*|w_{t-1},Z_{t-2}}\left(L_{V_{t+1}|w_t,X_t^*}D_{w_t|w_{t-1},X_t^*}L_{X_t^*|w_{t-1},Z_{t-2}}\right)^{-1}L_{V_{t+1}|w_t,X_t^*}\\
= \; & L_{W_{t+1}|w_t,X_t^*},
\end{aligned}$$

where the left hand side is identified.

Step 3 Moreover, the following equation

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

implies that for any given $w_t \in \mathcal{W}_t$,

$$f_{V_{t+1},W_t=w_t,W_{t-1},W_{t-2}} = L_{V_{t+1}|w_t,X_t^*} f_{W_t=w_t,X_t^*,W_{t-1},W_{t-2}}.$$

Therefore, we identify $f_{W_t=w_t,X_t^*,W_{t-1},W_{t-2}}$ for any given $w_t \in \mathcal{W}_t$ through

$$f_{W_t=w_t,X_t^*,W_{t-1},W_{t-2}} = L^{-1}_{V_{t+1}|w_t,X_t^*} f_{V_{t+1},W_t=w_t,W_{t-1},W_{t-2}}.$$

In summary, the densities $f_{W_{t+1}|W_t,X_t^*}$ and $f_{W_t,X_t^*,W_{t-1},W_{t-2}}$ are identified from $f_{W_{t+1},W_t,W_{t-1},W_{t-2}}$. Given the known function q in $Z_{t-2} = q\left(W_{t-2}\right)$, the identification of $f_{W_t,X_t^*,W_{t-1},W_{t-2}}$ implies that of $f_{W_t,X_t^*,W_{t-1},Z_{t-2}}$. Moreover, because the joint densities of (W_{t-1}, Z_{t-2}) and (W_{t-1}, W_{t-2}) are observed from the data, the conditional densities $f_{W_t,X_t^*|W_{t-1},Z_{t-2}}$ and $f_{W_t,X_t^*|W_{t-1},W_{t-2}}$ are also identified.

Step 4 In this final step, we show the identification of the density $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$. Assumption 1 implies

$$\begin{aligned} f_{W_t,X_t^*|W_{t-1},Z_{t-2}} &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*,Z_{t-2}} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\ &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^*. \end{aligned} \tag{42}$$

The left hand side has been identified in the previous step. Thus far, we have only used the four observations $W_{t+1}, W_t, W_{t-1}, W_{t-2}$. In order to identify the density $f_{X_{t-1}^*|W_{t-1},Z_{t-2}}$ on the right hand side of Eq. (42), we use one more period of the data, W_{t-3} or Z_{t-3}.

By repeating the argument in the previous three steps, but lagged one period, we can identify $f_{V_t|W_{t-1},X_{t-1}^*}$ and $f_{W_{t-1},X_{t-1}^*|W_{t-2},Z_{t-3}}$ from the observed density $f_{V_t,W_{t-1}|W_{t-2},Z_{t-3}}$. Therefore, $f_{X_{t-1}^*,W_{t-1},Z_{t-2}}$ is identified from

$$f_{X_{t-1}^*,W_{t-1},W_{t-2}} = \int f_{W_{t-1},X_{t-1}^*|W_{t-2},Z_{t-3}} f_{W_{t-2},Z_{t-3}} dz_{t-3}$$

with $Z_{t-2} = q\left(W_{t-2}\right)$ for the known function q. Because (W_{t-1}, Z_{t-2}) is observed in the data, the conditional density $f_{X_{t-1}^*|W_{t-1},Z_{t-2}}$ is also identified.

To proceed, we derive an operator equality corresponding to Eq. (42). For the left-hand side of this equation, we have

$$\begin{aligned}
f_{W_t,X_t^*|W_{t-1},Z_{t-2}} &= \int f_{W_t,X_t^*,X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\
&= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*,Z_{t-2}} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\
&= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\
&= \int f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} f_{X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\
&= \int f_{W_t|W_{t-1},X_t^*} f_{X_t^*,X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\
&= f_{W_t|W_{t-1},X_t^*} f_{X_t^*|W_{t-1},Z_{t-2}}. \qquad (43)
\end{aligned}$$

The operator corresponding to $f_{W_t|W_{t-1},X_t^*} f_{X_t^*|W_{t-1},Z_{t-2}}$ is $D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}}$, for given w_t, w_{t-1}. Eq. (31) above implies that

$$D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}} = L_{V_{t+1}|w_t,X_t^*}^{-1} L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}. \qquad (44)$$

Both elements on the right-hand side have been identified.

For the right-hand side of Eq. (42), we define the oeprators

$$\begin{aligned}
L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} &: \mathcal{L}^p\left(\mathcal{X}_{t-1}^*\right) \to \mathcal{L}^p\left(\mathcal{X}_t^*\right), \\
\left(L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} h\right)(x_t^*) &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}(w_t, x_t^*|w_{t-1}, x_{t-1}^*) h(x_{t-1}^*) dx_{t-1}^*
\end{aligned}$$

and

$$\begin{aligned}
L_{X_{t-1}^*|w_{t-1},Z_{t-2}} &: \mathcal{L}^p\left(q(\mathcal{W}_{t-2})\right) \to \mathcal{L}^p\left(\mathcal{X}_{t-1}^*\right), \\
\left(L_{X_{t-1}^*|w_{t-1},Z_{t-2}} h\right)(x_{t-1}^*) &= \int f_{X_{t-1}^*|W_{t-1},Z_{t-2}}(x_{t-1}^*|w_{t-1}, z) h(z) dz.
\end{aligned}$$

The operator $L_{X_{t-1}^*|w_{t-1},Z_{t-2}}$ is identified because its corresponding density $f_{X_{t-1}^*|w_{t-1},Z_{t-2}}$ has been identified before.

Hence, using Eq. (43), we can express Eq. (42) in operator notation as

$$D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}} = L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*|w_{t-1},Z_{t-2}}. \qquad (45)$$

Combining Eqs. (44) and (45) leads to

$$L_{w_t,X_t^*|w_{t-1},X_{t-1}^*}L_{X_{t-1}^*|w_{t-1},Z_{t-2}} = L_{V_{t+1}|w_t,X_t^*}^{-1}L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}. \tag{46}$$

By Assumption 5, the second element on the left-hand side is invertible. Hence, for all $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{W}_{t-1}$, the desired Markov transition density $f_{W_t=w_t,X_t^*|W_{t-1}=w_{t-1},X_{t-1}^*}$ and $L_{w_t,X_t^*|w_{t-1},X_{t-1}^*}$ are identified as

$$L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} = \left(L_{V_{t+1}|w_t,X_t^*}^{-1}L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}\right)L_{X_{t-1}^*|w_{t-1},Z_{t-2}}^{-1}.$$

■

B Special case: a discrete unobserved state variable

In this section, we illustrate our identification strategy in the special case where X_t^* is discrete:

$$\forall t,\ X_t^* \in \mathcal{X}^* \equiv \{1, 2, \ldots, J\}.$$

The main difference between this discrete case and the previous continuous case is that the linear integral operators are replaced by matrices, which may be more straightforward.

Since we assume the unobserved state variable X_t^* is discrete in this section, we first discretize the observed variable W_t and then use the discretized W_t to identify the distribution involving the latent X_t^*. Let $\mathcal{W}_t$ be the support of W_t and $\mathcal{W}_t^1$, $\mathcal{W}_t^2$, ..., $\mathcal{W}_t^J$ be a known partition of $\mathcal{W}_t$. We define a discrete variable $V_t \in \mathcal{X}_t^*$ such that $V_t = j$ if $W_t \in \mathcal{W}_t^j$, i.e.,

$$V_t = \sum_{j=1}^{J} j \times I\left(W_t \in \mathcal{W}_t^j\right),$$

where $I(\cdot)$ is the indicator function. This mapping corresponds to the known functions g and q in the continuous case, which also implies we use V_{t-2} as Z_{t-2} in the continuous case.

Given the proof of theorem 1, a number of equations and derivations are stated without proof in this section.

Step 1: Identification of $\mathbf{f_{V_{t+1}|W_t,X_t^*}}$. Eqs. (2) and (3) implies for any $v, z \in \mathcal{X}_t^*$, $w_t \in \mathcal{W}_t$, and $w_{t-1} \in \mathcal{W}_{t-1}$,

$$\begin{aligned} & f_{V_{t+1},W_t|W_{t-1},V_{t-2}}(v, w_t|w_{t-1}, z) \\ = & \sum_{x_t^* \in \mathcal{X}_t^*} f_{V_{t+1}|W_t,X_t^*}(v|w_t, x_t^*)\, f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*)\, f_{X_t^*|W_{t-1},V_{t-2}}(x_t^*|w_{t-1}, z). \end{aligned} \quad (47)$$

Define the J-by-J matrices

$$\begin{aligned} L_{V_{t+1},w_t|w_{t-1},V_{t-2}} &= \left[f_{V_{t+1},W_t|W_{t-1},V_{t-2}}(i, w_t|w_{t-1}, j)\right]_{i,j}, \\ L_{V_{t+1}|w_t,X_t^*} &= \left[f_{V_{t+1}|W_t,X_t^*}(i|w_t, j)\right]_{i,j}, \\ L_{X_t^*|w_{t-1},V_{t-2}} &= \left[f_{X_t^*|W_{t-1},V_{t-2}}(i|w_{t-1}, j)\right]_{i,j}, \end{aligned}$$

for $i, j = 1, 2, ..., J$ and a J-by-J diagonal matrix

$$D_{w_t|w_{t-1},X_t^*} = \begin{bmatrix} f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, 1) & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, J) \end{bmatrix}$$

for $w_t \in \mathcal{W}_t$, $w_{t-1} \in \mathcal{W}_{t-1}$. Given these definitions, we can write Eq. (47) in matrix notation as

$$L_{V_{t+1},w_t|w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}}. \quad (48)$$

for any $w_t \in \mathcal{W}_t$, $w_{t-1} \in \mathcal{W}_{t-1}$. Obviously, the unknown matrices on the right hand side are not uniquely determined by the observed matrix on the left hand side without further assumptions. Notice, however, that the matrix $L_{V_{t+1}|w_t,X_t^*}$ does not depend on w_{t-1} and $L_{X_t^*|w_{t-1},V_{t-2}}$ does not depend on w_t. This important fact in Eq. (48) may help the identification of $L_{V_{t+1}|w_t,X_t^*}$.

We assume that for any given $w_t \in \mathcal{W}_t$ there exists $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1})$ with $\overline{w}_t \neq w_t$ and $\overline{w}_{t-1} \neq w_{t-1} \in \mathcal{W}_{t-1}$ such that the matrices $L_{V_{t+1},w_t|w_{t-1},V_{t-2}}$, $L_{V_{t+1},\overline{w}_t|w_{t-1},V_{t-2}}$, $L_{V_{t+1},w_t|\overline{w}_{t-1},V_{t-2}}$, and $L_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},V_{t-2}}$ are all invertible. This assumption is testable from the data. Eq. (48)

then implies

$$\text{for } (w_t, w_{t-1}): \qquad L_{V_{t+1},w_t|w_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}}, \tag{49}$$

$$\text{for } (\overline{w}_t, w_{t-1}): \qquad L_{V_{t+1},\overline{w}_t|w_{t-1},V_{t-2}} = L_{V_{t+1}|\overline{w}_t,X_t^*} D_{\overline{w}_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},V_{t-2}}, \tag{50}$$

$$\text{for } (w_t, \overline{w}_{t-1}): \qquad L_{V_{t+1},w_t|\overline{w}_{t-1},V_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} L_{X_t^*|\overline{w}_{t-1},V_{t-2}}, \tag{51}$$

$$\text{for } (\overline{w}_t, \overline{w}_{t-1}): \qquad L_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},V_{t-2}} = L_{V_{t+1}|\overline{w}_t,X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L_{X_t^*|\overline{w}_{t-1},V_{t-2}}, \tag{52}$$

where all the left hand side matrices are observed. The key identification procedure includes three eliminations. First, eliminating matrix $L_{X_t^*|w_{t-1},V_{t-2}}$ in Eqs. (49) and (50) leads to

$$\begin{aligned} \mathbf{A} &\equiv L_{V_{t+1},w_t|w_{t-1},V_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|w_{t-1},V_{t-2}} \\ &= L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*}. \end{aligned} \tag{53}$$

Second, eliminating $L_{X_t^*|\overline{w}_{t-1},V_{t-2}}$ in Eqs. (51) and (52) results in

$$\begin{aligned} \mathbf{B} &\equiv L_{V_{t+1},w_t|\overline{w}_{t-1},V_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},V_{t-2}} \\ &= L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} D^{-1}_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*}. \end{aligned} \tag{54}$$

Notice that matrices $\mathbf{A}$ and $\mathbf{B}$ are still directly estimable from the data. Third, we eliminate $L_{V_{t+1}|\overline{w}_t,X_t^*}$ in Eqs. (53) and (54) to obtain

$$\mathbf{AB}^{-1} = L_{V_{t+1}|w_t,X_t^*} D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|w_t,X_t^*}, \tag{55}$$

where

$$D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} = D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*}.$$

Since $D_{w_t|w_{t-1},X_t^*}$ is diagonal, the matrix $D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}$ is also diagonal with j-th diagonal entry equal to

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

Therefore, Eq. (55) implies that the observed matrix $\mathbf{AB}^{-1}$ on the left hand side has an eigenvalue-eigenvector decomposition. Each value on the diagonal of $D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*}$ is an eigenvalue and each corresponding column of $L_{V_{t+1}|w_t,X_t^*}$ is a corresponding eigenvector. Each eigenvector is normalized because it must sum to 1.

One ambiguity left is the possibility that the eigenvalues may not be distinctive. Therefore, we need to assume that for any $w_t \in \mathcal{W}_t$, there exists a $(\overline{w}_t, w_{t-1}, \overline{w}_{t-1})$ such that $k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, j) < \infty$ for all $j \in \mathcal{X}_t^*$ and $k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, j_1) \neq k(w_t, \overline{w}_t, w_{t-1}, \overline{w}_{t-1}, j_2)$ for $j_1 \neq j_2$. Then all the unknowns on the right hand side of Eq. (55) are uniquely determined by the decomposition of the observed matrix on the left hand side. This matrix $L_{V_{t+1}|w_t,X_t^*}$ is identified up to the permutation of its columns, which implies the identification of $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ up to the value of x_t^*.

In order to identify how the USV X_t^* changes, we may wish to pin down its value. As shown in Hu (2007), there are various ways to pin down the value of x_t^*. For example, we may normalize the value of x_t^* be the median or another quantile of the distribution $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$. As required in Assumption 4, such a quantile needs to be different for a different value of x_t^*. In summary, the conditional density $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ is identified for any $w_t \in \mathcal{W}_t$.

Step 2: Identification of $\mathbf{f_{W_{t+1}|W_t,X_t^*}}$. We then show that the identification of $f_{V_{t+1}|W_t,X_t^*}$ implies that of $f_{W_{t+1}|W_t,X_t^*}$. Define for any given $w_{t+1} \in \mathcal{W}_{t+1}$, $w_t \in \mathcal{W}_t$, and $w_{t-1} \in \mathcal{W}_{t-1}$,

$$\begin{aligned}
\overrightarrow{f}_{W_{t+1},W_t|W_{t-1},V_{t-2}} &= \left[f_{W_{t+1},W_t|W_{t-1},V_{t-2}}(w_{t+1}, w_t|w_{t-1}, 1), ..., f_{W_{t+1},W_t|W_{t-1},V_{t-2}}(w_{t+1}, w_t|w_{t-1}, J)\right], \\
\overrightarrow{f}_{W_{t+1}|W_t,X_t^*} &= \left[f_{W_{t+1}|W_t,X_t^*}(w_{t+1}|w_t, 1), ..., f_{W_{t+1}|W_t,X_t^*}(w_{t+1}|w_t, J)\right].
\end{aligned}$$

One can show that for any $w_t \in \mathcal{W}_t$

$$\overrightarrow{f}_{W_{t+1}|W_t,X_t^*} = \overrightarrow{f}_{W_{t+1},W_t|W_{t-1},V_{t-2}} \left(L_{V_{t+1},w_t|w_{t-1},V_{t-2}}\right)^{-1} L_{V_{t+1}|w_t,X^*}.$$

Therefore, the density $f_{W_{t+1}|W_t,X_t^*}$ is identified.

Step 3: Identification of $\mathbf{f_{W_t,X_t^*,W_{t-1},V_{t-2}}}$. Moreover, the identification of $f_{V_{t+1}|W_t,X_t^*}$ also implies that of $f_{W_t,X_t^*,W_{t-1},V_{t-2}}$. Eq. (47) also implies

$$f_{V_{t+1},W_t,W_{t-1},V_{t-2}} = \sum_{X_t^* \in \mathcal{X}_t^*} f_{V_{t+1}|W_t,X_t^*} f_{W_t,X_t^*,W_{t-1},V_{t-2}}. \tag{56}$$

Define for any given $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{W}_{t-1}$,

$$L_{w_t,X_t^*,w_{t-1},V_{t-2}} = \left[f_{W_t,X_t^*,W_{t-1},V_{t-2}}(w_t, i|w_{t-1}, j)\right]_{i,j}.$$

Eq. (56) is equivalent to[12]

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

Therefore, the identification of $L_{V_{t+1}|w_t,X^*}$ implies that $L_{w_t,X_t^*,w_{t-1},V_{t-2}}$ is identified as $L^{-1}_{V_{t+1}|w_t,X^*}L_{V_{t+1},w_t|w_{t-1},V_{t-2}}$ for any $w_t \in \mathcal{W}_t$. Consequently, the density $f_{W_t,X_t^*,W_{t-1},V_{t-2}}$ is identified.

Step 4: Identification of $\mathbf{f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}}$. So far, we have only used the four observations $W_{t+1}, W_t, W_{t-1}, W_{t-2}$. In the last step, we use one more period of the data W_{t-3} to identify the desired joint density $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$.

Replacing t by $t-1$ in the previous three steps implies that the additional information from $\{W_t, W_{t-1}, W_{t-2}, W_{t-3}\}$ or the density $f_{W_t,W_{t-1},W_{t-2},W_{t-3}}$ identifies $f_{W_t|W_{t-1},X_{t-1}^*}$ and $f_{W_{t-1},X_{t-1}^*,W_{t-2},V_{t-3}}$.[13] In turn, we can identify the density $f_{X_{t-1}^*|W_{t-1},V_{t-2}}$ given the known mapping from W_{t-2} to V_{t-2}.

We then use the identified densities $f_{W_t,X_t^*,W_{t-1},V_{t-2}}$ and $f_{X_{t-1}^*|W_{t-1},V_{t-2}}$ to identify $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$. The Markov property implies

$$\begin{aligned} & f_{W_t,X_t^*|W_{t-1},V_{t-2}}\left(w_t, x_t^*|w_{t-1}, z\right) \\ = & \sum_{X_{t-1}^* \in \mathcal{X}_{t-1}^*} f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}\left(w_t, x_t^*|w_{t-1}, x_{t-1}^*\right) f_{X_{t-1}^*|W_{t-1},V_{t-2}}\left(x_{t-1}^*|w_{t-1}, z\right). \end{aligned} \tag{57}$$

Define for any $w_t \in \mathcal{W}_t$, and $w_{t-1} \in \mathcal{W}_{t-1}$

$$\begin{aligned} L_{w_t,X_t^*|w_{t-1},V_{t-2}} &= \left[f_{W_t,X_t^*|W_{t-1},V_{t-2}}(w_t, i|w_{t-1}, j)\right]_{i,j}, \\ L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} &= \left[f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}\left(w_t, i|w_{t-1}, j\right)\right]_{i,j}, \\ L_{X_{t-1}^*|w_{t-1},V_{t-2}} &= \left[f_{X_{t-1}^*|W_{t-1},V_{t-2}}\left(i|w_{t-1}, j\right)\right]_{i,j}, \end{aligned}$$

[12] In fact, $L_{w_t,X_t^*,w_{t-1},X_{t-2}} = D_{w_t|w_{t-1},X_t^*}L_{X_t^*|w_{t-1},X_{t-2}}$.

[13] In this discrete case, if our goal is only to identify $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ for period t, then it may be possible to dispense with Assumption 4 together, and pick two arbitrary orderings for X_t^* and X_{t-1}^* in recovering $L_{V_{t+1}|w_t,X_t^*}$ and $L_{V_t|w_{t-1},X_{t-1}^*}$. If we do this, we will not be able to pin down the exact value of X_t^* or X_{t-1}^*, but the recovered density of $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ will still be consistent with the two arbitrary orderings for X_t^* and X_{t-1}^* (in the sense that the implied transition matrix $X_t^*|X_{t-1}^*, w_{t-1}$ for every $w_{t-1} \in \mathcal{W}_{t-1}$ will be consistent with the true, but unknown ordering of X_t^* and X_{t-1}^*). But this will not suffice if our goal is 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 (say) values of X_t^* across these two periods, then we must invoke Assumption 4 to pin down values of X_t^* which are consistent across the two periods. We thank Thierry Magnac for this insight.

for $i, j = 1, 2, ..., J$. Then it is straightforward to show that Eq. (57) implies

$$L_{w_t, X_t^* | w_{t-1}, V_{t-2}} = L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*} L_{X_{t-1}^* | w_{t-1}, V_{t-2}}, \tag{58}$$

where the invertibility of $L_{V_{t+1}, w_t | w_{t-1}, V_{t-2}}$ in Eq. (48) implies that of $L_{w_t, X_t^* | w_{t-1}, V_{t-2}}$. That mean all the matrices in Eq. (58) are invertible. Therefore, $L_{w_t, X_t^* | w_{t-1}, X_{t-1}^*}$ is identified as

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

This results hold for any $w_t \in \mathcal{W}_t$, and $w_{t-1} \in \mathcal{W}_{t-1}$, and therefore, the density $f_{W_t, X_t^* | W_{t-1}, X_{t-1}^*}$ is identified. ■

References

Ackerberg, D. (2003): "Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Examination," *International Economic Review*, 44, 1007–1040.

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

Hu, Y. (2007): "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: a General Solution," *Journal of Econometrics*, forthcoming.

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

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

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

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.

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

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

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

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

Pesendorfer, M., and P. Schmidt-Dengler (2003): "Identification and Estimation of Dynamic Games," NBER Working Paper, #9726.

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.

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