# Nonparametric Identification of Dynamic Models with Unobserved State Variables*

Yingyao Hu
Dept. of Economics, Johns Hopkins University

Matthew Shum
Division of Humanities and Social Sciences, Caltech

First version: April 2008
This version: November 2008
**Incomplete mid-revision draft**

**Abstract**

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

## 1 Introduction

In this paper, we consider the identification of a Markov process $\{W_t, X_t^*\}$ when only $\{W_t\}$, a subset of the variables, is observed. In structural dynamic models, $W_t$ typically consists of the sequences 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@caltech.edu`. We thank Xiaohong Chen, Jeremy Fox, Han Hong, Ariel Pakes, and Susanne Schennach for their suggestions. Seminar participants at Harvard, LSE, NYU, Penn, Penn State, Toulouse, UCL and the Cowles 2008 Summer Conference at Yale provided useful comments.

We demonstrate two main results. First, in the non-stationary case, where the Markov law of motion $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$, can vary across periods $t$, we show that, for any period $t$, $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is identified from the observation of five periods of data $W_{t+1},\ldots,W_{t-3}$ under reasonable assumptions. Second, in the stationary case, where $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ is the same across all $t$, only four observations $W_{t+1},\ldots,W_{t-2}$, for some $t$, are required for identification.

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

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

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

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

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

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

examples given above; also see Rust (1994)), $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ can be factored into

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

The first term denotes the conditional choice probability for the agent's optimal choice in period $t$. The second term is the Markovian law of motion for the state variables $(M_t, X_t^*)$. 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 law of motion for the state variables are recovered, it is straightforward to use them as inputs in a CCP-based approach for estimating dynamic discrete-choice models. This approach was pioneered in Hotz and Miller (1993) and Hotz, Miller, Sanders, and Smith (1994). Subsequent methodological developments in this vein include Aguirregabiria and Mira (2002), (2007), Pesendorfer and Schmidt-Dengler (2007), Bajari, Benkard, and Levin (2007), Pakes, Ostrovsky, and Berry (2007), and Hong and Shum (2007).[2] Alternatively, it is possible to use our identification results for the CCP's and laws of motions for the state variables as a "first-step" in an argument for identification of the per-period utility functions, in the spirit of Magnac and Thesmar (2002) and Bajari, Chernozhukov, Hong, and Nekipelov (2007), who considered the case of dynamic discrete-choice models without serially correlated unobserved state variables.

A general criticism of these CCP-based methods is that they cannot accommodate unobservables which are persistent over time. However, there are some recent papers focusing on the CCP-based estimation of dynamic discrete-choice models, in the presence of a latent state variable $X_t^*$. 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

[2]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.

(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.

While these papers have focused on estimation, our focus is on identification. Kasahara and Shimotsu (2007, hereafter KS) 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, KS prove the nonparametric identification of the Markov kernel $W_{t+1}|W_t, X^*$ in this setting, using six periods of data. Similarly to KS, our identification arguments rely on a matrix diagonalization argument. However, our framework is more general than in KS, in that we allow $X_t^*$ to vary over periods, and also to be drawn from a continuous distribution.

Henry, Kitamura, and Salanie (2008, hereafter HKS) exploit exclusion restrictions to identify Markov regime-switching models with a discrete and latent state variable. As in HKS, our identification result uses the feature of first-order Markovian models that, conditional on $W_{t-1}$, $W_{t-2}$ is an "excluded variable" which affects $W_t$ only via the unobserved state $X_t^*$. However, our identification arguments, which rely on recent econometric results for nonclassical measurement error models, are quite distinct from those in HKS.

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

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

## 2 Overview of assumptions

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

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

Below, we introduce our four assumptions. These assumptions have roles of both (i) ruling out certain types of models; and (ii) establishing identification. The first assumption below restricts attention to certain classes of models, where Assumptions 2-4 establish identification for the restricted class of models.

## 2.1 Model restrictions

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

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

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

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

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

Assumption 1(ii) is a “limited feedback” assumption, because it rules out direct feedback from the last period’s USV, $X_{t-1}^*$, on the current value of the observed component $W_t$. When $W_t = (Y_t, M_t)$, where $Y_t$ denotes the agent’s action in period $t$, and $M_t$ denotes the period-$t$ observed state variable, Assumption 1 implies that:

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

In the bottom line of the above display, the limited feedback assumption eliminates $X_{t-1}^*$ as a conditioning variable in both terms. In Markovian dynamic optimization models, the first term (corresponding to the CCP) can be further simplified to $f_{Y_t|M_t,X_t^*}$, because the Markovian laws of motion for the state variables $(M_t, X_t^*)$ imply that the optimal policy function depends just on the current state variables, but not past values. Hence, Assumption 1 imposes weaker restrictions on the first term than Markovian dynamic optimization models.[3]

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

Although Assumption 1(i) implies a first order Markov process, our identification results can be extended to a general $k$-th order Markov process ($k < \infty$). In that case, the limited feedback assumption 1(ii) may be relaxed to

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

For example, in a 2nd-order Markov model, the analogous limited feedback condition allows $W_t$ to depend on $X_{t-1}^*$, i.e., $f_{W_t|W_{t-1},W_{t-2},X_t^*,X_{t-1}^*}$. The identification of the $k$-th order Markov law of motion $f_{W_t,X_t^*|W_{t-1},\ldots,W_{t-k},X_{t-1}^*,\ldots,X_{t-k}^*}$ from the $3k+2$ observations $W_{t+k},\ldots,W_{t-2k-1}$ can be shown under straightforward extensions of the assumptions presented in this section, and we do not explore this extension here.

[3] Moreover, if we move outside the class of these models, the above display also shows that Assumption 1 does not rule out the dependence of $Y_t$ on $Y_{t-1}$ or $M_{t-1}$, which corresponds to some models of state dependence. These may include linear or nonlinear panel data models with lagged dependent variables, and serially correlated errors, cf. Arellano and Honore (2000). Arellano (2003, chs. 7–8) considers linear panel models with lagged dependent variables and persistent unobservables, which is also related to our framework. Indeed, both the Markov and limited feedback assumptions are examples of "dynamic completeness" assumptions made in dynamic panel data models, which limit the dependence of the current dependent variables on lagged dependent and explanatory variables.

### 2.2 Notation

We assume that the unobserved state variable $X_t^*$ is scalar-valued, and is drawn from a continuous distribution.[4] Hence, we will employ linear operator-theoretic arguments for identification. For what follows, we introduce notation for linear integral operators. Let $R_1$, $R_2$ denote two random variables, with support $\mathcal{R}_1$ and $\mathcal{R}_2$, distributed according to the joint density $f_{R_1,R_2}(r_1, r_2)$.[5] Let $\mathcal{L}^p(\mathcal{X})$, $1 \leq p < \infty$ denote the space of function $h(\cdot)$ with $\int_{\mathcal{X}} |h(x)|^p dx < \infty$. For any $1 \leq p \leq \infty$, we define an integral operator $L_{R_1,R_2} : \mathcal{L}^p(\mathcal{R}_2) \to \mathcal{L}^p(\mathcal{R}_1)$ for any $h \in \mathcal{L}^p(\mathcal{R}_2)$,

$$\left(L_{R_1,R_2} h\right)(v) = \int f_{R_1,R_2}(v, z)\, h(z) dz.$$

Similarly, for a random variable $R_3$ with support $\mathcal{R}_3$ and density $f_{R_3}(r_3)$, we define the diagonal operator $D_{R_3} : \mathcal{L}^p(\mathcal{R}_3) \to \mathcal{L}^p(\mathcal{R}_3)$ for any $h \in \mathcal{L}^p(\mathcal{R}_3)$,

$$\left(D_{R_3} h\right)(v) = f_{R_3}(v)\, h(v).$$

**Scalarization of $W_t$** 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

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

where the function $g_{t+1} : \mathbb{R}^d \to \mathbb{R}$ is known and $d$ is the dimension of $W_t$. (When $W_{t+1}$ is a scalar, we define $g_{t+1}(w) = w$.) The restrictions imposed later on the function $g_{t+1}$ 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

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

with a known function $q_{t-2} : \mathbb{R}^d \to \mathbb{R}$. As we discuss later, we introduce the function $q_{t-2}$ only in order to avoid introducing additional notation involving the generalized inverse of an operator. In addition, the functions $g_{t+1}$ and $q_{t-2}$ do not need to depend on $t$.

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

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

### 2.3 Identification via spectral decomposition of observed densities

As will be clear from the next section, the crucial step in our identification argument is a spectral decomposition of a linear operator generated from $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$, which corresponds to the observed density of $V_{t+1}, W_t|W_{t-1}, Z_{t-2}$. The next two assumptions guarantee the validity and uniqueness of this spectral decomposition.

**Assumption 2** *Invertibility: for any $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{W}_{t-1}$,*
*(i) $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$ is one-to-one.*
*(ii) $L_{V_{t+1}|w_t,X_t^*}$ is one-to-one.*
*(iii) $L_{V_t|w_{t-1},Z_{t-2}}$ is one-to-one.*

**Assumption 3** *Uniqueness of spectral decomposition:*
*(i) Bounded eigenvalues: For any $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{W}_{t-1}$, there exist functions $C_1(w_t, w_{t-1})$ and $C_2(w_t, w_{t-1})$ such that the density $f_{W_t|W_{t-1},X_t^*}$ satisfies*

$$0 < C_1(w_t, w_{t-1}) \leq f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*) \leq C_2(w_t, w_{t-1}) < \infty, \quad \forall x_t^* \in \mathcal{X}_t^*. \tag{2}$$

*(ii) Distinct eigenvalues: For any $w_t \in \mathcal{W}_t$, there exists $w_{t-1} \in \mathcal{W}_{t-1}$ such that the density $f_{W_t|W_{t-1},X_t^*}$ satisfies for any $x_t^* \in \mathcal{X}_t^*$*

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

*where $m_t$ (resp. $m_{t-1}$) is a continuous component of $w_t$ (resp. $w_{t-1}$).*

Assumption 2 restricts three operators to be one-to-one, which implies that they are invertible.[6] The operators in (i) and (iii) correspond to densities which are observed in the data, so that the one-to-one assumptions on those operators could be tested. On the other hand, the density $f_{V_{t+1}|w_t,X_t^*}$, corresponding to the operator in (ii), is not directly observed, and so should be checked on a model-by-model basis.

Broadly speaking, Assumption 2(ii) rules 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

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

the one-to-one assumption. In Section 4, we consider this assumption in several example models.

Assumption 3 contains technical conditions which ensure the existence and uniqueness of this decomposition. We will elaborate upon the specific role that this assumption plays in the next section. Since the density $f_{W_t|W_{t-1},X_t^*}$ is not observed in the data, this assumption should be verified on a model-by-model basis. Generally, Assumption 3(ii) is 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.[7]

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

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

The functional $G$, which may depend on the value of $w_t$, can include the mean, mode, median, or another quantile of $f_{V_{t+1}|W_t,X_t^*}$. Having introduced our four main identification assumptions, we proceed in the next section to present our main identification result.

# 3 Main nonparametric identification results

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

**Lemma 1 (Representation of the observed density $f_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$):**

[7]It turns out that Assumptions 2 and 3, as stated here, are stronger than necessary. An earlier version of the paper (Hu and Shum (2008)) contained less restrictive versions of these assumptions. However, they are not as intuitive as the stronger versions presented here.

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

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

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

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

**Proofs:** in Appendix.

Since $L_{V_{t+1},w_t|w_{t-1},Z_{t-2}}$ and $L_{V_t|w_{t-1},Z_{t-2}}$ are observed, Lemma 2 implies that identification of $L_{V_{t+1}|w_t,X_t^*}$ and $L_{V_t|w_{t-1},X_{t-1}^*}$ are sufficient for that of $L_{w_t,X_t^*|w_{t-1},X_{t-1}^*}$. We show that identification in the following lemma.

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

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

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

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

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

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

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

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

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

eliminating $L_{X_t^*|w_{t-1},Z_{t-2}}$.[8] Similarly, postmultiplying Eq. (8) by the inverse of Eq. (9):

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

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

$$
\begin{aligned}
\mathbf{AB}^{-1} &\equiv L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}} \left( L_{V_{t+1},w_t|\overline{w}_{t-1},Z_{t-2}} L^{-1}_{V_{t+1},\overline{w}_t|\overline{w}_{t-1},Z_{t-2}} \right)^{-1} \\
&= L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*} \times \\
&\quad \times \left( L_{V_{t+1}|w_t,X_t^*} D_{w_t|\overline{w}_{t-1},X_t^*} D^{-1}_{\overline{w}_t|\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|\overline{w}_t,X_t^*} \right)^{-1} \\
&= L_{V_{t+1}|w_t,X_t^*} \left( D_{w_t|w_{t-1},X_t^*} D^{-1}_{\overline{w}_t|w_{t-1},X_t^*} D_{\overline{w}_t|\overline{w}_{t-1},X_t^*} D^{-1}_{w_t|\overline{w}_{t-1},X_t^*} \right) L^{-1}_{V_{t+1}|w_t,X_t^*} \\
&\equiv L_{V_{t+1}|w_t,X_t^*} D_{w_t,\overline{w}_t,w_{t-1},\overline{w}_{t-1},X_t^*} L^{-1}_{V_{t+1}|w_t,X_t^*}, \qquad \text{where} \qquad (12)
\end{aligned}
$$

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

---

[8] If we do not scalarize $W_{t-2}$ using the function $q_{t-2}$, the operator $L_{V_{t+1},w_t|w_{t-1},W_{t-2}}$ may be surjective, in which case Assumption 2(i) should be replaced by the condition that $L_{V_{t+1},w_t|w_{t-1},W_{t-2}} L^*_{V_{t+1},w_t|w_{t-1},W_{t-2}}$ is one-to-one, where $L^*$ denotes an adjoint operator. In that case, we consider

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

The last expression is the same as using $L^{-1}_{V_{t+1},\overline{w}_t|w_{t-1},Z_{t-2}}$.

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

Assumption 3 ensures that this decomposition is unique, by bounding the eigenvalues (part (i)) and ensuring their distinctiveness (part (ii)). Even given the unique decomposition, the eigenfunctions are only identified up to a normalization, and an arbitrary ordering. Since each eigenfunction $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)$ is a density function, it is appropriate to normalize each function so that it integrates to one: $\int f_{V_{t+1}|W_t,X_t^*}(x|w_t, x_t^*)dx = 1$. Moreover, Assumption 4 requires a functional $G$ to exist such that $G$ applied to the family of densities $f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, \cdot)$ to be monotonic in $X_t^*$, given $w_t$. Given this monotonicity, we can pin down the scale of $X_t^*$ by setting, $x_t^* = G\left[f_{V_{t+1}|W_t,X_t^*}(\cdot|w_t, x_t^*)\right]$ without loss of generality. Therefore, altogether the density $f_{V_{t+1}|W_t,X_t^*}$ or $L_{V_{t+1}|w_t,X_t^*}$ is nonparametrically identified for any given $w_t \in \mathcal{W}_t$ via the spectral decomposition in Eq. (12). ■

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

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

Next we present some important corollaries of our main identification result.

## 3.1 Initial conditions

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

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

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

**Proof:** in Appendix.

### 3.2 Stationarity

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

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

$$f_{W_t,X^*_t|W_{t-1},X^*_{t-1}} = f_{W_2,X^*_2|W_1,X^*_1}, \ \forall\ 2 \leq t \leq T.$$

In most dynamic optimization settings, this assumption is maintained in infinite-horizon models. Below, we show how our identification arguments can be easily extended to the stationary case. Given the foregoing discussion, we present the next corollary without proof.

**Corollary 2 (Identification of Markov law of motion, stationary case):** *Under assumptions 1, 2, 3, 4, and 5, the observed density* $f_{W_{t+1},W_t,W_{t-1},W_{t-2}}$ *for any* $t \in \{3, \ldots, T-1\}$ *uniquely determines the density* $f_{W_2,X^*_2|W_1,X^*_1}$.

In the stationary case, initial conditions are still a concern. The following corollary, analogous to Corollary 1 for the non-stationary case, postulates the identification of $f_{W_{t-2},X^*_{t-2}}$, the marginal distribution of the observables $W$ and unobservable $X^*$.

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

**Proof:** in Appendix.

# 4 Comments on Assumptions in Specific Examples

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

To this end, we present in this section two examples of dynamic models, both of which satisfy the first-order Markov and limited feedback assumptions in Assumption 1. 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 these two example models.

## 4.1 Example 1: A discrete model

As a first example, let $(W_t, X_t^*)$ denote a bivariate discrete first-order Markov process where $W_t$ and $X_t^*$ are both scalars, and binary:

$$\forall t,\ \mathrm{supp}X_t^* = \mathrm{supp}W_t \equiv \{0,1\}\,.$$

This is the simplest example of the models considered in our framework. In this example, no scalarization of $W_t$ is required. We assume that the laws of motion for both $W_t$ and $X_t^*$ exhibit state dependence:

$$\begin{aligned} Pr(W_t = 1|w_{t-1}, x_t^*) = p(w_{t-1}, x_t^*); \quad & Pr(W_t = 0|w_{t-1}, x_t^*) = 1 - p(W_{t-1}, X_t^*) \\ Pr(X_t^* = 1|x_{t-1}^*, w_{t-1}) = q(x_{t-1}^*, w_{t-1}); \quad & Pr(X_t^* = 0|x_{t-1}^*, w_{t-1}) = 1 - q(x_{t-1}^*, w_{t-1}). \end{aligned} \tag{13}$$

These laws of motion satisfy Assumption 1.

This model is a binary version of Abbring, Chiappori, and Zavadil's (2008) "dynamic moral hazard" model of auto insurance. In that model, $W_t$ is a binary indicator of claims occurrence, and $X_t^*$ is a binary effort indicator, with $X_t^* = 1$ denoting higher effort. The mutual dependence of effort and claims in the laws of motion (13) arise from moral hazard, and experience rating in insurance pricing.

The main difference between this discrete case and the previous continuous case is that the linear integral operators are replaced by matrices. The $L$ operators in the main proof correspond to 2×2 square matrices, and the $D$ operators are 2×2 diagonal matrices. Assumptions 2 and 3 are quite transparent to interpret in the matrix setting. Assumption 2 implies the invertibility of certain matrices. From Lemma 1, the following matrix equality holds, for all values of $(w_t, w_{t-1})$:

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

Given this equation, the invertibility of $L_{W_{t+1},w_t|w_{t-1},W_{t-2}}$ implies that $L_{W_{t+1}|w_t,X_t^*}$ and $L_{X_t^*|w_{t-1},W_{t-2}}$ are both invertible, and that all the elements in the diagonal matrix $D_{w_t|w_{t-1},X_t^*}$ are nonzero. Furthermore, by the matrix equality in Eq. (30) of the appendix, the invertibility of $L_{X_t^*|w_{t-1},W_{t-2}}$ also implies that of $L_{V_t|w_{t-1},Z_{t-2}}$. Hence, in this discrete model, Assumptions 2(ii) and 2(iii) are redundant, because they are implied by 2(i), which is testable from the observed data.

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

In this discrete case, Assumption 4 can be interpreted as an "ordering" assumption, which imposes an ordering on the columns of the $L_{W_{t+1}|w_t,X_t^*}$ matrix, corresponding to the eigenvectors of $\mathbf{AB^{-1}}$. If the goal is only to identify $f_{W_t,X_t^*|W_{t-1},X_{t-1}^*}$ for a single period $t$, then we could dispense with Assumption 4 altogether, and pick two arbitrary in recovering $L_{W_{t+1}|w_t,X_t^*}$ and $L_{W_t|w_{t-1},X_{t-1}^*}$. If we do this, we will not be able to pin down the exact value of $X_t^*$ or $X_{t-1}^*$, but the recovered density of $W_t, X_t^*|W_{t-1}, X_{t-1}^*$ will still be consistent with the two arbitrary orderings for $X_t^*$ and $X_{t-1}^*$ (in the sense that the implied transition matrix $X_t^*|X_{t-1}^*, w_{t-1}$ for every $w_{t-1} \in \mathcal{W}_{t-1}$ will be consistent with the true, but unknown ordering of $X_t^*$ and $X_{t-1}^*$).[9]

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

[9]We thank Thierry Magnac for this insight.

across the two periods. Hu (2008) suggests a number of ways to satisfy Assumption 4 in the discrete case. For example, a reasonable restriction, which satisfies Assumption 4, is that

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

which implies that claims occur less frequently with higher effort. This restriction can be ensured by appropriate restrictions on the values of the $p(\cdots)$ and $q(\cdots)$ functions in (13).

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

The second example model is 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.

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 $-\infty < L < U < +\infty$,

$$X_t^* = \begin{cases} 0.5X_{t-1}^* + 0.3\psi\left(M_{t-1}\right) + 0.2\nu_t & \text{if } Y_{t-1} = 0 \\ 0.8X_{t-1}^* + 0.2\nu_t & \text{if } Y_{t-1} = 1 \end{cases} \qquad \text{where} \tag{15}$$

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

$\nu_t$ 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 $f_{\nu_t}(\cdot)$.

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

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

$$u_0(S_t) = -c(M_t) + X_t^*; \quad u_1(S_t) = -RC. \tag{16}$$

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:

$$M_{t+1} = \begin{cases} M_t + \eta_{t+1} & \text{if } Y_t = 0 \\ \eta_{t+1} & \text{if } Y_t = 1 \end{cases} \tag{17}$$

where the incremental mileage $\eta_{t+1} > 0$ is a standard normal random variable, truncated to $[0, 1]$, with density $\tilde{\phi}(\eta) \equiv \frac{\phi(\eta)}{\Phi(1)-\Phi(0)}$, where $\phi$ and $\Phi$ denote, respectively, the standard normal density and CDF.[10] ■

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

$$u_0(S_t) = -c(M_t); \quad u_1(S_t) = -RC. \tag{18}$$

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

$$M_{t+1} = \begin{cases} M_t + \eta_{t+1} \cdot \exp(X_{t+1}^*) & \text{if } Y_t = 0 \\ \eta_{t+1} \cdot \exp(X_{t+1}^*) & \text{if } Y_t = 1. \end{cases} \tag{19}$$

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 $g_{t+1}(\cdot)$ and $q_{t-2}(\cdot)$ introduced at the beginning of Section 2, we use: $V_{t+1} = M_{t+1}$, $Z_{t-2} = M_{t-2}$. That is, we scalarize $W_t$ by just using the continuous component, which is 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, 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.

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

Given the assumptions so far, the conditional choice probabilities take the multinomial logit form (for $Y_t = 0, 1$): $P(Y_t|S_t) = \exp(V_{Y_t}(S_t)) / \left[\sum_{y=0}^{1} \exp(V_y(S_t))\right]$ where $V_y(S_t)$ is the choice-specific value function in period $t$, defined recursively by

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

**Assumption 1** The first-order Markov and limited feedback assumptions are satisfied in both specifications. Implicitly, the limited feedback assumption 1(ii) imposes a timing restriction, that $X^*_{t+1}$ is realized before $M_{t+1}$, so that $M_{t+1}$ depends on $X^*_{t+1}$. On the one hand, this is less restrictive than the assumption that $M_{t+1}$ evolves independently of both $X^*_t$ and $X^*_{t+1}$, which has been made in many applied settings, to enable the estimation of the $M_{t+1}$ law of motion directly from the data. On the other hand, the limited feedback assumption does rule out models such as $M_{t+1} = g(M_t, X^*_t) + \eta_{t+1}$, which is consistent with the alternative timing assumption that $X^*_{t+1}$ is realized after $M_{t+1}$.[11]

**Assumption 2** contains three invertibility assumptions. Assumption 2(i) requires 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 $g_{t+1}(W_{t+1})$, and $M_{t-2}$ for $q_{t-2}(W_{t-2})$.)

Consider Specification A, and consider $w_t$ such that $Y_t = 1$ (so that the engine is replaced in period $t$). In this case, $M_{t+1}|Y_t = 1$ follows a normal distribution truncated to $[0, 1]$, 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 is satisfied. The discussion of Assumption 2(iii) is very similar to that of 2(ii), and we omit it for convenience here.

Assumption 2(ii) requires that, for all $w_t$, the mapping $L_{M_{t+1}|w_t,X^*_t}$ is one-to-one. As before, consider a value $w_t$ such that $Y_t = 1$. In Specification A, $M_{t+1}|w_t, X^*_t$ is distributed according to a standard normal distribution truncated to $[0, 1]$, regardless of the value of $X^*_t$. Hence, the one-to-one requirement fails. For Specification B, however, $M_{t+1}$ is distributed

[11]However, as we noted in the Section 2, our results could be suitably extended to a second-order Markov models, which would allow for $M_{t+1} = g(M_t, M_{t-1}, X^*_{t+1}, X^*_t) + \eta_{t+1}$, under stronger identification assumptions.

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** guarantees the finiteness and distinctiveness of the eigenvalues in the decomposition of Eq. (12). Assumption 3(i), which ensures the finiteness of the eigenvalues, requires 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 $+\infty$. This density can be factored as $f_{W_t|W_{t-1},X_t^*} = f_{Y_t|M_t,X_t^*} \cdot f_{M_t|Y_{t-1},M_{t-1},X_t^*}$. The mileage law of motion $f_{M_t|Y_{t-1},M_{t-1},X_t^*}$ is a truncated normal 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 $\epsilon$'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.

Assumption 3(ii) ensures that the eigenvalues in the decomposition (12) are distinctive. Because of the factorization above, and the fact that the choice probabilities are bounded away from zero, a sufficient condition for Eq. (3) 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 $m_t \in [m_{t-1}, m_{t-1} + \exp(x_t^*)]$. 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 $x_t^*$. Consider the median, i.e., $\text{med}\,[f] = \inf\left\{\widetilde{x}^* : \int_{-\infty}^{\widetilde{x}^*} f(x)dx \geq 0.5\right\}$. Eqs. (15) and (19) 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 the 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 $\eta_{t+1}$ and $\nu_{t+1}$ and the value of $(y_t, m_t)$, we have

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

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

### 4.3 Example 3: generalized investment model

For the third example, we consider a dynamic model of firm R&D and product quality in the "generalized dynamic investment" framework described in Doraszelski and Pakes (2007). In this model, $Y_t$ measures a firm's R&D, and $X_t^*$ measures the firm's product quality, which evolves according to

$$X_{t+1}^* = X_t^* + h(Y_t) + \nu_{t+1}.$$

The observed state variables $M_t$ is installed base, which evolves as

$$M_{t+1} = (1-\delta) * M_t + k(X_{t+1}^*) + \exp(X_{t+1}^*)\xi_{t+1}.$$

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

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

In the above, the errors $(\xi_t, \nu_t, \gamma_t)$ are assumed to be mutually independent, each distributed $N(0,1)$ i.i.d. across periods. These errors are introduced just to induce randomness in

$(Y_t, M_t, X_t^*)$ conditional on $(Y_{t-1}, M_{t-1}, X_{t-1}^*)$.

As in the Rust example above, we scalarize $V_t = M_t$ for this model because, as noted in Levinsohn and Petrin (2000) and Ackerberg, Benkard, Berry, and Pakes (2007), $Y_t$ may be equal to zero for many values of $(M_t, X_t^*)$, and hence may not provide enough information on $X_t^*$.

Obviously, Assumption 1 is satisfied with the above assumptions. To verify that $L_{M_{t+1},w_t|w_{t-1},M_{t-2}}$ is one-to-one, for assumption 2(ii), note that $M_{t+1}$ depends on $X_{t+1}^*$, which is correlated with $M_{t-2}$. Similarly, $M_t$ depends on $X_t^*$, which is correlated with $M_{t-2}$, so that $L_{M_t|w_{t-1},M_{t-2}}$ is one-to-one, and satisfies Assumption 2(iii). For Assumption 2(ii), that $L_{M_{t+1}|w_t,X_t^*}$ is one-to-one, note that the conditional distribution of $M_{t+1}|w_t, X_t^*$ depends on $X_t^*$.

Next we consider Assumption 3. The bounded eigenvalues restriction in Assumption 3(i) can be ensured by truncating supports of $(\gamma_t, \xi_t, \nu_t)$, and the range of the $k(\cdot)$, $h(\cdot)$, and $c(\cdot)$ functions, similarly to what was done for the Rust example above. For part (ii) of this assumption, which guarantees the distinct eigenvalues, we derive that for any $(x_t^*, w_t)$, $\exists\ w_{t-1}$ such that

$$\begin{aligned}
&\frac{\partial^3}{\partial m_t \partial m_{t-1} \partial x_t^*} \ln f_{W_t|W_{t-1},X_t^*}(w_t|w_{t-1}, x_t^*) \\
=&\frac{\partial^3}{\partial m_t \partial m_{t-1} \partial x_t^*} \left[\ln f(y_t|m_t, x_t^*) + \ln f(m_t|w_{t-1}, x_t^*)\right] \neq 0.
\end{aligned}$$

For this model, this is satisfied because $m_t|m_{t-1}, x_t^* \sim \frac{1}{\exp(x_t^*)} \cdot \tilde{\phi}\left(\frac{m_t-(1-\delta)m_{t-1}-k(x_t^*)}{\exp(x_t^*)}\right)$, where $\tilde{\phi}(\cdot)$ denotes a truncated standard normal density function.

Finally, for the monotonicity assumption 4, we note that

$$E[M_{t+1}|m_t, y_t, x_t^*] = (1-\delta)m_t + E[k(X_{t+1}^*)|x_t^*, y_t]$$

so if $E[k(X_{t+1}^*)|x_t^*, y_t]$ is monotonic in $x_t^*$, we can use the mean as the $G$ functional, and pin down $x_t^* = E\left[f_{M_{t+1}|M_t,Y_t,X_t^*}(\cdot|m_t, y_t, x_t^*)\right]$.

# 5 Concluding remarks

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

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

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

# A Proofs

**Proof.** (Lemma 1) Assumption 1(i) implies that the observed density $f_{W_{t+1}, W_t, W_{t-1}, W_{t-2}}$ is equal to

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

(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

$$f_{W_{t+1},W_t|W_{t-1},W_{t-2}} = \int f_{W_{t+1}|W_t,X_t^*} f_{W_t|W_{t-1},X_t^*} f_{X_t^*|W_{t-1},W_{t-2}} dx_t^*. \tag{23}$$

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

$$L_{W_{t+1},w_t|w_{t-1},W_{t-2}} = L_{W_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},W_{t-2}}, \tag{24}$$

given values of $(w_t, w_{t-1}) \in \mathcal{W}_t \times \mathcal{W}_{t-1}$. To see this, note that, 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. (23) is equivalent to eq. (24).

Recall the scalarizing functions $g_{t+1}$ ,$q_{t-2} : \mathbb{R}^d \to \mathbb{R}$, and

$$V_{t+1} = g_{t+1}(W_{t+1}), \quad Z_{t-2} = q_{t-2}(W_{t-2}).$$

Hence, using Eq. (24), the joint density of $\{V_{t+1}, W_t, W_{t-1}, Z_{t-2}\}$ can be expressed in operator notation, for any $(x, w_t, w_{t-1}, z) \in g_{t+1}(\mathcal{W}_{t+1}) \times \mathcal{W}_t \times \mathcal{W}_{t-1} \times q_{t-2}(\mathcal{W}_{t-2})$, as

$$L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} = L_{V_{t+1}|w_t,X_t^*} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}}, \tag{25}$$

as postulated by the lemma. ■

**Proof.** (Lemma 2) Assumption 1 implies

$$\begin{aligned} f_{W_t,X_t^*|W_{t-1},Z_{t-2}} &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*,Z_{t-2}} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\ &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^*. \end{aligned} \tag{26}$$

To proceed, we derive an operator equality corresponding to Eq. (26). For the left-hand side of this equation, we have

$$\begin{aligned} f_{W_t,X_t^*|W_{t-1},Z_{t-2}} &= \int f_{W_t,X_t^*,X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\ &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*,Z_{t-2}} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\ &= \int f_{W_t,X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\ &= \int f_{W_t|W_{t-1},X_t^*,X_{t-1}^*} f_{X_t^*|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\ &= \int f_{W_t|W_{t-1},X_t^*} f_{X_t^*,X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^* \\ &= f_{W_t|W_{t-1},X_t^*} f_{X_t^*|W_{t-1},Z_{t-2}}. \end{aligned} \tag{27}$$

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 $w_t, w_{t-1}$. Hence, combining eq. (26) and eq. (27) leads to

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

which, in operator notation, is equivalent to

$$\begin{aligned} D_{w_t|w_{t-1},X_t^*} L_{X_t^*|w_{t-1},Z_{t-2}} &= L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*|w_{t-1},Z_{t-2}}. \\ \Leftrightarrow L_{w_t,X_t^*|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*|w_{t-1},Z_{t-2}} &= L_{V_{t+1}|w_t,X_t^*}^{-1} L_{V_{t+1},w_t|w_{t-1},Z_{t-2}} \end{aligned} \tag{28}$$

where the second line follows from Eq. (25).

By integrating out $X_t^*$ in Eq. (26) and then scalarizing $W_t$ to $V_t \equiv g_t(W_t)$, we obtain

$$f_{V_t|W_{t-1},Z_{t-2}} = \int f_{V_t|W_{t-1},X_{t-1}^*} f_{X_{t-1}^*|W_{t-1},Z_{t-2}} dx_{t-1}^*. \tag{29}$$

In operator notation, this is

$$\begin{aligned}
&L_{V_t|w_{t-1},Z_{t-2}} = L_{V_t|w_{t-1},X_{t-1}^*} L_{X_{t-1}^*|w_{t-1},Z_{t-2}} \\
\Leftrightarrow &L_{X_{t-1}^*|w_{t-1},Z_{t-2}} = L^{-1}_{V_t|w_{t-1},X_{t-1}^*} L_{V_t|w_{t-1},Z_{t-2}} \\
\Rightarrow &L^{-1}_{X_{t-1}^*|w_{t-1},Z_{t-2}} = L^{-1}_{V_t|w_{t-1},Z_{t-2}} L_{V_t|w_{t-1},X_{t-1}^*}
\end{aligned} \tag{30}$$

where the second line applies Assumption 2(ii), and the third line applies 2(ii) and 2(iii).

Hence, for all $w_t \in \mathcal{W}_t$ and $w_{t-1} \in \mathcal{W}_{t-1}$, the desired Markov law of motion operator $L_{w_t,X_t^*|w_{t-1},X_{t-1}^*}$ in eq. (28) can be written as

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

■

**Proof.** (Corollary 1)

From Lemma 3, we know that $f_{V_t|W_{t-1},X_{t-1}^*}$ is identified from the observed density $f_{V_t W_{t-1}|W_{t-2},Z_{t-3}}$, where $V_t$ is the "scalarized" version of $W_t$. The following equation

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

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

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

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

■

**Proof.** (Corollary 3)

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

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

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

# References


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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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