

# Moment Restrictions for Nonlinear Panel Data Models with Feedback

Stéphane Bonhomme, Kevin Dano, and Bryan S. Graham\*

July 15, 2025

## Abstract

Many panel data methods, while allowing for general dependence between covariates and time-invariant agent-specific heterogeneity, place strong *a priori* restrictions on *feedback*: how past outcomes, covariates, and heterogeneity map into future covariate levels. Ruling out feedback entirely, as often occurs in practice, is unattractive in many dynamic economic settings. We provide a general characterization of all *feedback and heterogeneity robust* (FHR) moment conditions for nonlinear panel data models and present constructive methods to derive feasible moment-based estimators for specific models. We also use our moment characterization to compute semiparametric efficiency bounds, allowing for a quantification of the information loss associated with accommodating feedback, as well as providing insight into how to construct estimators with good efficiency properties in practice. Our results apply both to the finite dimensional parameter indexing the parametric part of the model as well as to estimands that involve averages over the distribution of unobserved heterogeneity. We illustrate our methods by providing a complete characterization of all FHR moment functions in the multi-spell mixed proportional hazards model. We compute efficient moment functions for both model parameters and average effects in this setting.

**JEL Codes:** C23, C33

**Keywords:** Sequential Exogeneity, Feedback, Panel Data, Duration Models, Incidental Parameters, Semiparametric Efficiency Bounds.

---

\*Bonhomme: Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, e-mail: [sbonhomme@uchicago.edu](mailto:sbonhomme@uchicago.edu), web: <https://sites.google.com/site/stephanebonhommeresearch/>.

Dano: Department of Economics, Princeton University, Washington Rd, Princeton, NJ 08544, e-mail: [kdano@princeton.edu](mailto:kdano@princeton.edu), web: <https://kevindano.github.io>.

Graham: Department of Economics, University of California - Berkeley, 530 Evans Hall #3380, Berkeley, CA 94720-3880 and National Bureau of Economic Research, e-mail: [bgraham@econ.berkeley.edu](mailto:bgraham@econ.berkeley.edu), web: <http://bryangraham.github.io/econometrics/>.

We thank Manuel Arellano, Jinyong Hahn, Bo Honoré, Whitney Newey, Aureo de Paula, Martin Weidner, Tiemen Woutersen, and seminar participants at various places for comments. All the usual disclaimers apply. Portions of the research reported were undertaken while Bonhomme and Graham were visiting CEMFI in the Fall of 2022 with support from the Spanish State Research Agency under the María de Maeztu Unit of Excellence Programme (Project No: CEX2020-001104-M).

An econometrician randomly samples units from a population of interest. For each sampled unit  $i = 1, \dots, N$ , let  $Y_{it}$  denote a period  $t = 1, \dots, T$  outcome,  $X_{it}$  a corresponding vector of covariates, and  $A_i$ , a latent variable representing unmeasured unit-specific attributes. Importantly,  $A_i$  is constant over time and may freely covary with the regressors,  $X_{i1}, \dots, X_{iT}$ . An initial condition,  $Y_{i0}$ , is also observed. Panel data of this type feature prominently in empirical research in economics and other fields. That panel data offers the possibility to “control for” the correlated heterogeneity,  $A_i$ , is a key attraction.

While “fixed effects” panel data methods place no restrictions (beyond mild regularity conditions) on the joint distribution of the initial condition  $Y_{i0}$  and latent heterogeneity  $A_i$ , they generally *do* place strong restrictions on how the outcomes  $Y_{i1}, \dots, Y_{iT}$  and regressors  $X_{i1}, \dots, X_{iT}$  relate to each other. These restrictions involve more than substantive modeling assumptions; they also constrain what we will call the *feedback process*, whereby past outcomes and covariates  $Y_{it-1}, Y_{it-2}, \dots, Y_{i0}, X_{it-1}, X_{it-2}, \dots, X_{i1}$ , as well as heterogeneity  $A_i$ , influence current covariates  $X_{it}$ .

Feedback arises naturally in many dynamic economic problems. For example, a firm’s optimal investment rule typically varies with its current capital stock (and hence past investment decisions) as well as past productivity shocks (and hence its output history), see, e.g., Olley and Pakes (1996) and Blundell and Bond (2000). A doctor may adjust a patient’s treatment protocol in a way which depends on her perceptions of their health response to past treatments (e.g., Robins, 1986). A worker’s decision to participate in job training may, as is typically the focus in evaluation studies, influence their future labor market outcomes, but participation may also depend on their past labor market experiences (e.g., Ashenfelter, 1978; Ashenfelter and Card, 1985).

One approach to handling feedback, indeed the leading one in empirical work, rules it out *a priori*. This corresponds to maintaining the *strict exogeneity* assumption formulated by Chamberlain (1982a,b). Strict exogeneity assumptions underpin, albeit generally implicitly, many panel data based approaches to program evaluation (see Ghanem et al., 2022 on difference-in-differences methods). The overwhelming majority of nonlinear panel data estimators also require strict exogeneity (see Arellano and Honoré, 2001; Arellano and Bonhomme, 2011). Strict exogeneity, while a convenient assumption for estimation, is restrictive in many economic applications. Ironically, although Chamberlain (1982b, 1984) emphasized the testable implications of strict exogeneity, today the assumption is so common as to often go unmentioned in applied work.

A different approach, pioneered by Robins (1986), assumes that the feedback process is *homogeneous*. By homogeneous we mean that the mapping from past outcomes,  $Y_{it-1}, Y_{it-2}, \dots, Y_{i0}$  and covariates  $X_{it-1}, X_{it-2}, \dots, X_{i1}$  to the current covariate,  $X_{it}$ , *does not*

vary with  $A_i$ : it is identical across agents. This is a powerful simplification, leading to feasible nonparametric and semiparametric estimators (e.g., Robins, 2000). However, the restriction to homogeneous feedback, like strict exogeneity, is a strong assumption. It rules out, for example, a firm's investment rule varying with its persistent productivity level. Robin's (1986) setup is often plausible in environments where the researcher controls  $X_{it}$ , such as in a dynamic experiment. Strict exogeneity and homogeneous feedback are non-nested assumptions; but both restrictions correspond to a subset of the data generating processes covered by our results.

In this paper we study nonlinear panel data models with *unrestricted* heterogeneous feedback and correlated heterogeneity. Almost 25 years ago, surveying the then extant work on nonlinear panel data analysis, Arellano and Honoré (2001, p. 3265) observed:

The main limitation of much of the literature on nonlinear panel data methods is that it is assumed that the explanatory variables are strictly exogenous in the sense that some assumptions will be made on the errors conditional on all (including future) values of the explanatory variables.

Arellano and Honoré's (2001) observation remains largely true today. Chamberlain (2022), in a paper first circulated in the early 1990s, studied a class of panel data models with multiplicative heterogeneity defined by sequential moment restrictions. Certain panel data count models are covered by his results (see Chamberlain, 1992; Wooldridge, 1997; Blundell et al., 2002; Windmeijer, 2008). Feedback in dynamic *linear* panel data models with sequential moment restrictions is also well understood (e.g., Arellano and Bond, 1991; Arellano and Bover, 1995; Chamberlain, 1992; Hahn, 1997; Ai and Chen, 2012). However, outside the setting studied by Chamberlain (2022) that includes linear models as a special case, very little is known about panel data models with unrestricted feedback.<sup>1</sup>

We characterize the set of feedback and heterogeneity robust (FHR) moment conditions in nonlinear panel data models with feedback. We work in a likelihood setting where the outcome density depends on a finite-dimensional parameter, and both the feedback process and the unobserved heterogeneity distribution are unrestricted. Our results cover both the finite-dimensional parameter indexing the parametric part of the model, as well as estimands which involve averages over the distribution of unobserved heterogeneity and feedback (e.g., average partial effects, average treatment effects and other average effects). We also characterize semiparametric efficiency bounds for the common parameter and average effects. We

---

<sup>1</sup>Buchinsky et al. (2010) show how to compute semiparametric efficiency bounds in dynamic discrete choice models that feature feedback. Recently, in independent work, Botosaru et al. (2024) and Chesher et al. (2024) propose innovative approaches to static and dynamic nonlinear panel data models.

demonstrate how these results may be used to find feasible estimating equations with good efficiency properties in practice.

To illustrate the power of our approach, we include a complete characterization of all FHR moments in the multi-spell mixed proportional hazards (MPH) model with both feedback and lagged duration dependence. Heckman and Borjas (1980) emphasized the importance of incorporating these phenomena into duration analysis, although we are not aware of methods for doing so beyond those appearing in the unpublished dissertation of Woutersen (2000). Hahn (1994) studied efficiency bounds in the multi-spell MPH model under strict exogeneity and no lagged duration dependence (see also Ridder and Woutersen, 2003 for related work on estimation of MPH models).

In the next section we formally define the class of semiparametric nonlinear panel data models with feedback. Section 2 presents our first main result: a complete characterization of the set of all feedback and heterogeneity robust (FHR) moment conditions. This extends the characterization obtained by functional differencing (Bonhomme, 2012), which requires strict exogeneity, to models with unrestricted feedback. Section 3 provides a similar characterization for average effects. Section 4 presents the semiparametric efficiency bound analysis. There we demonstrate that the orthogonal complement of the nuisance tangent set coincides with the set of all FHR estimating equations. This result has important implications for efficient estimation. Specifically, we show how to construct FHR moment functions that have good efficiency properties, leading to locally efficient estimators in the sense of Newey (1990). Throughout we use the MPH model to illustrate key results in a concrete setting. Our MPH results are novel and of independent interest. Finally, in Section 6, we touch on a number of important additional issues, including existence of FHR moments, regularization, and additional examples (several of which are novel). A version of the paper that combines the main text, appendices, and all supplementary materials in a single file is available [here].

**Notation.** In what follows we generally suppress the  $i$  subscript when referring to a single random draw from the cross-sectional population. Hence, for example,  $Y_t$  denotes the period  $t$  outcome of a randomly sampled unit and  $A$  its unobserved, time-invariant, attribute. We let  $Z^t = (Z_t, Z_{t-1}, \dots)$  denote the entire observed history of  $Z_t$  and  $Z^{s:t} = (Z_s, \dots, Z_t)$  its history from periods  $s \leq t$  to  $t$ .

## 1 Panel data models with feedback

In this section we introduce the semiparametric panel data model with feedback, connect this model to the more restrictive one which maintains strict exogeneity, and formally state our

main research questions. Throughout this section, and those that follow, we illustrate key results in the context of a multi-spell mixed proportional hazards (MPH) model with lagged duration dependence and feedback.

## 1.1 Setup

Let  $\{(X_{i1}, \dots, X_{iT}, Y_{i0}, Y_{i1}, \dots, Y_{iT}, A_i)\}_{i=1}^\infty$  be an independently and identically distributed random sequence drawn from some distribution function  $F$ . The sole prior restriction on  $F$  is that the conditional density of  $Y_t$  at  $y_t$  given the past  $y^{t-1}$ , regressor history  $x^t$ , and latent unit-specific heterogeneity  $a$ , belongs to a known parametric family indexed by the unknown parameter  $\theta \in \Theta \subset \mathbb{R}^K$ :

$$f(y_t | y^{t-1}, x^t, a) = f_\theta(y_t | y^{t-1}, x^t, a) = f_\theta(y_t | y_{t-1}, x_t, a), \quad t = 1, \dots, T, \quad (1)$$

for some  $\theta \in \Theta$ . The density  $f_\theta(y_t | y_{t-1}, x_t, a)$  is the parametric component of our setup.<sup>2</sup>

Familiar nonlinear examples of (1) include binary choice logit models (Chamberlain, 1980, 2010; Bonhomme et al., 2023; Honoré and Weidner, 2024) and count models (Chamberlain, 1992, 2022; Wooldridge, 1997). Specific forms for  $f_\theta(y_t | y_{t-1}, x_t, a)$  also arise in the context of dynamic structural models (e.g., Aguirregabiria and Mira, 2010). Observe that the parametric families,  $f_\theta(y_t | y_{t-1}, x_t, a)$  and  $f_\theta(y_s | y_{s-1}, x_s, a)$  need not coincide for  $s \neq t$ ; this allows for time effects and other forms of nonstationarity.

**Example 1. (MIXED PROPORTIONAL HAZARDS (MPH) MODEL)** Let  $\{Y_t\}_{t=1}^T$  denote a sequence of durations, or spell lengths, with  $X_t$  a corresponding vector of beginning-of-spell covariates;  $Y_0$  is an “initial duration”. For example,  $Y_t$  might equal time to re-arrest following an agent’s  $t^{th}$  release from prison while  $X_t$  might include measures of their post-release support and supervision. Since support and supervision,  $X_t$ , might depend on the previous time to re-arrest,  $Y_{t-1}$ , as well as unmeasured agent attributes,  $A$ , heterogeneous feedback is plausible.

The MPH model was introduced by Lancaster (1979) and Nickell (1980) for single-spell data. Chamberlain (1985) studied identification and estimation with multi-spell data under strict exogeneity. Hahn (1994) derived the semiparametric efficiency bound for both the single- and multi-spell case (the latter under strict exogeneity). Heckman and Borjas (1980) emphasize the relevance of lagged duration dependence and feedback in labor market applications.

---

<sup>2</sup>While (1) imposes that only the contemporaneous regressor and the first lag of the outcome matter, additional lags could be easily accommodated in what follows.

The instantaneous conditional hazard rate is given by

$$\lambda(y_t | Y^{t-1} = y^{t-1}, X^t = x^t, A = a) = \lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x_t' \beta + a), \quad (2)$$

with  $\lambda_\alpha(y_t)$  a known parametric family of baseline hazard functions indexed by  $\alpha$  (we emphasize the Weibull case with  $\lambda_\alpha(y_t) = \alpha y_t^{\alpha-1}$  below). Under (2) the conditional density at  $Y_t = y_t$  is

$$f_\theta(y_t | y^{t-1}, x^t, a) = \lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x_t' \beta + a) \exp(-\rho_\theta(z_t) e^a), \quad (3)$$

for  $z_t = (y_t, y_{t-1}, x_t')'$ ,  $\rho_\theta(z_t) = \Lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x_t' \beta)$ ,  $\theta = (\alpha', \beta', \gamma)'$ , and  $\Lambda_\alpha(y_t) = \int_0^{y_t} \lambda_\alpha(u) du$  the integrated baseline hazard.

**Example 2. (POISSON MODEL)** Let  $\{Y_t\}_{t=1}^T$  denote a sequence of counts, for example the number of patents awarded to a firm in a year, as in Blundell et al. (2002), and  $X_t$  a vector of time-varying regressors. For  $t = 1, \dots, T$  we have

$$Y_t | Y^{t-1}, X^t, A \sim \text{Poisson}(\exp(\gamma Y_{t-1} + X_t' \beta + A)), \quad (4)$$

with  $Y_0$  an initial count. Chamberlain (1992) and Wooldridge (1997) proposed GMM estimators for  $\theta = (\beta', \gamma)'$  in this model. Windmeijer (2008) reviews extant results.

Returning to our general setup, sequentially factorizing the joint density of  $Y_0, Y_1, \dots, Y_T, X_1, \dots, X_T, A$  at  $y_0, y_1, \dots, y_T, x_1, \dots, x_T, a$  yields the following expression for the likelihood contribution of a single unit:

$$\begin{aligned} \ell(\theta, g, \pi, \nu | y^T, x^T) &= \left[ \prod_{t=2}^T f(x_t, y_t | x^{t-1}, y^{t-1}, a) \right] f(y_1 | y_0, x_1, a) f(a | y_0, x_1) f(y_0, x_1) \\ &= \underbrace{\left[ \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, a) \right]}_{\text{Parametric component}} \times \underbrace{\left[ \prod_{t=2}^T g(x_t | y^{t-1}, x^{t-1}, a) \right]}_{\text{Feedback process}} \\ &\quad \times \underbrace{\pi(a | y_0, x_1)}_{\text{Heterogeneity}} \times \underbrace{\nu(y_0, x_1)}_{\text{Initial condition}}, \end{aligned} \quad (5)$$

where the second equality follows by imposing the parametric assumption (1) and establishing the following notations:

- (i)  $g(x_t | y^{t-1}, x^{t-1}, a)$ , denotes the density of  $X_t$  at  $x_t$  given the past  $(y^{t-1}, x^{t-1})$  and heterogeneity  $a$ ;

- (ii)  $\pi(a|y_0, x_1)$ , the conditional density of  $A$  at  $a$  given the initial condition  $(y_0, x_1)$ ; and
- (iii)  $\nu(y_0, x_1)$ , the initial condition density.

While  $f_\theta(y_t|y^{t-1}, x^t, a)$  belongs to a parametric family, the remaining components, items (i) to (iii) above, are all unrestricted. We call this model the semiparametric panel data model with feedback.

In what follows we call the  $T-1$  densities  $g(x_T|y^{T-1}, x^{T-1}, a)$ ,  $g(x_{T-1}|y^{T-2}, x^{T-2}, a)$ , ...,  $g(x_2|y^1, x_1, a)$  the *feedback process*. This process describes how past values of the outcome influence current regressor values. Because it depends on  $A$ , the feedback process is *heterogeneous* across units. We do not impose any stationarity over  $t = 1, \dots, T$ , so  $g(x_t|y^{t-1}, x^{t-1}, a)$  and  $g(x_s|y^{s-1}, x^{s-1}, a)$  for  $s \neq t$  may differ arbitrarily. When  $X_t$  is a policy variable, such flexibility accommodates un-modeled regime shifts (e.g., as when a change in tax policy alters firm investment behavior). We use the notation  $g$  to denote a generic element of the set of all allowable feedback processes  $\mathcal{G}$ . We call  $\pi(a|y_0, x_1)$  the *heterogeneity distribution*; this density describes the distribution of unobserved heterogeneity,  $A$ , across units as well as any dependence of this heterogeneity on the initial condition,  $(y_0, x_1)$ . Let  $\pi$  denote a generic element of the set of all allowable heterogeneity distributions,  $\Pi$ . Finally we let  $\nu \in \mathcal{N}$  denote an element of the set of all possible initial condition densities.

Although not emphasized in our exposition, it is straightforward to incorporate strictly exogenous regressors into the feedback model. Similarly, additional sources of heterogeneity, beyond  $A$ , can enter the feedback process. This generality, while important in some applications, involves no new issues and clutters notation.<sup>3</sup> We also note that, although not emphasized in most of our examples,  $Y_t$  may be vector-valued in some settings.

Panel data models with feedback and heterogeneity arise frequently in economic applications. In structural dynamic choice models, agents' dynamic optimization typically leads to a likelihood function of the form (5), where  $Y_t$  contains choice variables of interest, as well as payoff variables, and  $X_t$  contains dynamic state variables and un-modeled choice variables; see Aguirregabiria and Mira (2010) for an exposition in the case of models with discrete out-

---

<sup>3</sup>To be specific: let  $W^T$  contain all leads and lags of a vector of strictly exogenous regressors and  $B$  an additional source of heterogeneity. The results that follow are easily modified to accommodate the richer model:

$$\ell(\theta, g, \pi, \nu | y^T, x^T, w^T) = \left\{ \int \int \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, w_t, a) \times \left[ \prod_{t=2}^T g(x_t | y^{t-1}, x^{t-1}, w^T, a, b) \right] \right. \\ \left. \times \pi(a, b | y_0, x_1, w^T) da db \right\} \times \nu(y_0, x_1, w^T),$$

where we assume that only the contemporaneous value of  $W_t$  enters the parametric part of the model for simplicity.

comes. The moment conditions we derive are robust to any possible process for the dynamic state variables and distribution of unobserved heterogeneity.

Feedback also arises in program evaluation settings. Let  $Y_t$  denote earnings and  $X_t$  recent past participation in a job-training program. Ashenfelter (1978) observed that Manpower Development and Training Act (MDTA) trainees had unusually low earnings in the year prior to undertaking training, consistent with a behavioral model where poor labor market outcomes (low values of  $Y_{t-1}$ ) induced agents to seek out training in the next period ( $X_t = 1$ ). In contrast, maintaining no feedback would require that, conditional on the latent attribute,  $A$ , a worker's labor market history has no bearing on the decision to undertake training; a rather strong assumption (Ashenfelter and Card, 1985).

**Remark 1.1. (STRICT EXOGENEITY)** In applications, researchers routinely maintain strict exogeneity of the regressors,  $X_t$ , conditional on the latent attribute,  $A$ . Strict exogeneity imposes independence of  $Y_t$  and  $(X_{t+1}, \dots, X_T)$  conditional on  $(Y_0, X_1, \dots, X_t, A)$ . Chamberlain (1982a) shows that strict exogeneity is equivalent to the following *no feedback* condition:

$$g(x_t | y^{t-1}, x^{t-1}, a) = g(x_t | y_0, x^{t-1}, a), \quad t = 2, \dots, T. \quad (6)$$

While (5) allows  $Y_s$  to influence  $X_t$  for  $s < t$ , (6) rules out such feedback (we allow dependence on the initial condition throughout). Restriction (6) is a counterpart of Granger's (1969) definition of "Y does not cause X" in the panel data setting with latent heterogeneity. If  $X_t$  is a choice variable, and  $Y^{t-1}$  is a component of the agent's begining-of-period  $t$  information set, then (6) typically implies strong restrictions on economic behavior (e.g., Ashenfelter and Card, 1985) and/or the structure of agent's information sets (e.g., Chamberlain, 1984, 1985).<sup>4</sup> Our approach, by accommodating unrestricted feedback and heterogeneity, allows the researcher to proceed without maintaining such assumptions.

## 1.2 The goal: feedback and heterogeneity robust estimation

In this paper we study estimation of the common parameter,  $\theta$ , in panel data models when (1) is the *only* prior restriction on  $F$  (except for some mild regularity conditions). We wish to construct estimators that are consistent irrespective of the precise instances of the

---

<sup>4</sup>Under strict exogeneity, the likelihood function becomes

$$\ell^{\text{SE}}(\theta, \pi, \nu | y^T, x^T) = \left\{ \int \left[ \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, a) \right] \pi(a | x_1, \dots, x_T, y_0) da \right\} \nu(y_0, x^T), \quad (7)$$

where now the distribution of unobserved heterogeneity  $\pi(a | x_1, \dots, x_T, y_0)$  is conditional on covariates in *all* periods.

feedback process,  $g$ , heterogeneity distribution,  $\pi$ , and initial condition,  $\nu$ , which describe the sampled data. In what follows we call such estimators *feedback and heterogeneity robust* (FHR). Because strict exogeneity obtains as a special case, any FHR estimator remains valid under strict exogeneity. FHR estimators are natural extensions of familiar “fixed effects” approaches to estimation in panel data models without feedback. We fully characterize the set of moment-based FHR estimators. We also derive semiparametric efficiency bounds for  $\theta$ . In particular, our results allow for a precise quantification of the information loss associated with accommodating unrestricted feedback relative to maintaining strict exogeneity.

One alternative to FHR estimation involves assuming that both the feedback process,  $g$ , and heterogeneity density,  $\pi$ , belong to parametric families indexed by some parameter vector  $\eta$ . With these additional maintained assumptions, the econometrician can then maximize the resulting likelihood, conditional on  $Y_0$  and  $X_1$ , with respect to both  $\theta$  and  $\eta$ . This transforms the problem from a semiparametric to a parametric one. Consistency of such parametric “random effects” estimators typically requires that the additionally maintained parametric restrictions on  $g$  and  $\pi$  hold in the sampled population. Consequently, such estimators are *not* generally feedback and heterogeneity robust.

We next formally define the FHR property. Let  $\theta \in \Theta$ , and  $\omega = (g, \pi, \nu) \in \Omega$  collect all the nuisance parameters in our model. We assume that all elements  $\omega \in \Omega$  have a common support, known to the econometrician (which may be unbounded and include the full real line, for example, for the support of the heterogeneity  $A$ ). Let  $\mathbb{E}_{\theta, \omega} [\cdot]$  denote an expectation taken under the DGP at  $(\theta, \omega)$ . Let  $\phi_\theta(y_0, y_1, \dots, y_T, x_1, \dots, x_T)$  be a function of the observed data indexed by  $\theta$ . We say that  $\phi_\theta$  is a *FHR moment function* if, for all  $\omega$  such that  $\phi_\theta$  is absolutely integrable under DGP  $(\theta, \omega)$ , we have

$$\mathbb{E}_{\theta, \omega} [\phi_\theta(Y_0, Y_1, \dots, Y_T, X_1, \dots, X_T)] = 0. \quad (8)$$

The main goal of the paper is to derive moment restrictions  $\phi_\theta$  that have the FHR property. We will first provide a complete characterization of FHR moment restrictions for  $\theta$  in Section 2.<sup>5</sup> Then, in Section 3 we will show how essentially the same analysis can be used to provide a characterization of FHR moment functions for average effects of the form

$$\mu(\theta, \omega) = \mathbb{E}_{\theta, \omega} [h_\theta(Y^T, X^T, A)],$$

where  $h_\theta(Y^T, X^T, A)$  is a known function of  $(Y^T, X^T, A)$  indexed by  $\theta$ . Average effects

---

<sup>5</sup>In fact, our characterization of FHR moment functions remains valid if  $\theta$  is infinite-dimensional (for example, if  $\theta$  contains nonparametric elements such as functions). However, the finite-dimensional  $\theta$  case contains many important models, and all the examples we will use as illustrations feature a finite-dimensional  $\theta$  vector. We also focus on this case in our analysis of efficiency.

include a variety of causal or structural parameters, such as average partial effects, as in Chamberlain (1984), and average structural functions, as in Blundell and Powell (2003).

Since any FHR moment function is also valid under strict exogeneity, the set of all FHR moment conditions for a given panel data model will be a subset of the corresponding set of moment conditions derived under strict exogeneity (and characterized in Bonhomme, 2012). In some cases the latter set may be non-trivial and the former empty. For example, Chamberlain (2010) showed point identification of the panel logit model under strict exogeneity, while Bonhomme et al. (2023) show a failure of identification in this model under feedback (see also Section 6 below).

## 2 FHR moment restrictions for common parameters

This section presents our first main result: a characterization of all FHR moment conditions for  $\theta$ . As an application, we also specialize our results to provide a constructive characterization of the set of all possible FHR moment conditions for the MPH model.

### 2.1 Main characterization of FHR moments for $\theta$

We begin by characterizing the set of FHR moments for  $\theta$ .

**Theorem 2.1.** (FHR MOMENT CHARACTERIZATION). *Let  $\theta \in \Theta$ . (A) Suppose that the following  $T$  restrictions hold: (i)*

$$\int \phi_\theta(y^T, x^T) \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{1:T} = 0, \quad (9)$$

*and (ii), for all  $s = 2, \dots, T$ ,*

$$\int \phi_\theta(y^T, x^T) \prod_{t=s}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{s:T} \text{ does not depend on } x^{s:T}. \quad (10)$$

*Then, for all  $\omega = (g, \pi, \nu) \in \Omega$  such that  $\mathbb{E}_{\theta, \omega} [|\phi_\theta(Y^T, X^T)|] < \infty$ , we have*

$$\mathbb{E}_{\theta, \omega} [\phi_\theta(Y^T, X^T)] = 0.$$

(B) Suppose that, (i) the root density  $\omega \mapsto \ell^{1/2}(\theta, \omega | y^T, x^T)$  is differentiable in quadratic mean at  $\omega^*$  for some  $\omega^* = (g^*, \pi^*, \nu^*) \in \Omega$ , and (ii) there is a neighborhood  $\mathcal{N}$  of  $\omega^*$  such that  $\sup_{\omega \in \mathcal{N}} \mathbb{E}_{\theta, \omega} [|\phi_\theta(Y^T, X^T)|^2] < \infty$ . Then, the converse is also true.

As an implication of Part (A) of Theorem 2.1, suppose the data is generated according to some  $(\theta_0, \omega_0)$ . Then any function  $\phi_\theta$  that satisfies (9) and (10) and is absolutely integrable under the population DGP  $(\theta_0, \omega_0)$  has zero mean at the true  $\theta_0$ , that is,

$$\mathbb{E}_{\theta_0, \omega_0}[\phi_{\theta_0}(Y^T, X^T)] = 0.$$

To see this, simply apply Part (A) of Theorem 2.1 with  $\theta = \theta_0$  and  $\omega = \omega_0$ .

The first condition for the FHR property, restriction (9), ensures robustness of  $\phi_\theta$  to the presence of heterogeneity of an unknown form. This condition is highlighted in Bonhomme (2012), in a setting with strictly exogenous covariates, as the key condition ensuring valid moment functions for  $\theta$ . Under strict exogeneity, (9) ensures that  $\phi_\theta(Y^T, X^T)$  is conditionally mean zero given  $X^T$  and  $A$ . By the law of iterated expectations, this suffices to ensure that it is unconditionally mean zero as well. The functional differencing approach then provides a general recipe for constructing functions  $\phi_\theta$  satisfying (9).

To illustrate how the presence of feedback modifies the interpretation of condition (9), consider the  $T = 2$  setting. In this case, the condition is

$$\int \left[ \int \phi_\theta(y_0, y_1, y_2, x_1, x_2) f_\theta(y_2 | y_1, x_2, a) dy_2 \right] f_\theta(y_1 | y_0, x_1, a) dy_1 = 0,$$

which coincides with

$$\mathbb{E} [\mathbb{E} [\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, Y_1, X_1, X_2 = x_2, A] | Y_0, X_1, A] = 0 \quad \text{for all } x_2. \quad (11)$$

Here the inner expectation corresponds to an average over  $y_2$  with respect to its model density,  $f_\theta(y_2 | y_1, x_2, a)$ . This inner expectation conditions on  $X_2 = x_2$ , while the outer expectation, which averages over  $y_1$  with respect to its model density,  $f_\theta(y_1 | y_0, x_1, a)$ , does not condition on  $x_2$ . This is a key difference between the heterogeneous feedback case, considered here, and the setting with strict exogeneity studied by Bonhomme (2012). Under strict exogeneity,  $Y_1$  is independent of  $X_2$  conditional on  $(Y_0, X_1, A)$ , so (11) coincides with

$$\mathbb{E} [\mathbb{E} [\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, Y_1, X_1, X_2, A] | Y_0, X_1, X_2, A] = 0,$$

with both the inner and outer expectations conditioning on  $X_2$ . By iterated expectations, this conditional expectation implies a zero mean condition given  $A$ , initial conditions, and the *entire sequence of covariates*,

$$\mathbb{E} [\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, X_1, X_2, A] = 0.$$

However, under feedback it is not generally the case that (11) can be written as a zero mean condition given the covariates sequence  $(X_1, X_2)$ .

In the presence of feedback, Theorem 2.1 additionally requires the  $T - 1$  conditions (10), to ensure that  $\phi_\theta$  is a valid moment function. To explain this additional requirement, consider again the  $T = 2$  setting, in which case (10) reads

$$\int \phi_\theta(y_0, y_1, y_2, x_1, x_2) f_\theta(y_2 | y_1, x_2, a) dy_2 \quad \text{does not depend on } x_2,$$

which corresponds to the single mean independence restriction

$$\mathbb{E} [\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, Y_1, X_1, X_2, A] = \mathbb{E} [\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, Y_1, X_1, A]. \quad (12)$$

Equation (12) implies that, at the population parameter,  $X_2$  does not predict  $\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2)$  conditional on  $Y_0, Y_1, X_1, A$ . Under this condition, which we interpret as “feedback robustness” of  $\phi_\theta$ , the conditioning on  $X_2 = x_2$  disappears in (11), which implies

$$\mathbb{E} [\mathbb{E} [\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, Y_1, X_1, A] | Y_0, X_1, A] = 0,$$

ensuring, by iterated expectations, that  $\phi_\theta$  is a valid moment function. Importantly,  $\phi_\theta$  is then a valid moment under both unrestricted heterogeneity and unrestricted feedback.

Lastly, Theorem 2.1 also establishes that (9) and (10) are, under suitable conditions, *necessary* for the FHR property. To show the necessity property, we assume quadratic mean differentiability at  $(\theta, \omega^*)$  as stated in Part (B) Condition (i). This is a standard regularity condition in semiparametric estimation, see for example Van der Vaart (2000, Ch. 25). Part (B) Condition (ii), which imposes square-integrability of the moment function in a neighborhood of  $\omega^*$ , is similarly standard; e.g., see the assumptions for the generalized information equality in Lemma 5.4 of Newey and McFadden (1994).<sup>6</sup> We will rely on differentiability in quadratic mean in our analysis of efficiency in Section 4.

## 2.2 Alternative characterization of FHR moments for $\theta$

The following corollary to Theorem 2.1 facilitates the construction of FHR moment functions in practice.

**Corollary 2.1.1. (ALTERNATIVE FHR MOMENT CHARACTERIZATION)** *Let  $\omega^* \in \Omega$ . An absolutely-integrable function  $\phi_\theta$  under  $(\theta, \omega^*)$  satisfies (9) and (10) if and only if there exist*

---

<sup>6</sup>One can show that, if  $\phi_\theta$  is bounded, then the necessity of (9) and (10) obtains directly without reference to  $\omega^*$  or differentiability in quadratic mean.

absolutely-integrable functions  $\psi_{\theta,t}$  under  $(\theta, \omega^*)$ , for  $t = 1, \dots, T - 1$ , such that:

$$\mathbb{E} [\phi_\theta(Y^T, X^T) | Y^{T-1}, X^T, A] = \sum_{t=1}^{T-1} \psi_{\theta,t}(Y^t, X^t, A), \quad (13)$$

with  $\mathbb{E} [\psi_{\theta,t}(Y^t, X^t, A) | Y^{t-1}, X^t, A] = 0$  for  $t = 1, \dots, T - 1$ .

To understand Corollary 2.1.1, it is helpful to return to the  $T = 2$  setting. In this case, letting

$$\psi_{\theta,1}(Y_0, Y_1, X_1, A) = \mathbb{E}[\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2, A) | Y_0, Y_1, X_1, A],$$

it follows from (9) that

$$\psi_{\theta,1}(Y_0, Y_1, X_1, A) = \mathbb{E}[\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2, A) | Y_0, Y_1, X_1, X_2, A],$$

which implies (13), whereas it follows from (10) that

$$\mathbb{E}[\psi_{\theta,1}(Y_0, Y_1, X_1, A) | Y_0, X_1, A] = 0$$

(a requirement for  $\psi_{\theta,1}$  given in the corollary).

The representation provided by Corollary 2.1.1 suggests a systematic recipe for constructing FHR moment functions. The first step involves finding functions  $\psi_{\theta,t}(Y^t, X^t, A)$  that are mean zero conditional on  $Y^{t-1}, X^t, A$  for  $t = 1, \dots, T - 1$ . This is straightforward since any function of  $(Y^t, X^t, A)$ , suitably de-meaned, automatically fulfills this requirement (see Section 6 for several examples). Then, given a collection of  $\psi_{\theta,t}(Y^t, X^t, A)$  functions, the second step requires solving a linear integral equation to recover a valid moment function  $\phi_\theta$ . Indeed, (13) can equivalently be written as

$$\int \phi_\theta(y^T, x^T) f_\theta(y_T | y_{T-1}, x_T, a) dy_T = \sum_{t=1}^{T-1} \psi_{\theta,t}(y^t, x^t, a), \quad (14)$$

which is known as an inhomogeneous Fredholm equation of the first kind. Note that the integral operator on the left-hand side of (14) is known given the parameter  $\theta$ . Solution methods for this type of linear integral equation are the subject of a large literature (e.g., Engl et al., 1996; Carrasco et al., 2007). In Section 6 we will show how to construct FHR moment functions in some specific models using Corollary 2.1.1, and discuss when such functions exist.

A special case of Corollary 2.1.1 obtains when one is able to find functions  $\eta_{\theta,t}$  such that

$$\mathbb{E} [\eta_{\theta,t}(Y^t, X^t) | Y^{t-1}, X^t, A] = b(Y_0, X_1, A), \quad (15)$$

for some function  $b$  of the heterogeneity and initial conditions. Then, the instrumented first difference

$$\phi_\theta(y^T, x^T) = [\eta_{\theta,T}(y^T, x^T) - \eta_{\theta,T-1}(y^{T-1}, x^{T-1})] \cdot m(y^{T-2}, x^{T-1})$$

satisfies (13), for  $\psi_{\theta,T-1}(y^{T-1}, x^{T-1}, a) = [b(y_0, x_1, a) - \eta_{\theta,T-1}(y^{T-1}, x^{T-1})] \cdot m(y^{T-2}, x^{T-1})$  and  $\psi_{\theta,t} = 0$  for all  $t < T - 1$  (for an arbitrary function  $m$ ). This provides a FHR moment function on  $\theta$ . More generally, one can check that

$$\phi_\theta(y^t, x^t) = [\eta_{\theta,t}(y^t, x^t) - \eta_{\theta,t-1}(y^{t-1}, x^{t-1})] \cdot m(y^{t-2}, x^{t-1}), \quad t = 2, \dots, T, \quad (16)$$

all satisfy (13), thus providing additional moments.

We now illustrate this particular recipe with the MPH model.

**Example 1** (Continued). (SIMPLE MOMENTS FOR THE MPH) It is convenient to first develop some additional implications of the MPH model under feedback. Recall the notations  $z_t = (y_t, y_{t-1}, x'_t)'$  and  $\rho_\theta(z_t) = \Lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x'_t \beta)$ . Adapting arguments appearing in Lancaster (1990), Hahn (1994) and Ridder and Woutersen (2003), it is straightforward to show that

$$\rho_\theta(Z_t) | Y^{t-1}, X^t, A \sim \text{Exponential}(e^A), \quad t = 1, \dots, T, \quad (17)$$

(see Lemma D.1 in Supplemental Appendix D). From this observation we have

$$\mathbb{E} [\rho_\theta(Z_t) | Y^{t-1}, X^t, A] = e^{-A}, \quad t = 1, \dots, T, \quad (18)$$

which coincides with (15) after setting  $\eta_{\theta,t}(Y^t, X^t) = \rho_\theta(Z_t)$  and  $b(Y_0, X_1, A) = e^{-A}$ . This yields the FHR moment functions

$$\phi_\theta(Y^t, X^t) = [\rho_\theta(Z_t) - \rho_\theta(Z_{t-1})] \cdot m(Y^{t-2}, X^{t-1}), \quad t = 2, \dots, T, \quad (19)$$

where  $m$  is an arbitrary function. In unpublished dissertation research, Woutersen (2000, Ch. 1) presented moments similar to (19).

Moments of the form (16) were considered by Arellano and Bond (1991) in the linear model context, by Chamberlain (1992, 2022) and Wooldridge (1997) for Poisson models, and by Al-Sadoon et al. (2017) for certain binary choice models. However, this family of estimating

equations does not exhaust all available FHR moments, and may lead to estimators with low levels of asymptotic precision in practice. By comparison, Theorem 2.1 and Corollary 2.1.1 characterize *all* available FHR moments, as we will now illustrate in the case of the MPH model. Moreover, as we will establish in later sections, our characterization can be used to derive efficient estimators (i.e., we extend to nonlinear models efficiency arguments which appear in the linear panel data literature such as in Arellano and Bover, 1995; Arellano and Honoré, 2001; Arellano, 2016).

### 2.3 FHR moments for $\theta$ in the MPH model

For specific models it is sometimes possible to use Theorem 2.1 to provide a direct characterization of all FHR moment functions. We show how this can be done in the MPH model with feedback in Lemma 2.2 below. Our result provides insight into the MPH model and simplifies the derivation of new FHR moments.

Our characterization makes use of several special features of the MPH model, which we introduce first (details can be found in Supplemental Appendix D). Let  $P_{\theta,t} = \rho_\theta(Z_t)$ . As indicated in (17), conditional on  $Y_0, X_1, A$ , the  $P_{\theta,t}$  for  $t = 1, \dots, T$  are independent exponential random variables; each with a common rate parameter of  $e^A$ .

Next, we define a one-to-one transformation of the vector  $P_\theta^T$  into a “forward orthogonal deviations” part and a “between” part, as follows:

$$\begin{aligned}\tilde{P}_{\theta,t} &= \frac{\rho_\theta(Z_t)}{\sum_{s=t}^T \rho_\theta(Z_s)}, \quad t = 1, \dots, T-1, \\ \bar{P}_\theta &= \sum_{t=1}^T \rho_\theta(Z_t).\end{aligned}\tag{20}$$

Observe that  $\tilde{P}_{\theta,t}$  involves the ratio of  $\rho_\theta(Z_t)$  to the sum of itself and the *future* values of  $\rho_\theta(Z_s)$  for  $s = t+1, \dots, T$ . Lemma D.2 in Supplemental Appendix D establishes that, conditionally on  $Y_0, X_1, A$ :

$$\begin{aligned}\tilde{P}_{\theta,t} &\sim \text{Beta}(1, T-t), \quad t = 1, \dots, T-1, \\ \bar{P}_\theta &\sim \text{Gamma}(T, e^A),\end{aligned}\tag{21}$$

where, throughout,  $\theta$  is the parameter indexing the DGP. Lemma D.2 additionally establishes that the elements of  $(\tilde{P}_{\theta,1}, \dots, \tilde{P}_{\theta,T-1}, \bar{P}_\theta)$  are mutually independent of one another.

Transformation (20) can be thought of as a MPH-specific analog of the forward orthogonal deviations transformation used by Arellano and Bover (1995) in the context of linear panel data models with predetermined regressors. It has the property that the random variables

$\tilde{P}_{\theta,t}$  are independent of both contemporaneous and lagged predetermined covariates as well as lagged values of the spell outcomes  $\{(X_{is}, Y_{is-1})\}_{s=1}^t$  (see part (i) of Lemma D.1). Appealing to the same analogy, we can think of  $\bar{P}_\theta$  as containing the “between” variation in  $P_\theta^T$ .

With these preliminaries in place, we can state the following FHR moment characterization for the MPH model with feedback.

**Lemma 2.2.** (CHARACTERIZATION OF FHR MOMENT RESTRICTIONS FOR THE MPH MODEL) *Consider the MPH model with  $T \geq 2$ . Let  $\omega^* \in \Omega$ . Then an absolutely-integrable  $\phi_\theta$  under  $(\theta, \omega^*)$  satisfies (9) and (10) if and only if there exist absolutely-integrable  $\psi_{\theta,t}$  under  $(\theta, \omega^*)$ , for  $t = 1, \dots, T$ , such that:*

$$\phi_\theta(Y^T, X^T) = \sum_{t=1}^{T-1} \psi_{\theta,t}(Y_0, \tilde{P}_\theta^t, \bar{P}_\theta, X^t),$$

where, for  $t = 1, \dots, T-1$ ,

$$\mathbb{E} \left[ \psi_{\theta,t}(Y_0, \tilde{P}_\theta^t, \bar{P}_\theta, X^t) \middle| Y_0, \tilde{P}_\theta^{t-1}, \bar{P}_\theta, X^t \right] = 0.$$

The proof of Lemma 2.2 is available in Supplemental Appendix B.2. It represents a useful simplification, induced by the special structure of the MPH model, relative to the general result of Theorem 2.1. Note in particular that the latent heterogeneity  $A$  does not appear in the characterization of Lemma 2.2. To illustrate, consider the  $T = 2$  setting. In this case the lemma implies that all FHR moment functions are of the form

$$\phi_\theta(Y_0, Y_1, Y_2, X_1, X_2) = \psi_{\theta,1}(Y_0, \tilde{P}_{\theta,1}, \bar{P}_\theta, X_1),$$

where  $\mathbb{E} \left[ \psi_{\theta,1}(Y_0, \tilde{P}_{\theta,1}, \bar{P}_\theta, X_1) \middle| Y_0, \bar{P}_\theta, X_1 \right] = 0$ . We seek functions  $\psi_{\theta,1}$  of the forward orthogonal deviations  $\tilde{P}_{\theta,1}$ , that are mean zero conditional on the between variation  $\bar{P}_\theta$ , as well as the initial condition  $(Y_0, X_1)$ . It is straightforward to construct such functions by de-meaning, and we will exploit this property in our analysis of efficiency.

For the general case of an arbitrary number  $T$  of periods,  $\phi_\theta$  is the sum of functions  $\psi_{\theta,t}(Y_0, \tilde{P}_\theta^t, \bar{P}_\theta, X^t)$ , for  $t = 1, \dots, T-1$ , that are mean independent of the between variation,  $\bar{P}_\theta$ , current and past values of the predetermined regressors,  $X^t$ , and past values of the spell outcomes,  $Y^{t-1}$ . This structure is reminiscent of how moment conditions are typically constructed for linear panel data models with predetermined regressors (see, especially, Arellano and Bover, 1995).

### 3 FHR moment restrictions for average effects

In this section we characterize all FHR moment conditions for average effects,

$$\mu(\theta, \omega) = \mathbb{E}_{\theta, \omega} [h_\theta(Y^T, X^T, A)],$$

where  $h_\theta(Y^T, X^T, A)$  is a known function of  $Y^T, X^T, A$  indexed by  $\theta$ .

**Example 1** (Continued). (AVERAGE EFFECTS IN THE MPH MODEL) Consider the *average structural hazard (ASH)* function

$$\bar{\lambda}(y_t|x_t, y_{t-1}) = \mathbb{E}_{\theta, \omega} [\lambda_\alpha(y_t) e^{x'_t \beta + \gamma y_{t-1} + A}]. \quad (22)$$

The ASH appears to be a new estimand, albeit a natural one given concerns about spurious duration dependence (Lancaster, 1990; Heckman, 1991). In the context of the MPH model, the ASH corresponds to the mean survival time for a unit exogenously assigned to policy  $X_t = x_t$  and history  $Y_{t-1} = y_{t-1}$ . Like other quantities relevant to causal analysis, the ASH depends on the (marginal) distribution of unobserved heterogeneity  $A$ .<sup>7</sup>

In nonlinear panel data models, knowledge of  $\theta$  does not suffice to identify the effect of an external manipulation of a regressor's value on the probability distribution of the outcome. This follows from the nonseparable way in which the unobserved heterogeneity enters such models. In contrast, estimands which average over the marginal distribution of  $A$ , such as (22), do provide easy-to-interpret *summaries* of such effects.

Until recently the identifiability of such averages was not well understood. Recent work by Honoré and Tamer (2006), Chernozhukov et al. (2013), Aguirregabiria and Carro (2021), Davezies et al. (2021), Dobronyi et al. (2021), Pakel and Weidner (2023) and others, however, has shown that average effects are (partially) identified in several specific settings of interest. The study of average effects in more general settings, such as those with feedback, as we consider here, remains underdeveloped (see, for example, Bonhomme et al., 2023).

#### 3.1 Characterization of FHR moments for $\mu(\theta, \omega)$

Our first result is an analog of Theorem 2.1 for  $\mu(\theta, \omega)$ .

---

<sup>7</sup>A related average effect of interest is the *average structural function (ASF)* (Blundell and Powell, 2004):

$$\mu(x_t, y_{t-1}) = \mathbb{E}_{\theta, \omega} [m(x_t, y_{t-1}, A; \theta)], \quad (23)$$

where  $m(x_t, y_{t-1}, a; \theta) = \mathbb{E}[Y_t | X_t = x_t, Y_{t-1} = y_{t-1}, A = a]$ .

**Theorem 3.1.** (CHARACTERIZATION OF MOMENT RESTRICTIONS FOR AVERAGE EFFECTS)  
*Let  $\theta \in \Theta$ . (A) Suppose that the following  $T$  restrictions hold: (i)*

$$\int (\varphi_\theta(y^T, x^T) - h_\theta(y^T, x^T, a)) \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{1:T} = 0, \quad (24)$$

and (ii), for all  $s = 2, \dots, T$ ,

$$\int (\varphi_\theta(y^T, x^T) - h_\theta(y^T, x^T, a)) \prod_{t=s}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{s:T} \text{ does not depend on } x^{s:T}. \quad (25)$$

Then, for all  $\omega \in \Omega$  such that  $\mathbb{E}_{\theta, \omega} [|\varphi_\theta(Y^T, X^T)|] < \infty$  and  $\mathbb{E}_{\theta, \omega} [|h_\theta(Y^T, X^T, A)|] < \infty$  we have

$$\mathbb{E}_{\theta, \omega} [\varphi_\theta(Y^T, X^T)] = \mu(\theta, \omega).$$

(B) Suppose (i) the root density  $\omega \mapsto \ell^{1/2}(\theta, \omega | y^T, x^T)$  is differentiable in quadratic mean at  $\omega^*$  for some  $\omega^* = (g^*, \pi^*, \nu^*) \in \Omega$ , (ii) there is a neighborhood  $\mathcal{N}$  of  $\omega^*$  such that  $\sup_{\omega \in \mathcal{N}} \mathbb{E}_{\theta, \omega} [\|\varphi_\theta(y^T, x^T)\|^2] < \infty$  and  $\sup_{\omega \in \mathcal{N}} \mathbb{E}_{\theta, \omega} [|h_\theta(Y^T, X^T, A)|^2] < \infty$ . Then, the converse is also true.

Note the strong parallel between this theorem and Theorem 2.1. As in the case of  $\theta$ , (24) and (25) imply that, under absolute integrability, the true value  $\mu_0 = \mu(\theta_0, \omega_0)$  satisfies a moment condition. Indeed, if  $\mathbb{E}_{\theta_0, \omega_0} [|\varphi_{\theta_0}(Y^T, X^T)|] < \infty$  and  $\mathbb{E}_{\theta_0, \omega_0} [|h_{\theta_0}(Y^T, X^T, A)|] < \infty$ , then Part (A) implies

$$\mathbb{E}_{\theta_0, \omega_0} [\varphi_{\theta_0}(Y^T, X^T)] = \mu_0.$$

This moment condition has the FHR property: (24) ensures that  $\varphi_\theta$  is robust to an unknown distribution of unobserved heterogeneity, whereas the  $T - 1$  conditions (25) endow  $\varphi_\theta$  with robustness to heterogeneous feedback of an unknown form.

We additionally state the following corollary, which mimics Corollary 2.1.1 and suggests a systematic recipe for constructing functions  $\varphi_\theta$ .

**Corollary 3.1.1.** (ALTERNATIVE CHARACTERIZATION OF MOMENT RESTRICTIONS FOR AVERAGE EFFECTS) *Let  $\omega^* \in \Omega$ . Suppose that  $h_\theta$  is absolutely integrable under  $(\theta, \omega^*)$ . An absolutely-integrable  $\varphi_\theta$  satisfies (24)-(25) if and only if there exist absolutely-integrable  $\zeta_{\theta, t}$ , for  $t = 1, \dots, T - 1$ , such that:*

$$\mathbb{E} [\varphi_\theta(Y^T, X^T) - h_\theta(Y^T, X^T, A) | Y^{T-1}, X^T, A] = \sum_{t=1}^{T-1} \zeta_{\theta, t}(Y^t, X^t, A), \quad (26)$$

with  $\mathbb{E} [\zeta_{\theta,t}(Y^t, X^t, A) | Y^{t-1}, X^t, A] = 0$  for  $t = 1, \dots, T - 1$ .

### 3.2 FHR moments for the average structural hazard

The conditional hazard function

$$\lambda(y_t | y_{t-1}, x_t, a) = \lambda_\alpha(y_t) e^{x'_t \beta + \gamma y_{t-1} + a}$$

gives the instantaneous exit rate of a unit, at duration  $y_t$ , with lagged duration  $y_{t-1}$ , beginning-of-spell covariate  $x_t$ , and latent attribute  $a$ . Unfortunately, although easily identified, the observed hazard function

$$\lambda(y_t | y_{t-1}, x_t) = \lambda_\alpha(y_t) e^{x'_t \beta + \gamma y_{t-1}} \mathbb{E}[e^A | Y_t > y_t, Y_{t-1} = y_{t-1}, X_t = x_t]$$

suffers from spurious duration dependence (see Lancaster, 1990; Heckman, 1991): units with higher values of  $A$  will exit earlier, implying that the mean  $\mathbb{E}[e^A | Y_t > y_t, Y_{t-1} = y_{t-1}, X_t = x_t]$  declines with  $y_t$ . In contrast, the average structural hazard (ASH),

$$\bar{\lambda}(y_t | y_{t-1}, x_t) = \lambda_\alpha(y_t) e^{x'_t \beta + \gamma y_{t-1}} \mathbb{E}[e^A], \quad (27)$$

which equals the (expected) hazard function for a randomly sampled unit when externally assigned lagged duration  $y_{t-1}$  and covariate  $x_t$ , does not suffer from heterogeneity bias.

Since  $\bar{P}_\theta | Y_0, X_1, A \sim \text{Gamma}(T, e^A)$  we have

$$\mathbb{E}\left[\bar{P}_\theta^\delta | Y_0, X_1, A\right] = e^{-\delta A} \frac{\Gamma(T + \delta)}{\Gamma(T)} \quad (28)$$

for  $\delta > -T$ . Using (28) with  $\delta = -1$  shows that  $\mathbb{E}[e^A] = \mathbb{E}\left[\frac{T-1}{\bar{P}_\theta}\right]$ , so

$$\bar{\lambda}(y_t | y_{t-1}, x_t) = \lambda_\alpha(y_t) e^{x'_t \beta + \gamma y_{t-1}} \mathbb{E}\left[\frac{T-1}{\bar{P}_\theta}\right]. \quad (29)$$

We have thus found one FHR moment function for the ASH. Now, for any other FHR moment function for the ASH,  $\varphi_\theta$ , we have that

$$\phi_\theta(Y^T, X^T) = \varphi_\theta(Y^T, X^T) - \lambda_\alpha(y_t) e^{x'_t \beta + \gamma y_{t-1}} \frac{T-1}{\bar{P}_\theta}$$

is a FHR moment function for  $\theta$  (where here  $y_t, y_{t-1}, x_t$  are fixed values at which the ASH is evaluated). Since all such moment functions  $\phi_\theta$  are fully characterized in the MPH case by

Lemma 2.2, this characterizes all FHR moment functions for the ASH. It turns out, as we will shortly see, that estimation based upon the sample analog of (29) is efficient when  $\theta$  is replaced by an efficient estimate.

## 4 Efficient moment restrictions

Theorems 2.1 and 3.1 characterize the complete set of moment conditions available for estimating  $\theta$  and  $\mu(\theta, \omega)$ . When this set is nonempty, estimation at parametric rates may be (and often is) feasible. Moreover, many valid moment restrictions may be available (see Lemma 2.2 for the case of the MPH model). In such settings semiparametric efficiency bound theory provides a useful tool for selecting specific moments for estimation purposes.

In this section we derive the form of the efficient scores for  $\theta$  and  $\mu(\theta, \omega)$  and, consequently, their corresponding semiparametric efficiency bounds. As we shall see, the characterizations presented earlier facilitate the derivation of these bounds. We illustrate the application of our results via a detailed analysis of efficiency bounds for the MPH model.<sup>8</sup>

### 4.1 Efficiency for common parameters and average effects

We begin with a brief review of basic concepts in semiparametric efficiency theory; see, for example, Van der Vaart (2000, Chap. 25) and Newey (1990). Let  $\theta_0$ ,  $g_0$ ,  $\pi_0$ , and  $\nu_0$  denote, respectively, the common parameter, feedback process, heterogeneity distribution, and initial condition prevailing in the sampled population. Let  $\omega = (g, \pi, \nu) \in \Omega$ , with true value  $\omega_0 = (g_0, \pi_0, \nu_0)$ , denote the nonparametric components of the model.<sup>9</sup> A regular parametric submodel is defined by a likelihood function for a single random draw,  $\ell(\theta, \omega_\eta | y^T, x^T)$ , where  $\omega_{\eta_0} = \omega_0$  for some scalar  $\eta_0$ . The likelihood satisfies mean-square differentiability of its square root with respect to  $(\theta, \eta)$ , with its information matrix nonsingular. The semiparametric variance bound is the supremum of the Cramer Rao bounds for  $\theta$  over all such regular parametric submodels.

Let  $S^\theta$  denote the score for  $\theta$ , for a submodel evaluated at  $\theta = \theta_0$  and  $\eta = \eta_0$ :

$$S^\theta(Y^T, X^T) = \frac{\partial \ln \ell(\theta_0, \omega_0 | Y^T, X^T)}{\partial \theta},$$

---

<sup>8</sup>Hahn (1994) derived the information bound for  $\theta$  in the single-spell case studied by Elbers and Ridder (1982) and Heckman and Singer (1984). He also derived the bound for the multi-spell case under strict exogeneity (see also Chamberlain, 1985). Relative to prior work, our analysis covers the case with feedback and, additionally, considers average effects.

<sup>9</sup>To characterize the efficiency bound for  $\theta$ , by ancillarity it is sufficient to consider the conditional density given  $(y_0, x_1)$  (see, e.g., Hahn, 1994).

where we leave the dependence of  $S^\theta$  on  $(\theta_0, \omega_0)$  implicit. Likewise, let  $S^\eta$  denote the score for  $\eta$ . The nonparametric tangent set  $\mathcal{T}_{\theta_0, \omega_0, K}$  is the mean-square closure of the  $K \times 1$  linear combinations of scores  $S^\eta$  across all regular parametric submodels (where  $K$  is the dimension of  $\theta$ ). Let  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$  denote the orthocomplement of  $\mathcal{T}_{\theta_0, \omega_0, K}$  in the Hilbert space of square-integrable mean-zero functions with inner product  $\langle s_1, s_2 \rangle = \mathbb{E}_{\theta_0, \omega_0} [s_1(Y^T, X^T)' s_2(Y^T, X^T)]$ . By definition this set consists of all  $K \times 1$  elements  $\phi_\theta$  such that  $\langle \phi, s \rangle = 0$  for all  $s \in \mathcal{T}_{\theta_0, \omega_0, K}$ .

The next theorem provides a characterization of the orthocomplement of the tangent set,  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$ , which is key to the analysis of efficiency in our context.

**Theorem 4.1. (ORTHOCOMPLEMENT OF TANGENT SET)**  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$  is the linear span of square-integrable,  $K$ -dimensional moment functions  $\phi_\theta$  that satisfy (9) and (10).

An implication of Theorem 4.1 is that, if  $\phi_{\theta_0} \in \mathcal{T}_{\theta_0, \omega_0, K}^\perp$ , then it is also an element of the orthogonal complement of the nuisance tangent set associated with *any other* data generating process  $(\theta_0, \omega_*)$  with  $\omega_* \neq \omega_0$  (i.e., we also have  $\phi_{\theta_0} \in \mathcal{T}_{\theta_0, \omega_*, K}^\perp$ ). This follows from the fact that  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$  consists of the set of FHR moments characterized in Theorem 2.1 earlier; a set which does not vary with  $\omega_0$ . Indeed, it is precisely this feature of the model which makes feedback and heterogeneity robust estimation possible. Knowledge, whether *a priori* or up to sampling error, of the form of the feedback process and/or heterogeneity distribution is not required for consistent estimation. This is a crucial feature of the class of models we study in this paper.

One subtlety is that, although the elements of  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$  do not depend on the precise instance of  $\omega_0$ , the definition of the reference Hilbert space does depend on it. Let  $\phi_{\theta_0}$  be a valid FHR moment that is absolutely integrable under  $(\theta_0, \omega^*)$ . Then, while  $\phi_{\theta_0}(Y^T, X^T)$  has zero mean under both  $(\theta_0, \omega_0)$  and  $(\theta_0, \omega^*)$ , in general its variance differs under  $\omega_0$  and  $\omega^*$ . Consequently, the ability to precisely estimate  $\theta_0$  using a particular  $\phi_\theta$  generally varies with the population feedback process and heterogeneity distribution, although the validity of  $\phi_\theta$  as a moment function does not. This connects to the discussion of locally efficient estimation below.

To understand Theorem 4.1 it is helpful to consider the implications of restricting  $\omega_0$  such that it belongs to a parametric family (indexed by, say,  $\eta$ ). This is the approach taken in, for example, parametric random-effects analysis (e.g., Chamberlain, 1980, 1985). In that setting the residualized score,  $\tilde{S}^\theta = S^\theta - \mathbb{E}[S^\theta S^{\eta'}] \times \mathbb{E}[S^\eta S^{\eta'}]^{-1} S^\eta$ , will generally vary with  $\eta$ : consistent estimation of  $\theta$  requires knowledge of  $\eta$  (up to sampling error), and it requires correct specification of the parametric models of feedback and heterogeneity. This is not required in our approach; indeed (elements of)  $\omega$  may even be unidentified, while  $\theta$  remains  $\sqrt{N}$ -estimable.

Theorem 4.1 is reminiscent of the situation which arises in average treatment effect (ATE) estimation under unconfoundedness with a *known* propensity score (Robins et al., 1994; Hahn, 1994). In that setting the set of consistent estimating equations for the ATE does not depend on the form of the conditional distribution of the potential outcomes given covariates. Theorem 4.1 extends prior work for the case of strictly exogenous regressors showing that  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$  is characterized by moment functions that have zero means conditional on *all* covariates, initial conditions, and heterogeneity (see, e.g., Hahn, 1994 for the MPH model, and Dano, 2023 for dynamic logit models).

Of course, in many semiparametric estimation problems, consistent estimation of (features of) the nonparametric model component *is* a requirement for consistent estimation of  $\theta$ . Examples include the binary choice model with random utility components drawn from an unknown distribution independent of the regressors (see Newey, 1990) and ATE estimation under unconfoundedness with an *unknown* propensity score.

**Efficient score for  $\theta$ .** Under regularity conditions,<sup>10</sup> the semiparametric variance bound for  $\theta_0$  is equal to the inverse of the variance of the efficient score

$$\phi_{\theta_0, \omega_0}^{\text{eff}}(Y^T, X^T) = \Pi(S^\theta(Y^T, X^T) | \mathcal{T}_{\theta_0, \omega_0, K}^\perp), \quad (30)$$

where  $\Pi(\cdot | \mathcal{T}_{\theta_0, \omega_0, K}^\perp)$  denotes the orthogonal projection onto  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$ . Note that the projection is well defined since  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$  is closed and linear.<sup>11</sup> As a result, the efficient moment restriction for  $\theta_0$  is

$$\mathbb{E}_{\theta_0, \omega_0} [\phi_{\theta_0, \omega_0}^{\text{eff}}(Y^T, X^T)] = 0. \quad (31)$$

The efficient score  $\phi_{\theta_0, \omega_0}^{\text{eff}}$  in (30) is defined through a projection onto the orthocomplement  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$ , which we have fully characterized in Theorem 4.1. Below we show, via examples, how to use this observation to calculate – whether analytically or numerically – efficient scores in models with unknown heterogeneity and feedback.

**Efficient score for  $\mu$ .** We now turn to an analysis of efficiency for average effects. Suppose there exists a moment function  $\varphi_\theta$  that identifies an  $L \times 1$  average effect of interest  $\mu(\theta_0, \omega_0)$

---

<sup>10</sup>Namely that  $\ell(\theta, \omega | y^T, x^T)$  is smooth in a neighborhood of  $(\theta_0, \omega_0)$ , and that the information matrix is nonsingular (see for example Theorem 3.2 in Newey, 1990).

<sup>11</sup>An equivalent and more frequent formulation of the efficient score is  $\phi_{\theta_0}^{\text{eff}} = S^\theta - \Pi(S^\theta | \mathcal{T}_{\theta_0, \omega_0, K}^\perp)$  (e.g., Hahn, 1994), which is interpreted as the residual from the population regression of  $S^\theta$  on the nuisance tangent set. The equivalent formulation based on  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$  is convenient to work with in our setting; see van der Laan and Robins (2003, pp. 57-58).

given by<sup>12</sup>

$$\mu(\theta_0, \omega_0) = \mathbb{E}_{\theta_0, \omega_0} [h_{\theta_0}(Y^T, X^T, A)] = \mathbb{E}_{\theta_0, \omega_0} [\varphi_{\theta_0}(Y^T, X^T)], \quad (32)$$

and suppose that  $\theta_0$  is identified from (31). Let

$$\varphi_{\theta_0, \omega_0}^{\text{eff}}(Y^T, X^T) = \varphi_{\theta_0}(Y^T, X^T) - \Pi(\varphi_{\theta_0}(Y^T, X^T) | \mathcal{T}_{\theta_0, \omega_0, L}^{\perp}),$$

where the orthocomplement of the tangent set,  $\mathcal{T}_{\theta_0, \omega_0, L}^{\perp}$ , is given by Theorem 4.1, except for the fact that the relevant dimension is  $L$  instead of  $K$ .

Theorem 1 in Brown and Newey (1998) shows that the joint efficient moment restrictions for  $\theta_0$  and  $\mu_0 = \mu(\theta_0, \omega_0)$  are given by (31) and

$$\mathbb{E}_{\theta_0, \omega_0} [\varphi_{\theta_0, \omega_0}^{\text{eff}}(Y^T, X^T) - \mu_0] = 0. \quad (33)$$

The semiparametric variance bound for  $\mu_0$  equals the lower  $L \times L$  block of the asymptotic variance of the joint GMM estimator  $(\hat{\theta}, \hat{\mu})$  based on (31) and (33).

The construction of  $\varphi_{\theta_0, \omega_0}^{\text{eff}}$  relies on a function  $\varphi_{\theta_0}$  that satisfies (32). In some models one can find such a function (a Riesz representer), which does not depend on the nonparametric component  $\omega_0$ . This can be done by exploiting Theorem 3.1. See, for example, the discussion of the average structural hazard function in the MPH earlier. Moreover,  $\varphi_{\theta_0, \omega_0}^{\text{eff}}$  is not affected by the particular choice of  $\varphi_{\theta_0}$ .<sup>13</sup> Of course not all average effects will have non-zero efficiency bounds, but when a Riesz representer for an effect of interest is available, the bound can be calculated using extant results about expectations (Brown and Newey, 1998).

## 4.2 Moment restrictions based on working models

Although the characterization of feasible moment conditions provided by Theorem 4.1 is invariant to the specific instance of  $\omega_0$  indexing the sampled population, the projection (30) generally *does* vary with  $\omega_0$ . Consequently, although knowledge of  $\omega_0$  is not required for consistent estimation, it is generally valuable for improving asymptotic precision. Moreover, constructing an estimator which attains the semiparametric efficiency bound for all possible feedback processes,  $g_0$ , heterogeneity distributions,  $\pi_0$ , and initial conditions,  $\nu_0$  (i.e., for all  $\omega_0 \in \Omega$ ) generally requires nonparametrically estimating features of these model components.

---

<sup>12</sup>Standard implicit smoothness conditions are required, namely that, for a regular parametric submodel,  $\sup_{(\theta, \eta) \in \mathcal{N}} \mathbb{E}_{\theta, \omega_\eta} [\|\varphi_\theta(Y^T, X^T)\|^2] < \infty$  in a neighborhood  $\mathcal{N}$  of  $(\theta_0, \eta_0)$ , such that a generalized information equality holds (see Brown and Newey, 1998).

<sup>13</sup>This follows from noting that  $\varphi_{\theta_0}(Y^T, X^T) - \Pi(\varphi_{\theta_0}(Y^T, X^T) | \mathcal{T}_{\theta_0, \omega_0, L}^{\perp}) = \mu_0 + \Pi(\varphi_{\theta_0}(Y^T, X^T) | \mathcal{T}_{\theta_0, \omega_0, L}^{\perp})$ , and that, if  $\varphi_{\theta_0}^1$  and  $\varphi_{\theta_0}^2$  both satisfy (32), then  $\Pi(\varphi_{\theta_0}^1(Y^T, X^T) | \mathcal{T}_{\theta_0, \omega_0, L}^{\perp}) = \Pi(\varphi_{\theta_0}^2(Y^T, X^T) | \mathcal{T}_{\theta_0, \omega_0, L}^{\perp})$  as we show in Supplemental Appendix Lemma B.1.

This may be practically difficult, or even impossible, in short panels as considered here.

An alternative approach involves constructing *locally efficient* estimators (e.g., Newey, 1990; Graham et al., 2012). Let  $\tilde{\omega} = (\tilde{g}, \tilde{\pi}, \tilde{\nu}) \in \Omega$  denote candidate “working models” for the feedback process, heterogeneity distribution, and initial condition. We show how to construct method-of-moments estimators that (i) attain the bound for  $\theta_0$  (or  $\mu_0$ ) when these working models “happen to characterize the sampled population” (i.e.,  $\tilde{\omega} = \omega_0$ , but this is not part of the prior restriction) and (ii) remain  $\sqrt{N}$ -consistent irrespective of the true  $\omega_0$  characterizing the sampled population (i.e., when  $\tilde{\omega} \neq \omega_0$ ). A key property in our setting is that consistency holds for arbitrary  $\tilde{\omega}$  (i.e., our working models may be misspecified), only subject to regularity conditions.

Given working models  $\tilde{\omega}$ , let

$$\tilde{S}^\theta(Y^T, X^T) = \frac{\partial \ln \ell(\theta_0, \tilde{\omega} | Y^T, X^T)}{\partial \theta}$$

denote the score for  $\theta_0$ . Next define the counterpart, *under the working models*, to the efficient score  $\phi_{\theta_0}^{\text{eff}}$  for  $\theta_0$  as

$$\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y^T, X^T) = \tilde{\Pi} \left( \tilde{S}^\theta(Y^T, X^T) | \mathcal{T}_{\theta_0, \tilde{\omega}, K}^\perp \right),$$

where  $\tilde{\Pi}$  denotes the projection operator under the working models, that is,

$$\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}} = \arg \min_{\phi \in \mathcal{T}_{\theta_0, \tilde{\omega}, K}^\perp} \mathbb{E}_{\theta_0, \tilde{\omega}} \left[ \left( \tilde{S}^\theta(Y^T, X^T) - \phi(Y^T, X^T) \right)^2 \right]. \quad (34)$$

We next proceed similarly for  $\mu(\theta, \omega)$ : the counterpart to the efficient score  $\varphi_{\theta_0, \omega_0}^{\text{eff}}$  is

$$\tilde{\varphi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y^T, X^T) = \varphi_{\theta_0}(Y^T, X^T) - \tilde{\Pi}(\varphi_{\theta_0}(Y^T, X^T) | \mathcal{T}_{\theta_0, \tilde{\omega}, L}^\perp). \quad (35)$$

Finally, consider the following moment restrictions for  $\theta_0$  and  $\mu_0 = \mu(\theta_0, \omega_0)$ :

$$\mathbb{E}_{\theta_0, \omega_0} \left[ \tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y^T, X^T) \right] = 0, \quad (36)$$

$$\mathbb{E}_{\theta_0, \omega_0} \left[ \tilde{\varphi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y^T, X^T) - \mu_0 \right] = 0, \quad (37)$$

where we note that the expectations are taken under the true DGP  $(\theta_0, \omega_0)$ .

We can now state the following result.

**Theorem 4.2.** *For any working models  $\tilde{\omega} \in \Omega$  such that  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  and  $\tilde{\varphi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  are absolutely integrable under DGP  $(\theta_0, \omega_0)$ , the moment restrictions (36) and (37) hold. Moreover, if  $\tilde{\omega} = \omega_0$ , then (36) and (37) coincide with the efficient moment restrictions for  $\theta_0$  and  $\mu_0$ .*

Theorem 4.2 articulates a *locally* efficient approach to estimation. The moment functions  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  and  $\tilde{\varphi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  have the FHR property, irrespective of whether the working models used to derive them actually characterize the sampled population. However, if  $\tilde{\omega} = \omega_0$  “happens to hold” in the sampled population, then (36) and (37) coincide with the efficient moment restrictions for  $\theta_0$  and  $\mu_0$  (when  $\tilde{\omega} = \omega_0$  is *not* part of the prior restriction used to calculate the efficiency bound).

The functions  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  and  $\tilde{\varphi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  are FHR because calculation (34) returns an element in the orthogonal complement of the nuisance tangent set by construction. Calculation (34) provides a principled way to select a particular FHR moment, one that is optimal if the working model happens to hold in the sampled population. Heuristically, the method-of-moments estimator based upon (36) and (37) will be more precise when the working models are “approximately true”, but this – to repeat – is not required for consistency.

In practice  $\tilde{\omega}$  may be a fixed set of working models chosen by the researcher. Alternatively the researcher may posit that these models belong to parametric families  $\omega_\eta$  indexed by an unknown finite-dimensional (not necessarily scalar) parameter  $\eta$ . These models may be misspecified, in the sense that there may not exist any  $\eta_0$  such that  $\omega_{\eta_0} = \omega_0$ . Nevertheless the moments (36) and (37) will be valid for any  $\eta$ . There are different strategies for picking a particular  $\eta$ . First, the researcher may choose a particular (non-stochastic)  $\eta$  via introspection. Second, she might maximize the likelihood under the working models with respect to  $\theta$  and  $\eta$ . While the resulting estimate of  $\theta$  will generally be inconsistent, the corresponding estimate of  $\eta$  can be used to define the working models  $\tilde{\omega}$  under which  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  and  $\tilde{\varphi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  are calculated.

A third approach is to select  $\eta$  by maximizing an empirical counterpart to the information for  $\theta_0$ , as in Lindsay (1985). Let  $\hat{\eta}$  be such an estimator of  $\eta$ , and let  $\eta^*$  be its large- $N$  probability limit. If one defines  $\tilde{\omega} = \omega_{\eta^*}$ , then Theorem 4.2 implies that (36)-(37) are satisfied at true parameter values  $\theta_0$  and  $\mu_0$  (under absolute integrability of the functions). This suggests that the GMM estimators  $\hat{\theta}$  and  $\hat{\mu}$  based on (36) and (37) that uses  $\omega_{\hat{\eta}}$  in lieu of  $\tilde{\omega}$  is consistent and asymptotically normal under standard conditions. We leave details about efficient estimation, using working models of increasing dimensions (i.e., “sieves”), to future work.

Lastly, it is important to stress that our approach based on working models is fundamentally different from (parametric) random-effects maximum likelihood estimation. Indeed, plugging in misspecified parametric models  $\omega_\eta$  into the likelihood function (5), and maximizing that likelihood with respect to  $\theta$  and  $\eta$ , generally results in an inconsistent estimator of  $\theta_0$ . In contrast, the moment restrictions (36) and (37) remain valid even when the working models are (globally) misspecified.

## 5 Efficiency in the multi-spell MPH with unrestricted feedback

In this section we specialize the general results presented above in order to analyze semiparametric efficiency in the MPH model. We focus on the  $T = 2$  special case in what follows.

### 5.1 Efficiency bounds analysis

**Efficient score for  $\theta$ .** Using Lemma 2.2 and Theorem 4.1, we show in Supplemental Appendix D that, for  $T = 2$ , the efficient score for  $\theta$  in the MPH model with feedback equals

$$\phi_{\theta_0}^{\text{eff}}(Y^2, X^2) = \mathbb{E} \left[ S^\theta(Y^2, X^2) \mid Y_0, \tilde{P}_{\theta_0,1}, \bar{P}_{\theta_0}, X_1 \right] - \mathbb{E} \left[ S^\theta(Y^2, X^2) \mid Y_0, \bar{P}_{\theta_0}, X_1 \right]. \quad (38)$$

With some additional algebra we show that the efficient score for the  $\beta$  subvector is

$$\begin{aligned} \phi_{\theta_0}^{\text{eff},\beta}(Y^2, X^2) &= -X_1 \left( \tilde{P}_{\theta_0,1} - \frac{1}{2} \right) \bar{P}_{\theta_0} \mathbb{E} [e^A \mid Y_0, \bar{P}_{\theta_0}, X_1] \\ &\quad + \mathbb{E} \left[ X_2 \left( 1 - (1 - \tilde{P}_{\theta_0,1}) \bar{P}_{\theta_0} e^A \right) \mid Y_0, \tilde{P}_{\theta_0,1}, \bar{P}_{\theta_0}, X_1 \right] \\ &\quad - \mathbb{E} \left[ X_2 \left( 1 - (1 - \tilde{P}_{\theta_0,1}) \bar{P}_{\theta_0} e^A \right) \mid Y_0, \bar{P}_{\theta_0}, X_1 \right]. \end{aligned} \quad (39)$$

The first term in (39) does not involve the feedback process and resembles the efficient score for  $\beta$  under strict exogeneity originally derived by Hahn (1994):

$$\phi_{\theta_0}^{\text{eff,SE},\beta}(Y^2, X^2) = (X_2 - X_1) \left( \tilde{P}_{\theta_0,1} - \frac{1}{2} \right) \bar{P}_{\theta_0} \mathbb{E} [e^A \mid Y_0, \bar{P}_{\theta_0}, X_1, X_2]. \quad (40)$$

The second and third terms of (39), in contrast, are specific to the feedback case, involving averages over the second-period covariate  $X_2$ . More generally, the efficient score for  $\theta$  under strict exogeneity equals:

$$\phi_{\theta_0}^{\text{eff,SE}}(Y^2, X^2) = \mathbb{E} \left[ S^\theta(Y^2, X^2) \mid Y_0, \tilde{P}_{\theta_0,1}, \bar{P}_{\theta_0}, X_1, X_2 \right] - \mathbb{E} \left[ S^\theta(Y^2, X^2) \mid Y_0, \bar{P}_{\theta_0}, X_1, X_2 \right]. \quad (41)$$

In the presence of feedback and latent heterogeneity,  $X_2$  is an endogenous variable and cannot be conditioned on.

**Efficient estimation of the ASH.** An interesting average effect in the context of the MPH is the average structural hazard (ASH) defined in (22). The latter is identified by the FHR moment function in (29), namely  $\varphi_{\theta_0}(Y^T, X^T) = \lambda_{\alpha_0}(y_t) e^{x'_t \beta_0 + \gamma_0 y_{t-1} \frac{T-1}{\bar{P}_{\theta_0}}}$ . Applying Lemma D.3

in the Supplemental Appendix, which characterizes projections onto the orthocomplement of the tangent set in the MPH model, one can readily show that  $\Pi(\varphi_{\theta_0}(Y^T, X^T) | \mathcal{T}_{\theta_0, \omega_0, L}^\perp) = 0$ . It then follows from the discussion in Section 4.1 that the efficient moment function for the ASH is

$$\varphi_{\theta_0, \omega_0}^{\text{eff}}(Y^T, X^T) = \varphi_{\theta_0}(Y^T, X^T) = \lambda_{\alpha_0}(y_t) e^{x'_t \beta_0 + \gamma_0 y_{t-1}} \frac{T-1}{\bar{P}_{\theta_0}}.$$

In turn, a semiparametrically efficient estimator of the ASH is

$$\widehat{\lambda}(y_t | x_t, y_{t-1}; \widehat{\theta}) = \lambda_{\widehat{\alpha}}(y_t) e^{x'_t \widehat{\beta} + \widehat{\gamma} y_{t-1}} \frac{1}{N} \sum_{i=1}^N \frac{T-1}{\bar{P}_{i\widehat{\theta}}},$$

where  $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}', \widehat{\gamma})'$  is semiparametrically efficient for  $\theta$ .<sup>14</sup>

## 5.2 Numerical illustrations

In this subsection we summarize our findings from two numerical experiments designed to (i) illustrate the efficiency loss associated with accommodating feedback, and (ii) assess the performance of locally efficient estimators based on working models. Our context is the MPH model. In both experiments we impose a Weibull baseline hazard, set  $T = 2$ , and fix the common parameter at  $\theta_0 = (\alpha_0, \gamma_0, \beta_0) = (\frac{3}{4}, \frac{3}{4} \ln 2, -\frac{1}{10})$ . Our experiments are meant to numerically approximate the asymptotic precision of various methods of estimation, not to assess the accuracy of such approximations in small samples. Details on implementation can be found in Supplemental Appendix E.

The initial duration is drawn from an exponential distribution:  $Y_0 \sim \text{Exponential}(\frac{3}{2})$ , and the first-period covariate is a randomized binary treatment:  $X_1 \sim \text{Bernoulli}(\frac{1}{2})$ . The heterogeneity distribution equals  $V = e^A \sim \text{Gamma}(5, 5)$ , independent of  $Y_0, X_1$ . The second-period covariate,  $X_2$ , is a Bernoulli random variable with success probability  $p(Y_0, Y_1, X_1, V)$ , specified differently across the two experiments to reflect alternative assumptions about the DGP:

- In Experiment (A), we set  $p(Y_0, Y_1, X_1, V) = 1 - \exp(-(Y_0 + X_1)V)$ , which produces a DGP with strictly exogenous covariates but correlated unobserved heterogeneity.

---

<sup>14</sup>Note that this coincides with the method of moments estimator of Brown and Newey (1998) when the condition  $\varphi_{\theta_0}(Y^T, X^T) - \mu_0 \in \mathcal{T}_{\theta_0, \omega_0, L}$  holds, where  $\mu_0 = \mu(\theta_0, \omega_0)$  denotes the average effect of interest and  $\varphi_{\theta_0}$  is an identifying moment function. This condition is satisfied in the case of the ASH for the choice  $\varphi_{\theta_0}(Y^T, X^T) = \lambda_{\alpha_0}(y_t) e^{x'_t \beta_0 + \gamma_0 y_{t-1}} \frac{T-1}{\bar{P}_{\theta_0}}$ .

- In Experiment (B), we instead let  $p(Y_0, Y_1, X_1, V) = 1 - \exp(-(Y_0 + X_1 + Y_1)V)$ , which introduces feedback from past outcomes to future covariates, while continuing to include correlated heterogeneity.

One goal of our experiments is to assess the efficiency loss that arises when a researcher wishes to accommodate the possibility of heterogeneous feedback, but no such feedback is actually present in the sampled population. Put differently, this exercise gives us a sense of the benefit, in terms of asymptotic precision, of using the strict exogeneity assumption when it is valid (as is the case for the DGP in Experiment (A)).

We compare the asymptotic standard errors of two GMM estimators: the first uses the efficient score under strict exogeneity (41), while the second uses the efficient score under feedback (38).

The close connection between our FHR moment characterization (Theorem 2.1) and the relevant semiparametric efficiency bound theory (Theorem 4.1), raises interesting practical questions regarding estimation. While many possible FHR moments are available in the MPH setting, the precision with which they recover  $\theta$  varies with the population values of the feedback process and heterogeneity distribution.

The complex form of the efficient score (those for the baseline hazard parameter,  $\alpha$ , and the coefficient on the lagged duration,  $\gamma$  – both reported in Supplemental Appendix D – are particularly complicated), suggests that crafting a globally efficient estimator would be difficult. As a principled, yet practical, alternative, we instead explore the properties of a locally efficient estimator based upon simple working models for  $\tilde{\omega} = (\tilde{g}, \tilde{\pi})$  (a model for  $\nu$  is not needed in this case).

Our chosen working models are deliberately rudimentary, intended to illustrate how a researcher might build parsimonious working models while retaining favorable efficiency properties in more realistic settings. Specifically, in contrast to what prevails in the sampled population, the working model for the feedback process maintains that  $X_2 \sim \text{Bernoulli}(p)$ , for a constant  $p$ . Observe that the working model for the feedback process involves no feedback. For the heterogeneity distribution, we calculate the locally efficient score under a vague Gamma prior of  $\tilde{\pi}(v) = \frac{1}{v} \mathbb{1}\{v > 0\}$ , independent of initial conditions.

Putting all these pieces together yields the following locally efficient score for  $\beta$ :

$$\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}, \beta}(Y^2, X^2) = 2(\mathbb{E}_{\tilde{g}}[X_2] - X_1) \left( \tilde{P}_{\theta_0, 1} - \frac{1}{2} \right) = 2(p - X_1) \left( \tilde{P}_{\theta_0, 1} - \frac{1}{2} \right),$$

which is simpler than its optimal counterpart (39) yet, of course, still feedback and heterogeneity robust. Additional details, along with the full expression for the score  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y_0, \tilde{P}_{\theta_0, 1}, \bar{P}_{\theta_0}, X_1)$ , are provided in Supplemental Appendix E.

As a final point of comparison, we also report the limiting standard errors of a just-identified GMM estimator that employs the moment function:

$$\phi_{\theta_0}(Y^2, X^2) = \begin{pmatrix} 2 + \ln(1 - \tilde{P}_{\theta_0,1}) + \ln(\tilde{P}_{\theta_0,1}) \\ X_1 \left( \ln(1 - \tilde{P}_{\theta_0,1}) - \ln(\tilde{P}_{\theta_0,1}) \right) \\ Y_0 \left( \ln(1 - \tilde{P}_{\theta_0,1}) - \ln(\tilde{P}_{\theta_0,1}) \right) \end{pmatrix}. \quad (42)$$

The first entry in (42) is a mean-zero function that exploits the fact that  $\tilde{P}_{\theta_0,1} \sim U[0, 1]$  and leverages symmetry (it is also, coincidentally, a component of the efficient score for  $\alpha$  under the Weibull baseline hazard, derived and presented in Supplemental Appendix D). The second and third entries correspond to the log-transformed analogs of (19). While this set of moment conditions lacks an overt efficiency justification, it reflects a common strategy of choosing a small set of “simple” moments for estimation purposes. We include it primarily to benchmark the efficiency gains provided by the locally efficient approach based on working models.

Table 1 reports the asymptotic standard errors for each estimate of  $\theta_0$  in Experiment (A). We make several observations. First, comparing the first and second rows reveals that accommodating feedback – when it is, in fact, absent from the DGP – results in some efficiency loss for the slope coefficient  $\beta$ , with a 29% increase in standard error. Strict exogeneity is a strong assumption and imposing it, when it is valid to do so, improves asymptotic precision.

Although allowing for feedback degrades the precision with which we can learn  $\beta$ , this is not really the case for  $\alpha$  (the Weibull baseline hazard parameter) and  $\gamma$  (the lagged duration dependence parameter) in design (A). This finding is consistent with the structure of the efficient scores: those for  $\alpha$  and  $\gamma$  are very similar under both strict exogeneity and feedback.<sup>15</sup> By contrast, the efficient score for  $\beta$  differs markedly across the two settings.

A second observation, inspecting the third row of the table, is that the precision loss associated with using the locally efficient estimator based on the working models  $\tilde{\omega}$  is moderate. Recall that our working models do not characterize the sampled population in design (A). For  $\beta$  and  $\gamma$ , respectively, we observe a 28% and 26% increase in standard error when comparing rows 2 and 3. The efficiency losses are concentrated on the coefficients for the predetermined covariate and the lagged dependent variable, with only minimal deterioration for the parameter of the baseline hazard  $\alpha$ .

Finally, the approach based on working models leads to large improvements relative to using the “simple” moment functions (42). As seen in row 4, the standard errors associated with the simple GMM estimator are substantially larger for all three parameters. For exam-

---

<sup>15</sup>Compare (77) to (80) and (75) to (79) in Supplemental Appendix D.

ple, the standard error for  $\gamma$  is nearly 7 times higher than that obtained using the estimator based on the working models  $\tilde{\omega}$  (compare rows 3 and 4, column 3).

Table 1: Asymptotic standard errors relative to the semiparametric efficiency bound with strict exogeneity in Experiment (A)

	$\alpha$	$\beta$	$\gamma$
Eff.score SE	1.0	1.0	1.0
Eff.score FB	1.000	1.292	1.003
Locally Eff.score FB	1.035	1.660	1.264
Simple moments	2.711	2.954	8.767

Notes: SE denotes strict exogeneity, FB denotes feedback. The DGP satisfies strict exogeneity.

In Experiment (B), we repeat our analysis, but for a population where heterogeneous feedback is, in fact, present. Accordingly, Table 2 compares the asymptotic standard errors of the locally efficient estimator based on  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}, \beta}$  to that of the “simple” GMM estimator based on (42) under the second DGP described above (an estimate based on the efficient score under strict exogeneity would not be consistent in this design). As a benchmark, we use the globally semiparametrically efficient estimator based upon the true efficient score that uses (38). The fourth column of Table 2 also reports the corresponding asymptotic standard errors for the average structural hazard (ASH) when using the efficient moment function presented in the previous subsection. The key takeaways mirror those of Table 1. First, the efficiency loss from relying on locally efficient scores is modest, both for common parameters and for average effects. Second, and perhaps more importantly, locally efficient estimators significantly outperform the alternative of using a set of “simple” moments, reaffirming the practical advantages of approaches based on working models.

Table 2: Asymptotic standard errors relative to the semiparametric efficiency bound with feedback in Experiment (B)

	$\alpha$	$\beta$	$\gamma$	ASH
Eff.score FB	1.0	1.0	1.0	1.0
Locally Eff.score FB	1.030	1.188	1.317	1.012
Simple moments	3.372	3.223	9.965	2.675

Notes: FB denotes feedback. The DGP does not satisfy strict exogeneity. The ASH is evaluated at  $y_1 = y_0 = x_1 = 1$ .

## 6 FHR moment restrictions in other models

Our characterizations can be used to find FHR moment functions for many models. We have already analyzed the MPH model in detail. In this section we provide additional analytical examples and discuss how to obtain moment functions more generally. Given that the mathematical structure for model parameters and average effects is similar, in this section we focus our discussion on  $\theta$ .

### 6.1 Existence of moment functions

For certain models, it may be that the only solution to the system of equations in Theorem 2.1, or equivalently in Corollary 2.1.1, is the degenerate one,  $\phi_\theta = 0$ . We now provide two examples where only trivial moment functions exist. For simplicity we focus on the  $T = 2$  case.

**Example 3. (BINARY CHOICE MODEL)** Consider a binary choice logit model with continuous heterogeneity and sequentially exogenous covariates:

$$\Pr(Y_t = 1 \mid Y_{t-1} = y_{t-1}, X_t = x_t, A = a; \theta) = \frac{\exp(\gamma y_{t-1} + \beta' x_t + a)}{1 + \exp(\gamma y_{t-1} + \beta' x_t + a)}, \quad t = 1, 2.$$

Any valid moment function of  $\theta = (\gamma, \beta')'$  needs to satisfy (10), that is,

$$\sum_{y_2=0}^1 \phi_\theta(y_0, y_1, y_2, x_1, x_2) \frac{\exp(\gamma y_1 + \beta' x_2 + a)^{y_2}}{1 + \exp(\gamma y_1 + \beta' x_2 + a)} \text{ does not depend on } x_2.$$

Suppose that  $\beta \neq 0$ . The only functions  $\phi_\theta$  satisfying this restriction do not depend on  $y_2$  or  $x_2$ . Hence, by (9) we obtain

$$\sum_{y_1=0}^1 \phi_\theta(y_0, y_1, x_1) \frac{\exp(\gamma y_0 + \beta' x_1 + a)^{y_1}}{1 + \exp(\gamma y_0 + \beta' x_1 + a)} = 0,$$

from which we obtain that  $\phi_\theta = 0$ . This shows the absence of non-trivial moment restrictions for  $\theta$  in this model. Building on Chamberlain (2022, 2023), Bonhomme et al. (2023) study the failure of point-identification in binary choice models with sequentially exogenous covariates, and show how to compute identified sets on the parameters and average effects. For this reason, our examples in the next subsections will focus on continuous outcomes.<sup>16</sup>

---

<sup>16</sup>As Bonhomme et al. (2023) note, imposing assumptions on the feedback process, such as Markovianity, may lead to non-trivial moment restrictions in discrete choice and other models where an approach allowing for fully unrestricted feedback is uninformative. Extending our approach to accommodate such additional

**Example 4. (RANDOM COEFFICIENTS MODEL)** Consider the Gaussian linear random coefficients model

$$Y_t = \gamma Y_{t-1} + B' X_t + C + \varepsilon_t, \quad \varepsilon_t | Y^{t-1}, X^t, A \sim \mathcal{N}(0, \sigma^2), \quad (43)$$

where  $A = (B', C)'$  is multidimensional. Any moment function on  $\theta = (\gamma, \sigma^2)'$  needs to satisfy

$$\int \phi_\theta(y_0, y_1, y_2, x_1, x_2) \exp\left(-\frac{1}{2\sigma^2} (y_2 - \gamma y_1 - b' x_2 - c)^2\right) dy_2 \text{ does not depend on } x_2. \quad (44)$$

Note that, if (44) holds, then, for all  $b$  and  $x_2, \tilde{x}_2$ ,

$$\phi_\theta(y_0, y_1, y_2 + b' x_2, x_1, x_2) = \phi_\theta(y_0, y_1, y_2 + b' \tilde{x}_2, x_1, \tilde{x}_2).$$

This implies that  $\phi_\theta$  does not depend on  $y_2$  or  $x_2$ . Then (9) implies

$$\int \phi_\theta(y_0, y_1, x_1) \exp\left(-\frac{1}{2\sigma^2} (y_1 - \gamma y_0 - b' x_1 - c)^2\right) dy_1 = 0,$$

from which it follows that  $\phi_\theta = 0$ . This shows the only FHR moment function on  $\theta$  in this model is the null function. This negative result echoes an example in Chamberlain (2022). For this reason, our examples in the next subsections, which all involve scalar outcomes, will feature one-dimensional unobserved heterogeneity.

It is important to note that, even when there exist non-zero functions  $\phi_\theta$ ,  $\theta$  may fail to be identified. For example, in the MPH model the covariates may be collinear, in which case identification fails. This is of course not specific to our setting. As in any nonlinear GMM problem, identification needs to be verified on a case-by-case basis, and while rank conditions for local identification of  $\theta$  are available, verifying global identification may be difficult. Conversely, it may also be that the only  $\phi_\theta$  satisfying the conditions of Theorem 2.1 is  $\phi_\theta = 0$ , yet  $\theta$  is point-identified.<sup>17</sup> However, in that case an implication of our analysis in Section 4 is that such identification is necessarily irregular and the semiparametric efficiency bound for  $\theta$  is zero (Chamberlain, 1986). Lastly, even if point-identification fails identified sets may be informative, as shown in Lee (2020) and Bonhomme et al. (2023).

---

assumptions is an interesting topic for future work.

<sup>17</sup>As an example, let  $T = 2$  and let  $Y_t = \theta + AX_t + \varepsilon_t$  with  $X_t$  continuously distributed on  $\mathbb{R}$  and  $\varepsilon_t | Y^{t-1}, X^t, A \sim \mathcal{N}(0, 1)$ . Applying a similar logic to that in model (44), one can show that  $\phi_\theta = 0$  since the assumptions imply that  $P(X_2 = 0) = P(X_1 = 0) = 0$ . However, this is a case where the parameter of interest  $\theta$  is *identified at 0* since  $\lim_{x \rightarrow 0} \mathbb{E}[Y_1 | X_1 = x] = \theta$  (Graham and Powell, 2012).

## 6.2 Obtaining new moment conditions

We now illustrate how researchers can apply the two-step procedure underlying Corollary 2.1.1 to derive new moment conditions. In the first step, we construct a function  $\psi_\theta = \sum_{t=1}^{T-1} \psi_{\theta,t}$ , for  $\psi_{\theta,t}$  such that

$$\mathbb{E} [\psi_{\theta,t}(Y^t, X^t, A) | Y^{t-1}, X^t, A] = 0, \quad t = 1, \dots, T-1. \quad (45)$$

In the second step, in the final time period, we invert the linear integral equation

$$\int \phi_\theta(y^T, x^T) f_\theta(y_T | y_{T-1}, x_T, a) dy_T = \psi_\theta(y^{T-1}, x^{T-1}, a), \quad (46)$$

to get a FHR moment function  $\phi_\theta(y^T, x^T)$ . Naturally, this last task requires the function  $\psi_\theta(y^{T-1}, x^{T-1}, a)$  to lie in the range of the integral operator induced by the parametric model. In many examples of interest, equation (46) can actually be inverted in closed form, yielding explicit expressions for functions  $\phi_\theta$ . As an initial example, consider the Poisson regression model introduced earlier.

**Example 2** (Continued). (MOMENTS FOR THE POISSON MODEL) Recall, for  $t = 1, \dots, T$ ,

$$Y_t | Y^{t-1}, X^t, A \sim \text{Poisson}(\exp(\gamma Y_{t-1} + X'_t \beta + A)),$$

with both feedback and heterogeneity unrestricted. For simplicity, consider the case  $T = 2$  and denote  $Z_t = (X'_t, Y_{t-1})'$  and  $\theta = (\beta', \gamma)'$ . Following the logic laid out above, we start out by picking a moment function  $\psi_\theta$  such that  $\mathbb{E}[\psi_\theta(Y_0, Y_1, X_1, A) | Y_0, X_1, A] = 0$ . An example of such a function is provided by the score for period  $t = 1$  in the likelihood which conditions on  $A$ ,  $\psi_\theta(y_0, y_1, x_1, a) = z_1(y_1 - \exp(z'_1 \theta + a))$ .

Next, the challenge is to solve for  $\phi_\theta$  in (46); here corresponding to finding a solution to

$$\sum_{y_2=0}^{\infty} \phi_\theta(y_0, y_1, y_2, x_1, x_2) \exp(-\exp(z'_2 \theta + a)) \frac{\exp(z'_2 \theta + a)^{y_2}}{y_2!} = \psi_\theta(y_0, y_1, x_1, a).$$

After multiplying by  $\exp(\exp(z'_2 \theta + a))$  and letting  $v = \exp(a)$ , this is equivalent to

$$\sum_{y_2=0}^{\infty} \phi_\theta(y_0, y_1, y_2, x_1, x_2) \exp(y_2 z'_2 \theta) \frac{v^{y_2}}{y_2!} = \psi_\theta(y_0, y_1, x_1, \ln v) e^{v \exp(z'_2 \theta)}.$$

This formulation reveals that, for a solution to exist, we require that  $v \mapsto \psi_\theta(y_0, y_1, x_1, \ln v) e^{v \exp(z'_2 \theta)}$  admits a Taylor series at  $v = 0$ , with coefficients given by

$\phi_\theta(y_0, y_1, y_2, x_1, x_2) \exp(y_2 z'_2 \theta)$ . Appealing to the uniqueness of the Taylor series, we then infer that:

$$\phi_\theta(y_0, y_1, y_2, x_1, x_2) = \frac{\partial^{y_2}}{\partial v^{y_2}} \Big|_{v=0} \left[ \psi_\theta(y_0, y_1, x_1, \ln v) e^{v \exp(z'_2 \theta)} \right] \exp(-y_2 z'_2 \theta). \quad (47)$$

By Corollary 2.1.1, all FHR moment functions for  $\theta$  take the form (47), for some appropriate mean-zero function  $\psi_\theta$ . For instance, in the case where  $\psi_\theta$  is the score in period  $t = 1$ , we obtain

$$\phi_\theta(y_0, y_1, y_2, x_1, x_2) = z_1 \left( y_1 - y_2 e^{(z_1 - z_2)' \theta} \right),$$

which is proportional to the moment function of Chamberlain (1992) and Wooldridge (1997). However, using (47) provides many additional valid moment restrictions on  $\theta$ . For example, by the second-moment properties of the Poisson distribution,

$$\psi_\theta(y_0, y_1, x_1, a) = \left[ y_1 (y_1 - 1) - \exp(z'_1 \theta + a)^2 \right] \cdot m(z_1) \quad (48)$$

is analytic in  $v = \exp(a)$  and also satisfies (45), from which we get the moment

$$\phi_\theta(y_0, y_1, y_2, x_1, x_2) = \left[ y_1 (y_1 - 1) - \frac{y_2 (y_2 - 1) \exp(z'_1 \theta)^2}{\exp(z'_2 \theta)^2} \right] \cdot m(z_1). \quad (49)$$

Additional FHR moment functions, based upon higher order moments of the Poisson distribution, are straightforward to construct.<sup>18</sup>

**Example 5. (MIXED INTERACTIVE HAZARD (MIH) MODEL)** As an example of a new model, one for which no valid FHR moment conditions are known, consider the following Mixed Interactive Hazards (MIH) model. The MIH model relaxes the proportionality assumption of the MPH model; the conditional density of the  $t$ -th spell equals:

$$f_\theta(y_t | y_{t-1}, x_t, a) = \exp(\gamma y_{t-1} + x'_t \beta + a + (x'_t \delta) \cdot a) \lambda_\alpha(y_t) \\ \times \exp(-\exp(\gamma y_1 + x'_1 \beta + a + (x'_1 \delta) \cdot a) \Lambda_\alpha(y_t)), \quad (50)$$

where  $\theta = (\alpha', \gamma, \beta', \delta')'$ . Note that (50) simplifies to (3) when  $\delta = 0$ . However,  $\delta \neq 0$  allows for more general interaction effects between the covariate and unobserved heterogeneity. The MIH model is an example of a “generalized hazards” model (Bonev, 2020). Let  $\psi_\theta(y_0, y_1, x_1, a)$

---

<sup>18</sup>Moreover, by Theorem 4.1, the functions  $\phi_\theta$  in (47) span the orthocomplement of the tangent set. This suggests that one could compute the efficient score for  $\theta$  by projecting the  $\theta$ -score on that set of functions, although we leave the derivation of the precise form of the efficient score to future work.

be a function satisfying (45). The integral equation (46), after employing the change of variable  $y_2 \mapsto p_2$ , equals

$$\int_0^\infty \phi_\theta(y_0, y_1, p_2, x_1, x_2) e^{(1+x'_2\delta)a} \exp\left(-e^{(1+x'_2\delta)a} p_2\right) dp_2 = \psi_\theta(y_0, y_1, x_1, a).$$

Multiplying both sides by  $e^{-(1+x'_2\delta)a}$ , this is equivalent to

$$\mathcal{L}[\phi_\theta(y_0, y_1, \cdot, x_1, x_2)]\left(e^{(1+x'_2\delta)a}\right) = e^{-(1+x'_2\delta)a} \psi_\theta(y_0, y_1, x_1, a),$$

where  $\mathcal{L}[g](s) = \int_0^{+\infty} g(z) \exp(-sz) dz$  denotes the Laplace transform operator. Letting  $s = e^{(1+x'_2\delta)a}$ , we effectively wish to solve

$$\mathcal{L}[\phi_\theta(y_0, y_1, \cdot, x_1, x_2)](s) = s^{-1} \psi_\theta\left(y_0, y_1, x_1, \frac{\ln s}{1+x'_2\delta}\right),$$

and, provided  $s \mapsto s^{-1} \psi_\theta\left(y_0, y_1, x_1, \frac{\ln s}{1+x'_2\delta}\right)$  lies in the range of  $\mathcal{L}$ , we can back out  $\phi_\theta$  using the inverse Laplace transform. To illustrate, take

$$\psi_\theta(y_0, y_1, x_1, a) = p_1 - \exp(-(1+x'_1\delta)a),$$

from which we obtain

$$\mathcal{L}[\phi_\theta(y_0, y_1, \cdot, x_1, x_2)](s) = s^{-1} p_1 - s^{-\frac{1+x'_1\delta}{1+x'_2\delta}-1},$$

which, provided  $\frac{1+\delta'x_1}{1+\delta'x_2} > -1$ ,<sup>19</sup> admits the solution

$$\phi_\theta(y_0, y_1, y_2, x_1, x_2) = p_1 - \frac{p_2^{\frac{1+x'_1\delta}{1+x'_2\delta}}}{\Gamma\left(1 + \frac{1+x'_1\delta}{1+x'_2\delta}\right)}.$$

where  $\Gamma$  is the Gamma function. This gives the FHR moment function

$$\phi_\theta(y_0, y_1, y_2, x_1, x_2) = \left[ p_1 - \frac{p_2^{\frac{1+x'_1\delta}{1+x'_2\delta}}}{\Gamma\left(1 + \frac{1+x'_1\delta}{1+x'_2\delta}\right)} \right] \cdot m(y_0, x_1). \quad (51)$$

---

<sup>19</sup>A similar restriction on predetermined covariates features in the moment restrictions of the censored regression model of Honoré and Hu (2004).

More generally, one can obtain closed-form expressions if we choose  $\psi_\theta$  as a polynomial function of  $p_1$ .<sup>20</sup>

Example 5 illustrates how, using Corollary 2.1.1, one can derive moment functions by operator inversion. When suitable functions  $\psi_\theta$  exist, closed-form inversion delivers explicit moment functions. In other settings, it may be that the inverse is not available in closed form, and numerical inversion techniques need to be used (see, e.g., Engl et al., 1996).

### 6.3 Irregular moment conditions

In this section we have described an approach, based on Corollary 2.1.1, to construct moment functions  $\phi_\theta$  when those are available. However, for those functions to be helpful for estimation they need to be sufficiently regular. In this last part we provide examples that show how irregularity may arise, and how regularization techniques can help.

**Example 1** (Continued). (IRREGULARITY IN THE MPH MODEL) As an example, consider applying Corollary 2.1.1 to the MPH model, where we again focus on the two-period case for simplicity. We wish to solve for  $\phi_\theta$  in

$$\mathcal{L} [\phi_\theta(y_0, y_1, \cdot, x_1, x_2)] (e^a) = e^{-a} \psi_\theta(y_0, y_1, x_1, a),$$

that is, letting  $s = e^a$ , in

$$\mathcal{L} [\phi_\theta(y_0, y_1, \cdot, x_1, x_2)] (s) = s^{-1} \psi_\theta(y_0, y_1, x_1, \ln s).$$

Suppose in this case that we take  $\psi_\theta$  to be the score of the parametric model with respect to  $\theta$ , that is,

$$\psi_\theta(y_0, y_1, x_1, a) = z_1(1 - p_1 e^a).$$

Now, the solution to

$$\mathcal{L} [\phi_\theta(y_0, y_1, \cdot, x_1, x_2)] (s) = s^{-1} z_1(1 - p_1 s)$$

---

<sup>20</sup>For example, we can use for any  $b > 0$ ,

$$\psi_\theta(y_0, y_1, x_1, a) = p_1^b - \exp(-b(1 + x'_1 \delta)a) \Gamma(1 + b),$$

which has zero mean, and gives the FHR moment functions

$$\phi_\theta(y_0, y_1, y_2, x_1, x_2) = \left[ p_1^b - \frac{\Gamma(1 + b)}{\Gamma\left(1 + b \frac{1 + x'_1 \delta}{1 + x'_2 \delta}\right)} p_2^{\frac{1 + x'_1 \delta}{1 + x'_2 \delta}} \right] \cdot m(y_0, x_1),$$

which provides a continuum of possible moment functions on  $\theta$ .

is

$$\phi_\theta(y_0, y_1, y_2, x_1, x_2) = z_1(1 - p_1 \cdot \delta(p_2)),$$

where  $\delta(\cdot)$  is Dirac's delta. While  $\phi_\theta$  has zero expectation, it is a highly irregular function that cannot be directly used in GMM estimation. A possible strategy to address this issue is to regularize  $\phi_\theta$  by replacing  $\delta(p_2)$  with  $h^{-1}\kappa(p_2/h)$ , where  $\kappa$  is a nonparametric kernel and  $h > 0$  a bandwidth parameter. However, for fixed  $h$ , the regularized function  $\phi_\theta$  is no longer mean-zero, necessitating  $h$  to shrink to zero as the sample size increases to ensure consistent estimation of  $\theta$ . In the MPH model, it turns out that these difficulties can be entirely avoided. A rich set of regular moment functions exists, and, in fact, we have characterized the efficient moment function for this model.<sup>21</sup>

The regularization strategy we have outlined in the context of the MPH model can be useful in more complex models. A general strategy, when solving for  $\phi_\theta$  in the integral equation (13), is to use a regularized inverse of the relevant integral operator as in Carrasco et al. (2007). We now describe an example where this strategy can be successfully applied.

**Example 6. (NONLINEAR REGRESSION MODEL)** Consider the nonlinear panel data regression model

$$Y_t = m_\beta(Y_{t-1}, X_t, A) + \varepsilon_t, \quad \varepsilon_t | Y^{t-1}, X^t, A \sim \mathcal{N}(0, \sigma^2), \quad (52)$$

where  $a \mapsto m_\beta(y_{t-1}, x_t, a)$  is differentiable and strictly increasing. Here  $\theta = (\beta', \sigma^2)'$ . Let  $\lambda > 0$ , and let

$$K_\lambda(z) = \frac{1}{2\pi} \int \lambda \kappa(\lambda\tau) \exp\left(\lambda i\tau z + \frac{1}{2}\sigma^2\tau^2\right) d\tau,$$

where  $\kappa$  is the Fourier transform of a kernel function, satisfying  $\kappa(0) = 1$  and  $\kappa'(0) = 0$ ,  $|\kappa|$  is integrable, and where  $i$  is the imaginary number. A possible choice is  $\kappa(\tau) = \mathbf{1}\{\tau \in (-1, 1)\}$ , which is the Fourier transform of the sinc kernel.  $K_\lambda(z)$  corresponds to the deconvolution kernel introduced by Stefanski and Carroll (1990) in the context of a deconvolution problem with normal measurement error. Given any mean-zero function  $\psi_\theta(y_0, y_1, x_1, a)$ , we define

$$\phi_\theta^\lambda(y_0, y_1, y_2, x_1, x_2) = \int \psi_\theta(y_0, y_1, x_1, a) \frac{\partial m_\beta(y_1, x_2, a)}{\partial a} K_\lambda\left(\frac{m_\beta(y_1, x_2, a) - y_2}{\lambda}\right) da. \quad (53)$$

Applying Corollary 2.1.1 and a regularization strategy, we show in Supplemental Appendix

---

<sup>21</sup>These issues are not unique to the feedback setting. In the context of panel logit models with strictly exogenous covariates, the conditional likelihood estimator of Honoré and Kyriazidou (2000) can also be interpreted as relying on an irregular moment condition and requires kernel methods. Only recently did Honoré and Weidner (2024) demonstrate the existence of regular moment functions for this class of models. We are grateful to Manuel Arellano for this observation.

C that

$$\mathbb{E}_{\theta,\omega} [\phi_\theta^\lambda(Y_0, Y_1, Y_2, X_1, X_2)] \rightarrow 0 \text{ as } \lambda \rightarrow 0. \quad (54)$$

In this sense,  $\phi_\theta^\lambda$  provides an approximately valid moment function for  $\theta$ . As a special case, consider the linear Gaussian model where  $m_\beta(y_{t-1}, x_t, a) = \gamma y_{t-1} + x_t' \beta + a$ , and take  $\psi_\theta(y_0, y_1, x_1, a) = z_1(y_1 - \gamma y_0 - x_1' \beta - a)$ . Then (53) simplifies to

$$\begin{aligned} \phi_\theta^\lambda(y_0, y_1, y_2, x_1, x_2) &= \int z_1(y_1 - \gamma y_0 - x_1' \beta - a) K_\lambda \left( \frac{\gamma y_1 + x_2' \beta + a - y_2}{\lambda} \right) da \\ &= z_1 [(y_1 - \gamma y_0 - x_1' \beta) - (y_2 - \gamma y_1 - x_2' \beta)], \end{aligned}$$

which corresponds to the Arellano-Bond moment function for the case  $T = 2$ . Note that  $\phi_\theta^\lambda$  does not depend on  $\lambda$  in this case, so the regularization is immaterial.

## References

- Aguirregabiria, V. and Carro, J. M. (2021). Identification of average marginal effects in fixed effects dynamic discrete choice models. *arXiv preprint arXiv:2107.06141*.
- Aguirregabiria, V. and Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38 – 67.
- Ai, C. and Chen, X. (2012). The semiparametric efficiency bound for models of sequential moment restrictions containing unknown functions. *Journal of Econometrics*, 170(2):442–457.
- Al-Sadoon, M. M., Li, T., and Pesaran, H. (2017). Exponential class of dynamic binary choice panel data models with fixed effects. *Econometric Reviews*, 36(6-9).
- Arellano, M. (2016). Modelling optimal instrumental variables for dynamic panel data models. *Research in Economics*, 70(2):238 – 261.
- Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297.
- Arellano, M. and Bonhomme, S. (2011). Nonlinear panel data analysis. *Annu. Rev. Econ.*, 3(1):395–424.
- Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1):29–51.
- Arellano, M. and Honoré, B. (2001). Panel data models: some recent developments. In *Handbook of econometrics*, volume 5, pages 3229–3296. Elsevier.

- Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *Review of Economics and Statistics*, 60(1):47 – 57.
- Ashenfelter, O. and Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics*, 67(4):648 – 660.
- Blundell, R. and Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3):321 – 340.
- Blundell, R., Griffith, R., and Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics*, 108(1):113 – 131.
- Blundell, R. and Powell, J. L. (2003). *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, volume 2, chapter Endogeneity in nonparametric and semiparametric regression models, pages 312 – 357. Cambridge University Press.
- Blundell, R. W. and Powell, J. L. (2004). Endogeneity in semiparametric binary response models. *The Review of Economic Studies*, 71(3):655–679.
- Bonev, P. (2020). *Nonparametric identification in nonseparable duration models with unobserved heterogeneity*. School of Economics and Political Science, Department of Economics ....
- Bonhomme, S. (2012). Functional differencing. *Econometrica*, 80(4):1337 – 1385.
- Bonhomme, S., Dano, K., and Graham, B. S. (2023). Identification in a binary choice panel data model with a predetermined covariate. *SERIES: Journal of the Spanish Economic Association*, 14(3-4):315 – 351.
- Botosaru, I., Loh, I., and Muris, C. (2024). An adversarial approach to identification.
- Brown, B. W. and Newey, W. K. (1998). Efficient semiparametric estimation of expectations. *Econometrica*, 66(2):453–464.
- Buchinsky, M., Hahn, J., and Kim, K. i. (2010). Semiparametric information bound of dynamic discrete choice models. *Economic Letters*, 108(2):109 – 112.
- Carrasco, M., Florens, J.-P., and Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics*, 6:5633–5751.
- Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47(1):225 – 238.
- Chamberlain, G. (1982a). The general equivalence of granger and sims causality. *Econometrica*, 50(3):569 – 581.

- Chamberlain, G. (1982b). Multivariate regression models for panel data. *Journal of Econometrics*, 18(1):5 – 46.
- Chamberlain, G. (1984). *Handbook of Econometrics*, volume 2, chapter Panel Data, pages 1247 – 1318. North-Holland, Amsterdam.
- Chamberlain, G. (1985). *Longitudinal Analysis of Labor Market Data*, chapter Heterogeneity, omitted variable bias, and duration dependence, pages 3 – 38. Cambridge University Press, Cambridge.
- Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, 32(2):189 – 218.
- Chamberlain, G. (1992). Comment: sequential moment restrictions in panel data. *Journal of Business and Economic Statistics*, 10(2):20 – 26.
- Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica*, 78(1):159–168.
- Chamberlain, G. (2022). Feedback in panel data models. *Journal of Econometrics*, 226(1):4 – 20.
- Chamberlain, G. (2023). Identification in dynamic binary choice models. *SERIES*, 14(3):247–251.
- Chernozhukov, V., Fernández-Val, I., Hahn, J., and Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2):535–580.
- Chesher, A., Rosen, A. M., and Zhang, Y. (2024). Robust analysis of short panels. *arXiv preprint arXiv:2401.06611*.
- Dano, K. (2023). Transition probabilities and identifying moments in dynamic fixed effects logit models. *arXiv preprint arXiv:2303.00083*.
- Davezies, L., D'Haultfoeuille, X., and Laage, L. (2021). Identification and estimation of average marginal effects in fixed effects logit models. *arXiv preprint arXiv:2105.00879*.
- Dobronyi, C., Gu, J., et al. (2021). Identification of dynamic panel logit models with fixed effects. *arXiv preprint arXiv:2104.04590*.
- Elbers, C. and Ridder, G. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *The Review of Economic Studies*, 49(3):403–409.
- Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of inverse problems*, volume 375. Springer Science & Business Media.
- Ghanem, D., Sant'Anna, P. H., and Wüthrich, K. (2022). Selection and parallel trends. *arXiv preprint arXiv:2203.09001*.

- Graham, B. S., Pinto, C., and Egel, D. (2012). Inverse probability tilting for moment condition models with missing data. *Review of Economic Studies*, 79(3):1053 – 1079.
- Graham, B. S. and Powell, J. L. (2012). Identification and estimation of average partial effects in ‘irregular’ correlated random coefficient panel data models. *Econometrica*, 80(5):2105 – 2152.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424 – 438.
- Hahn, J. (1994). The efficiency bound of the mixed proportional hazard model. *The Review of Economic Studies*, 61(4):607–629.
- Hahn, J. (1997). A note on the efficient semiparametric estimation of some exponential panel models. *Econometric Theory*, 13(4):583 – 588.
- Heckman, J. J. (1991). Identifying the hand of past: distinguishing state dependence from heterogeneity. *American Economic Review*, 81(2):75 – 79.
- Heckman, J. J. and Borjas, G. J. (1980). Does unemployment cause future unemployment? definitions, questions and answers from a continuous time model of heterogeneity and state dependence. *Economica*, 47(187):247–283.
- Heckman, J. J. and Singer, B. (1984). The identifiability of the proportional hazard model. *Review of Economic Studies*, 51(2):231 – 241.
- Honoré, B. E. and Hu, L. (2004). Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics*, 122(2):293–316.
- Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874.
- Honoré, B. E. and Tamer, E. (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74(3):611–629.
- Honoré, B. E. and Weidner, M. (2024). Moment conditions for dynamic panel logit models with fixed effects. *Review of Economic Studies*, page rdae097.
- Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, 47(4):939 – 956.
- Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Number 17 in Econometric Society Monographs. Cambridge University Press, Cambridge.
- Lee, W. (2020). *Identification and estimation of dynamic random coefficient models*. PhD thesis, The University of Chicago.

- Lindsay, B. G. (1985). Using empirical partially bayes inference for increased efficiency. *The Annals of Statistics*, pages 914–931.
- Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of applied econometrics*, 5(2):99–135.
- Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.
- Nickell, S. (1980). Estimating the probability of leaving unemployment. *Econometrica*, 47(5):1249 – 1266.
- Olley, S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.
- Pakel, C. and Weidner, M. (2023). Bounds on average effects in discrete choice panel data models. *arXiv preprint arXiv:2309.09299*.
- Ridder, G. and Woutersen, T. M. (2003). The singularity of the information matrix of the mixed proportional hazards model. *Econometrica*, 71(5):1579 – 1589.
- Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512.
- Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 95–133. Springer.
- Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866.
- Stefanski, L. A. and Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics*, 21(2):169–184.
- van der Laan, M. J. and Robins, J. M. (2003). *United Methods for Censored Longitudinal Data*. Springer-Verlag, New York.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.
- Windmeijer, F. (2008). *The Econometrics of Panel Data*, volume 46 of *Advanced Studies in Theoretical and Applied Econometrics*, chapter GMM for panel data count models, pages 603 – 624. Springer, 3rd edition.

Wooldridge, J. M. (1997). Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory*, 13(5):667–678.

Woutersen, T. M. (2000). *Essays on the integrated hazard and orthogonality concepts*. Brown University.

## A Proofs of Main Results

### A.1 Proof of Theorem 2.1

**Part (A).** Suppose that  $\phi_\theta(Y^T, X^T)$  is absolutely integrable under DGP  $(\theta, \omega)$ . Then, by the law of iterated expectations, we have

$$\begin{aligned}\mathbb{E}_{\theta, \omega}[\phi_\theta(Y^T, X^T)] &= \mathbb{E}_{\theta, \omega} [\mathbb{E}_{\theta, \omega}[\phi_\theta(Y^T, X^T) | Y^{T-1}, X^{T-1}, A]] \\ &= \mathbb{E}_{\theta, \omega} \left[ \int \phi_\theta(Y^{T-1}, y_T, X^{T-1}, x_T) f_\theta(y_T | Y_{t-1}, x_T, A) g(x_T | Y^{T-1}, X^{T-1}, A) dy_T dx_T \right] \\ &= \mathbb{E}_{\theta, \omega} \left[ \int \phi_\theta(Y^{T-1}, y_T, X^{T-1}, x_T) f_\theta(y_T | Y_{t-1}, x_T, A) dy_T \right],\end{aligned}$$

for any arbitrary  $x_T$  value, where the last equality follows from (10) for  $s = T$ . By successive applications of the law of iterated expectations and (10) for  $s = T - 1, \dots, 2$ , we get

$$\mathbb{E}_{\theta, \omega}[\phi_\theta(Y^T, X^T)] = \mathbb{E}_{\theta, \omega} \left[ \int \phi_\theta(Y_0, y^{1:T}, X_1, x^{2:T}) \prod_{t=2}^T f_\theta(y_t | y_{t-1}, x_t, A) f_\theta(y_1 | Y_0, X_1, A) dy^{1:T} \right]$$

for any collection of regressor values  $x^{2:T}$ . Finally, using (9) implies  $\mathbb{E}_{\theta, \omega}[\phi_\theta(Y^T, X^T)] = 0$ .

**Part (B).** Suppose that, for all  $\omega \in \Omega$  such that  $\mathbb{E}_{\theta, \omega} [|\phi_\theta(Y^T, X^T)|] < \infty$  we have  $\mathbb{E}_{\theta, \omega}[\phi_\theta(Y^T, X^T)] = 0$ . Let  $\eta \mapsto \omega_\eta$  denote a smooth path such that  $\omega_{\eta^*} = \omega^*$  at some  $\eta^*$ . By (B)(ii), there exists a  $\kappa$ -ball around  $\eta^*$ ,  $B_\kappa(\eta^*)$ , such that for all  $\eta \in B_\kappa(\eta^*)$ ,  $\mathbb{E}_{\theta, \omega_\eta} [|\phi_\theta(Y^T, X^T)|] < \infty$  and thus  $\mathbb{E}_{\theta, \omega_\eta}[\phi_\theta(Y^T, X^T)] = 0$ . Next, by (B)(i) and (B)(ii), we can apply Lemma 5.4 in Newey and McFadden (1994), and conclude that  $\eta \mapsto \mathbb{E}_{\theta, \omega_\eta} [\phi_\theta(Y^T, X^T)]$  is differentiable at  $\eta^*$  with derivative  $\mathbb{E}_{\theta, \omega^*}[\phi_\theta(Y^T, X^T) S^\eta(Y^T, X^T)']$  where  $S^\eta(Y^T, X^T)$  denotes the score at  $\eta^*$ . Hence, since  $\mathbb{E}_{\theta, \omega_\eta}[\phi_\theta(Y^T, X^T)] = 0$  for all  $\eta \in B_\kappa(\eta^*)$ , we have  $\mathbb{E}_{\theta, \omega^*}[\phi_\theta(Y^T, X^T) S^\eta(Y^T, X^T)'] = 0$ . Moreover, by differentiability in quadratic mean,  $S^\eta$  is square-integrable under DGP  $(\theta, \omega^*)$ . Claim (B) then follows from Lemma A.1, whose proof is in Supplemental Appendix B.1.

**Lemma A.1.** Suppose that  $\mathbb{E}_{\theta,\omega^*}[\|\phi_\theta(Y^T, X^T)\|^2] < \infty$  and  $\mathbb{E}_{\theta,\omega^*}[\phi_\theta(Y^T, X^T)S^\eta(Y^T, X^T)'] = 0$  for all score functions  $S^\eta$  of smooth parametric submodels. Then (9)-(10) hold.

## A.2 Proof of Corollary 2.1.1

Assume (13). Then, since  $\sum_{t=1}^{T-1} \psi_{\theta,t}(y^t, x^t, a)$  does not depend on  $x_T$ , (10) holds for  $s = T$ . Next, integrating (13) with respect to  $f_\theta(y_{T-1} | y_{T-2}, x_{T-1}, a)$  implies that

$$\int \phi_\theta(y^T, x^T) f_\theta(y_T | y_{T-1}, x_{T-1}, a) f_\theta(y_{T-1} | y_{T-2}, x_{T-1}, a) dy^{T-1:T} = \sum_{t=1}^{T-2} \psi_{\theta,t}(y^t, x^t, a),$$

where we have used the fact that  $\mathbb{E} [\psi_{\theta,T-1}(Y^{T-1}, X^{T-1}, A) | Y^{T-2}, X^{T-1}, A] = 0$ . In particular, this implies (10) for  $s = T - 1$ . Further integrating with respect to  $f_\theta(y_{T-2} | y_{T-3}, x_{T-2}, a)$  then yields

$$\int \phi_\theta(y^T, x^T) \prod_{t=T-2}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{T-2:T} = \sum_{t=1}^{T-3} \psi_{\theta,t}(y^t, x^t, a),$$

since  $\mathbb{E} [\psi_{\theta,T-2}(Y^{T-2}, X^{T-2}, A) | Y^{T-3}, X^{T-2}, A] = 0$ . This implies (10) for  $s = T - 2$ . Continuing this reasoning, we easily conclude that for  $s = 2, \dots, T$ ,

$$\int \phi_\theta(y^T, x^T) \prod_{t=s}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{s:T} = \sum_{t=1}^{s-1} \psi_{\theta,t}(y^t, x^t, a),$$

implying (10) for  $s = 2, \dots, T$ . Finally, since  $\mathbb{E} [\psi_{\theta,1}(Y_0, Y_1, X_1, A) | Y_0, X_1, A] = 0$ , integrating the identity

$$\int \phi_\theta(y^T, x^T) \prod_{t=2}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{2:T} = \psi_{\theta,1}(y_0, y_1, x_1, a)$$

with respect to  $f_\theta(y_1 | y_0, x_1, a)$  yields (9).

Conversely, suppose that (9) and (10) hold. Equation (10) for  $s = T$  implies that we can write:

$$\int \phi_\theta(y^T, x^T) f_\theta(y_T | y_{T-1}, x_T, a) dy_T = \bar{\phi}_{\theta,T-1}(y^{T-1}, x^{T-1}, a), \quad (55)$$

for some function  $\bar{\phi}_{\theta,T-1}(y^{T-1}, x^{T-1}, a)$  that does not depend on  $x_T$ . Then, (10) for  $s = T - 1$

entails that:

$$\mathbb{E} [\bar{\phi}_{\theta,T-1}(Y^{T-1}, X^{T-1}, A) | Y^{T-2}, X^{T-1}, A] = \bar{\phi}_{\theta,T-2}(Y^{T-2}, X^{T-2}, A)$$

for some function  $\bar{\phi}_{\theta,T-2}(Y^{T-2}, X^{T-2}, A)$  that does not depend on  $X^{T-1:T}$ . Equivalently, we can write:

$$\bar{\phi}_{\theta,T-1}(Y^{T-1}, X^{T-1}, A) = \bar{\phi}_{\theta,T-2}(Y^{T-2}, X^{T-2}, A) + \psi_{\theta,T-1}(Y^{T-1}, X^{T-1}, A),$$

with  $\mathbb{E} [\psi_{\theta,T-1}(Y^{T-1}, X^{T-1}, A) | Y^{T-2}, X^{T-1}, A] = 0$ . Next, (10) for  $s = T - 2$  implies that:

$$\bar{\phi}_{\theta,T-2}(Y^{T-2}, X^{T-2}, A) = \bar{\phi}_{\theta,T-3}(Y^{T-3}, X^{T-3}, A) + \psi_{\theta,T-2}(Y^{T-2}, X^{T-2}, A),$$

for some function  $\bar{\phi}_{\theta,T-3}(Y^{T-3}, X^{T-3}, A)$  that does not depend on  $X^{T-2:T}$ , with  $\mathbb{E} [\psi_{\theta,T-2}(Y^{T-2}, X^{T-2}, A) | Y^{T-3}, X^{T-2}, A] = 0$ . Continuing this argument based on restriction (10), we conclude that, for  $s = 2, \dots, T - 1$ ,

$$\bar{\phi}_{\theta,s}(Y^s, X^s, A) = \bar{\phi}_{\theta,s-1}(Y^{s-1}, X^{s-1}, A) + \psi_{\theta,s}(Y^s, X^s, A),$$

such that  $\mathbb{E} [\psi_{\theta,s}(Y^s, X^s, A) | Y^{s-1}, X^s, A] = 0$ . Furthermore,

$$\bar{\phi}_{\theta,1}(Y_0, Y_1, X_1, A) = \psi_{\theta,1}(Y_0, Y_1, X_1, A),$$

with  $\mathbb{E} [\psi_{\theta,1}(Y_0, Y_1, X_1, A) | Y_0, X_1, A] = 0$  by (9). Collecting terms, we have shown that

$$\int \phi_{\theta}(y^T, x^T) f_{\theta}(y_T | y_{T-1}, x_T, a) dy_T = \bar{\phi}_{\theta,T-1}(y^{T-1}, x^{T-1}, a) = \sum_{t=1}^{T-1} \psi_{\theta,t}(y^t, x^t, a),$$

with the property that  $\mathbb{E} [\psi_{\theta,t}(Y^t, X^t, A) | Y^{t-1}, X^t, A] = 0$  for  $t = 1, \dots, T - 1$ . Hence (13).

### A.3 Proof of Theorem 3.1

Parts (A) and (B) follow from replacing  $\phi_{\theta}(y^T, x^T)$  by  $\varphi_{\theta}(y^T, x^T) - h_{\theta}(y^T, x^T, a)$  in the proof of Theorem 2.1, and from applying the triangle inequality to  $\varphi_{\theta}(y^T, x^T) - h_{\theta}(y^T, x^T, a)$ .

### A.4 Proof of Corollary 3.1.1

The proof is the same as the proof of Corollary 2.1.1, except for the fact that  $\phi_{\theta}(y^T, x^T)$  is replaced by  $\varphi_{\theta}(y^T, x^T) - h_{\theta}(y^T, x^T, a)$ .

## A.5 Proof of Theorem 4.1

Let  $\ell(\theta, \omega_\eta | y^T, x^T)$  denote a smooth parametric submodel, where  $\omega_\eta = (g_\eta, \pi_\eta, \nu_\eta)$  and  $\omega_{\eta_0} = \omega_0$  for some scalar  $\eta_0$ . Following the logic of Lemma A.1, but now using  $\eta_0$  in lieu of  $\eta^*$ , each submodel yields a  $(T + 1)$  dimensional score vector

$$S^\eta(y^T, x^T) = (S^{\eta, \pi}(y^T, x^T), S^{\eta, g, 2}(y^T, x^T), \dots, S^{\eta, g, T}(y^T, x^T), S^{\eta, \nu}(y^T, x^T))'$$

where the first component  $S^{\eta, \pi}(y^T, x^T)$  corresponds to the heterogeneity component; the next  $T - 1$  components  $S^{\eta, g, t}$ ,  $t = 2, \dots, T$ , are the scores for the feedback process at each period; and the final term  $S^{\eta, \nu}(y^T, x^T)$  is the score for the initial condition.

By definition, the nonparametric tangent set  $\mathcal{T}_{\theta_0, \omega_0, K}$  is the mean-square closure of elements  $AS^\eta$ , where  $A$  is a constant  $K \times (T + 1)$  matrix. Its orthocomplement is

$$\mathcal{T}_{\theta_0, \omega_0, K}^\perp = \{\phi \in \mathbb{R}^K \mid \mathbb{E}[\phi] = 0, \mathbb{E}[\phi' \phi] < \infty \text{ with } \mathbb{E}[\phi' s] = 0, \text{ for all } s \in \mathcal{T}_{\theta_0, \omega_0, K}\},$$

or, equivalently,

$$\begin{aligned} \mathcal{T}_{\theta_0, \omega_0, K}^\perp = \{\phi \in \mathbb{R}^K \mid \mathbb{E}[\phi] = 0, \mathbb{E}[\phi' \phi] < \infty \text{ with } \mathbb{E}[\phi S^\eta] = 0, \\ \text{for all scores } S^\eta \text{ of smooth parametric submodels}\}, \end{aligned}$$

by Lemma A.1 in Newey (1990). We have  $\phi_{\theta_0} \in \mathcal{T}_{\theta_0, \omega_0, K}^\perp$  if and only if  $\mathbb{E}[\phi_{\theta_0}(y^T, x^T) S^\eta(y^T, x^T)'] = 0$ . Thus, by Lemma A.1, we conclude that  $\mathcal{T}_{\theta_0, \omega_0, K}^\perp$  consists of the set of functions  $\phi_{\theta_0} \in \mathbb{R}^K$  such that each component satisfies conditions (9) and (10) of Theorem 2.1.

## A.6 Proof of Theorem 4.2

By construction,  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y^T, X^T) = \tilde{\Pi}(\tilde{S}^\theta(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \tilde{\omega}, K}^\perp)$  is an element of  $\mathcal{T}_{\theta_0, \tilde{\omega}, K}^\perp$ . Thus, it follows from Theorem 4.1 that  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  satisfies conditions (9) and (10). The moment restriction (36) is then a consequence of  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  being absolutely integrable under DGP  $(\theta_0, \omega_0)$  and Theorem 2.1. By the same argument,  $\mathbb{E}_{\theta_0, \omega_0} [\tilde{\Pi}(\varphi_{\theta_0}(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \tilde{\omega}, L}^\perp)] = 0$ . The moment condition (37) then follows from the definition of  $\tilde{\varphi}_{\theta_0, \tilde{\omega}}^{\text{eff}}$  in (35) and (32). Lastly, when  $\tilde{\omega} = \omega_0$  we have  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}} = \phi_{\theta_0, \omega_0}^{\text{eff}}$  and  $\tilde{\varphi}_{\theta_0, \tilde{\omega}}^{\text{eff}} = \varphi_{\theta_0, \omega_0}^{\text{eff}}$ , yielding the efficient moment functions for  $\theta_0$  and  $\mu_0$ , respectively.

# ONLINE SUPPLEMENTAL MATERIAL

## B Proofs of auxiliary results

### B.1 Proof of Lemma A.1

Let  $\ell(\theta, \omega_\eta | y^T, x^T)$  denote a smooth parametric submodel where  $\omega_\eta = (g_\eta, \pi_\eta, \nu_\eta)$  and  $\omega_{\eta^*} = \omega^*$ . We have  $(T+1)$  types of scores associated with submodels: one associated with the initial condition density,  $\nu_\eta$ , one associated with the heterogeneity distribution,  $\pi_\eta$ , and  $T-1$  scores associated with the feedback processes for  $X_2, \dots, X_T$ ,  $g_\eta$ . We begin with the heterogeneity component, which yields scores of the form:

$$S^{\eta, \pi}(y^T, x^T) = \frac{\int \nabla_\eta \ln \pi^*(a | y_0, x_1) p^*(y^{1:T}, x^{2:T}, a) da}{\int p^*(y^T, x^T, a) da},$$

where

$$p^*(y^T, x^T, a) = \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, a) \prod_{t=2}^T g^*(x_t | y^{t-1}, x^{t-1}, a) \pi^*(a | y_0, x_1),$$

and  $\nabla_\eta \ln \pi^*$  denotes the score of the submodel  $\eta \mapsto \pi_\eta$  at  $\omega^*$ .

Next we consider the scores for the  $(T-1)$  feedback components:

$$S^{\eta, g, t}(y^T, x^T) = \frac{\int \nabla_\eta \ln g^*(x_t | y^{t-1}, x^{t-1}, a) p^*(y^T, x^T, a) da}{\int p^*(y^T, x^T, a) da}, \quad t = 2, \dots, T.$$

Finally, for the initial condition, the scores take the form:

$$S^{\eta, \nu}(y^T, x^T) = \nabla_\eta \ln \nu^*(y_0, x_1).$$

Let  $S^\eta(y^T, x^T) = (S^{\eta, \pi}(y^T, x^T), S^{\eta, g, 2}(y^T, x^T), \dots, S^{\eta, g, T}(y^T, x^T), S^{\eta, \nu}(y^T, x^T))'$ . Without loss of generality, suppose  $\phi_\theta$  is scalar. By assumption, we have

$$\begin{aligned} 0 &= \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) S^{\eta, \pi}(Y^T, X^T)] \\ &= \int \phi_\theta(y^T, x^T) \nabla_\eta \ln \pi^*(a | y_0, x_1) p^*(y^T, x^T, a) \nu^*(y_0, x_1) da dy^T dx^T, \end{aligned} \tag{56}$$

and, for  $t = 2, \dots, T$ ,

$$\begin{aligned} 0 &= \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) S^{\eta, g, t}(Y^T, X^T)] \\ &= \int \phi_\theta(y^T, x^T) \nabla_\eta \ln g^*(x_t | y^{t-1}, x^{t-1}, a) p^*(y^T, x^T, a) \nu^*(y_0, x_1) dady^T dx^T, \end{aligned} \quad (57)$$

and

$$\begin{aligned} 0 &= \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) S^{\eta, \nu}(Y^T, X^T)] \\ &= \int \phi_\theta(y^T, x^T) \nabla_\eta \ln \nu^*(y_0, x_1) p^*(y^T, x^T, a) \nu^*(y_0, x_1) dady^T dx^T. \end{aligned} \quad (58)$$

Note that equation (57) for  $t = T$ , after integrating over  $y_T$ , coincides with:

$$\begin{aligned} 0 &= \int \nabla_\eta \ln g^*(x_T | y^{T-1}, x^{T-1}, a) \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-1} = y^{T-1}, X^T = x^T, A = a] \\ &\quad \times \prod_{t=1}^{T-1} f_\theta(y_t | y_{t-1}, x_t, a) \prod_{t=2}^T g^*(x_t | y^{t-1}, x^{t-1}, a) \pi^*(a | y_0, x_1) \nu^*(y_0, x_1) dady^{T-1} dx^T. \end{aligned}$$

Now, since  $\nabla_\eta \ln g^*(x_T | y^{T-1}, x^{T-1}, a)$  is unrestricted except for the fact that it has zero mean conditional on  $(y^{T-1}, x^{T-1}, a)$  and is square-integrable, we can specifically choose

$$\begin{aligned} \nabla_\eta \ln g^*(x_T | y^{T-1}, x^{T-1}, a) &= \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-1} = y^{T-1}, X^T = x^T, A = a] \\ &\quad - \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-1} = y^{T-1}, X^{T-1} = x^{T-1}, A = a], \end{aligned}$$

in which case (57) (for  $t = T$ ) evaluates to (for  $\mathbb{V}_{\theta, \omega^*}$  the variance under  $(\theta, \omega^*)$ ):

$$\begin{aligned} \int \mathbb{V}_{\theta, \omega^*} (\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-1} = y^{T-1}, X^T = x^T, A = a] | Y^{T-1} = y^{T-1}, X^{T-1} = x^{T-1}, A = a) \\ \times \prod_{t=1}^{T-1} f_\theta(y_t | y_{t-1}, x_t, a) \prod_{t=2}^T g^*(x_t | y^{t-1}, x^{t-1}, a) \pi^*(a | y_0, x_1) \nu^*(y_0, x_1) dady^{T-1} dx^T = 0, \end{aligned}$$

which holds if, and only if,  $\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-1} = y^{T-1}, X^T = x^T, A = a]$  does not depend on  $x_T$ . Conversely, if  $\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-1} = y^{T-1}, X^T = x^T, A = a]$  does not depend on  $x_T$ , it is clear that equation (57) for  $t = T$  is satisfied since the elements  $\nabla_\eta \ln g^*(x_T | y^{T-1}, x^{T-1}, a)$  have zero mean conditional on  $(y^{T-1}, x^{T-1}, a)$ . Therefore, (57) for  $t = T$  is equivalent to the requirement that  $\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-1} = y^{T-1}, X^T = x^T, A = a]$  does not depend on  $x_T$ .

Next, consider (57) for  $t = T - 1$ . Exploiting the result above, we can integrate over  $y_{T-1:T}$

and  $x_T$  to get:

$$0 = \int \nabla_\eta \ln g^*(x_{T-1} | y^{T-2}, x^{T-2}, a) \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-2} = y^{T-2}, X^{T-1} = x^{T-1}, A = a] \\ \times \prod_{t=1}^{T-2} f_\theta(y_t | y_{t-1}, x_t, a) \prod_{t=2}^{T-1} g^*(x_t | y^{t-1}, x^{t-1}, a) \pi^*(a | y_0, x_1) \nu^*(y_0, x_1) dy^{T-2} dx^{T-1}.$$

Observe that the lack of restrictions on  $\nabla_\eta \ln g^*(x_{T-1} | y^{T-2}, x^{T-2}, a)$ , besides it being mean zero conditional on  $(y^{T-2}, x^{T-2}, a)$  and being square-integrable, implies, by a logic analogous to that used for the  $t = T$  case, that (57) for  $t = T - 1$  is equivalent to the requirement that

$$\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y^{T-1} = y^{T-1}, X^{T-1} = x^{T-1}, A = a] \\ = \int \phi_\theta(y^T, x^T) \prod_{t=T-1}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{T-1:T}$$

does not depend on  $x^{T-1:T}$ . By inductive reasoning, we conclude that, for all  $s = 2, \dots, T$ ,

$$\int \phi_\theta(y^T, x^T) \prod_{t=s}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{s:T} \text{ does not depend on } x^{s:T}. \quad (59)$$

That is,  $\phi_\theta$  satisfies (10).

Next, consider equation (56). Using the results immediately above we get, after integrating over  $y^{1:T}$  and  $x^{2:T}$ :

$$\int \nabla_\eta \ln \pi^*(a | y_0, x_1) \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1, A = a] \pi^*(a | y_0, x_1) \nu^*(y_0, x_1) da dy_0 dx_1 = 0,$$

and since  $\nabla_\eta \ln \pi^*(a | y_0, x_1)$  is unrestricted (beyond having zero mean conditional on  $(y_0, x_1)$  and being square-integrable), choosing specifically

$$\nabla_\eta \ln \pi^*(a | y_0, x_1) = \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1, A = a] \\ - \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1]$$

implies that

$$\int \mathbb{V}_{\theta, \omega^*} (\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1, A = a] | Y_0 = y_0, X_1 = x_1) \\ \times \pi^*(a | y_0, x_1) \nu^*(y_0, x_1) da dy_0 dx_1 = 0.$$

Thus,  $\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1, A = a]$  does not depend on  $a$ . Conversely, if

$\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1, A = a]$  does not depend on  $a$ , then (56) also holds since  $\nabla_\eta \ln \pi^*(a | y_0, x_1)$  has mean-zero conditional on  $(y_0, x_1)$ . We therefore conclude that (56) is equivalent to

$$\begin{aligned} & \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1, A = a] \\ &= \int \phi_\theta(y^T, x^T) \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, a) \prod_{t=2}^T g^*(x_t | y^{t-1}, x^{t-1}, a) dy^{1:T} dx^{2:T} \\ &= \int \phi_\theta(y^T, x^T) \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{1:T} \quad (\text{by (59) for } s = 2) \end{aligned}$$

is a constant independent of  $a$ . Hence, we can write

$$\int \phi_\theta(y^T, x^T) \prod_{t=1}^T f_\theta(y_t | y_{t-1}, x_t, a) dy^{1:T} = \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1]. \quad (60)$$

Next, we also note that (58) can simply be rewritten as:

$$0 = \int \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1] \nabla_\eta \ln \nu^*(y_0, x_1) \nu^*(y_0, x_1) dy_0 dx_1.$$

Since  $\nabla_\eta \ln \nu^*(y_0, x_1)$  is unrestricted besides being mean-zero and square-integrable, a similar reasoning to before implies that (58) is equivalent to  $\mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T) | Y_0 = y_0, X_1 = x_1] = \mathbb{E}_{\theta, \omega^*} [\phi_\theta(Y^T, X^T)] = 0$ . This result in combination with (60) implies that  $\phi_\theta$  satisfies (9).

## B.2 Proof of Lemma 2.2

Recall the bijective transformation between  $(y_1, \dots, y_T)$  and  $(p_1, \dots, p_T)$  given by

$$p_t = \Lambda_\alpha(y_t) e^{\gamma y_{t-1} + x'_t \beta}, \quad t = 1, \dots, T,$$

as well as the second (bijective) transformation between  $(p_1, \dots, p_T)$  and  $(\tilde{p}_1, \dots, \tilde{p}_{T-1}, \bar{p})$  given by

$$\begin{aligned} \tilde{p}_t &= \frac{p_t}{\sum_{s=t}^T p_s}, \quad t = 1, \dots, T-1, \\ \bar{p} &= \sum_{t=1}^T p_t. \end{aligned}$$

The distribution of  $(P_1, \dots, P_T)$  is stated in Lemma D.1 and that of  $(\tilde{P}_1, \dots, \tilde{P}_{T-1}, \bar{P})$  in Lemma D.2.

From these two bijections we get the following two equivalent representations of  $\phi_\theta$ :

$$\begin{aligned} \xi_\theta(y_0, p^T, x^T) = & \phi_\theta(y_0, \Lambda_\alpha^{-1}(p_1 \exp(-\gamma y_0 - x'_1 \beta)), \\ & \dots, \Lambda_\alpha^{-1}(p_T \exp(-\gamma y_{T-1} - x'_T \beta)), x^T) \end{aligned} \quad (61)$$

$$\psi_\theta(y_0, \tilde{p}^{T-1}, \bar{p}, x^T) = \xi_\theta(y_0, \tilde{p}_1 \bar{p}, (1 - \tilde{p}_1) \tilde{p}_2 \bar{p}, \dots, \prod_{s=1}^{T-2} (1 - \tilde{p}_s) \tilde{p}_{T-1} \bar{p}, \prod_{s=1}^{T-1} (1 - \tilde{p}_s) \bar{p}, x^T). \quad (62)$$

In the present context, condition (10) of Theorem 2.1 for  $s = T$  corresponds to a requirement on (61) of

$$\int_0^{+\infty} \xi_\theta(y_0, p^T, x^T) e^a e^{-e^a p_T} dp_T \text{ does not depend on } x_T,$$

where we have used the change of variable,  $p_T = \Lambda_\alpha(y_T) \exp(\gamma y_{T-1} + x'_T \beta)$  (see Lemma D.1). This implies that  $\mathcal{L}[p_T \mapsto \xi_\theta(y_0, p^{T-1}, p_T, x^{T-1}, x_T)][e^a] = \mathcal{L}[p_T \mapsto \xi_\theta(y_0, p^{T-1}, p_T, x^{T-1}, \tilde{x}_T)][e^a]$  for any  $\tilde{x}_T$ , where  $\mathcal{L}$  denotes the Laplace transform operator. By the uniqueness of the Laplace transform, it follows that  $\xi_\theta(y_0, p^{T-1}, p_T, x^{T-1}, x_T) = \xi_\theta(y_0, p^{T-1}, p_T, x^{T-1}, \tilde{x}_T)$ , meaning that  $\xi_\theta$  does not depend on  $x_T$ . Hereafter we therefore suppress the dependence of  $\phi_\theta, \xi_\theta, \psi_\theta$  on  $x_T$ .

Next, consider the decomposition

$$\begin{aligned} \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) = & \\ & \left( \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) - \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{T-2}, \bar{P}, X^{T-1} \right] \right) \\ & + \left( \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{T-2}, \bar{P}, X^{T-1} \right] - \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{T-3}, \bar{P}, X^{T-2} \right] \right) \\ & + \left( \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{T-3}, \bar{P}, X^{T-2} \right] - \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{T-4}, \bar{P}, X^{T-3} \right] \right) \\ & \vdots \\ & + \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}_1, \bar{P}, X^2 \right], \end{aligned}$$

or, succinctly,

$$\psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) = \sum_{t=1}^{T-2} \psi_{\theta,t}(Y_0, \tilde{P}^t, \bar{P}, X^{t+1}) + \psi_{\theta,T-1}(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}),$$

where

$$\psi_{\theta,T-1}(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) = \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) - \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{T-2}, \bar{P}, X^{T-1} \right],$$

for all  $t = 2, \dots, T-2$ ,

$$\begin{aligned} & \psi_{\theta,t}(Y_0, \tilde{P}^t, \bar{P}, X^{t+1}) \\ &= \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^t, \bar{P}, X^{t+1} \right] - \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{t-1}, \bar{P}, X^t \right], \end{aligned}$$

and

$$\psi_{\theta,1}(Y_0, \tilde{P}_1, \bar{P}, X^2) = \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}_1, \bar{P}, X^2 \right].$$

The law of iterated expectations readily implies that:

$$\begin{aligned} \mathbb{E} \left[ \psi_{\theta,T-1}(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{T-2}, \bar{P}, X^{T-1} \right] &= 0, \\ \mathbb{E} \left[ \psi_{\theta,t}(Y_0, \tilde{P}^t, \bar{P}, X^{t+1}) \middle| Y_0, \tilde{P}^{t-1}, \bar{P}, X^t \right] &= 0, \quad t = 2, \dots, T-2. \end{aligned}$$

It only remains to show that (i) for all  $t = 1, \dots, T-1$ ,  $\psi_{\theta,t}(Y_0, \tilde{P}^t, \bar{P}, X^{t+1})$  does not depend on  $X_{t+1}$  and (ii)

$$\mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \bar{P}, X_1 \right] = 0.$$

We start by establishing (i). To that end, fix any  $s \in \{2, \dots, T-1\}$ . By Theorem 2.1 equation (10), we have

$$\int_0^\infty \xi_\theta(y_0, p^T, x^{T-1}) e^a e^{-e^a \sum_{t=s}^T p_t} dp^{s:T} \text{ does not depend on } x^{s:T},$$

using the change of variables,  $p_t = \Lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x'_t \beta)$  for  $t = s, \dots, T$ . Next, recall that  $\tilde{p}_t = \frac{p_t}{\sum_{k=t}^T p_k}$ ,  $t = s, \dots, T-1$ , and introduce  $\bar{p}_s = \sum_{k=s}^T p_k$ . By Lemma D.2 and a second change of variables, the previous condition is equivalent to

$$\int_0^\infty \Upsilon_\theta(y_0, p^{s-1}, \bar{p}_s, x^{T-1}) \frac{e^{(T-s+1)a}}{\Gamma(T-s+1)} (\bar{p}_s)^{T-s} e^{-e^a \bar{p}_s} d\bar{p}_s \text{ does not depend on } x^{s:T},$$

where

$$\begin{aligned} \Upsilon_\theta(y_0, p^{s-1}, \bar{p}_s, x^{T-1}) &= \int_0^1 \xi_\theta(y_0, p^{s-1}, \tilde{p}_s \bar{p}_s, (1 - \tilde{p}_s) \tilde{p}_{s+1} \bar{p}_s, \dots, \prod_{k=s}^{T-2} (1 - \tilde{p}_k) \tilde{p}_{T-1} \bar{p}_s, \prod_{k=s}^{T-1} (1 - \tilde{p}_k) \bar{p}_s, x^{T-1}) \\ &\quad \times \prod_{t=s}^{T-1} \frac{\Gamma(T-t+1)}{\Gamma(T-t)} (1 - \tilde{p}_t)^{T-t-1} d\tilde{p}^{s:T-1}. \end{aligned}$$

Hence,

$$\mathcal{L}[\bar{p}_s \mapsto (\bar{p}_s)^{T-s} \Upsilon_\theta(y_0, p^{s-1}, \bar{p}_s, x^{s-1}, x^{s:T-1})][e^a] = \mathcal{L}[\bar{p}_s \mapsto (\bar{p}_s)^{T-s} \Upsilon_\theta(y_0, p^{s-1}, \bar{p}_s, x^{s-1}, \tilde{x}^{s:T-1})][e^a],$$

for any possible values  $\tilde{x}^{s:T-1}$ . By uniqueness of the Laplace transform, we conclude that  $\Upsilon_\theta$  does not depend on  $x^{s:T-1}$ . Since there is a bijective transformation between  $(p^{s-1}, \bar{p}_s)$  and  $(\tilde{p}^{s-1}, \bar{p})$  given by  $\tilde{p}_t = \frac{p_t}{\sum_{k=t}^{s-1} p_k + \bar{p}_s}$ ,  $t = 1, \dots, s-1$ , and  $\bar{p} = \sum_{k=1}^{s-1} p_k + \bar{p}_s$ , we have:

$$\begin{aligned} \Upsilon_\theta(y_0, p^{s-1}, \bar{p}_s, x^{T-1}) &= \int_0^1 \psi_\theta(y_0, \tilde{P}^{T-1}, \bar{P}, x^{T-1}) \prod_{t=s}^{T-1} \frac{\Gamma(T-t+1)}{\Gamma(T-t)} (1 - \tilde{p}_t)^{T-t-1} d\tilde{p}^{s:T-1} \\ &= \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0 = y_0, \tilde{P}^{s-1} = \tilde{p}^{s-1}, \bar{P} = \bar{p}, X^s = x^s \right], \end{aligned}$$

which does not depend on  $x^{s:T-1}$ . The second equality follows from the fact that the “forward orthogonal transforms”  $\tilde{P}^{s:T-1}$  are independent of  $\tilde{P}^{s-1}, \bar{P}, X^s$  by part (i) of Lemma D.1 and Lemma D.2. This shows that, for  $s \in \{3, \dots, T-1\}$ ,

$$\begin{aligned} &\psi_{\theta,s-1}(Y_0, \tilde{P}^{s-1}, \bar{P}, X^s) \\ &= \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{s-1}, \bar{P}, X^s \right] - \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}^{s-2}, \bar{P}, X^{s-1} \right] \end{aligned}$$

does not depend on  $X^{s:T}$ , and that, for  $s = 2$

$$\psi_{\theta,1}(Y_0, \tilde{P}_1, \bar{P}, X^2) = \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \middle| Y_0, \tilde{P}_1, \bar{P}, X^2 \right]$$

does not depend on  $X^{2:T}$ . This proves (i).

Finally, by Theorem 2.1 equation (9), we have

$$\int_0^\infty \xi_\theta(y_0, p^T, x^{T-1}) e^a e^{-e^a \sum_{t=1}^T p_t} dp^T = 0,$$

where we have once more used the transformation  $p_t = \Lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x'_t \beta)$  for  $t = 1, \dots, T$ . Consider now,  $\tilde{p}_t = \frac{p_t}{\sum_{k=t}^T p_k}$ ,  $t = 1, \dots, T-1$  and  $\bar{p} = \sum_{k=1}^T p_k$ . By Lemma D.2

and a second change of variables, the previous condition is equivalent to

$$\int_0^\infty \Upsilon_\theta(y_0, \bar{p}, x^{T-1}) \frac{e^{Ta}}{\Gamma(T)} \bar{p}^{T-1} e^{-e^a \bar{p}} d\bar{p} = 0,$$

where now  $\Upsilon_\theta(y_0, \bar{p}, x^{T-1}) = \left( \int_0^1 \psi_\theta(y_0, \tilde{p}^{T-1}, \bar{p}, x^{T-1}) \prod_{t=1}^{T-1} \frac{\Gamma(T-t+1)}{\Gamma(T-t)} (1 - \tilde{p}_t)^{T-t-1} d\tilde{p}^{T-1} \right)$ . Hence,

$$\mathcal{L}[\bar{p} \mapsto \bar{p}^{T-1} \Upsilon_\theta(y_0, \bar{p}, x^{T-1})][e^a] = 0,$$

which by uniqueness of the Laplace transform implies that

$$0 = \Upsilon_\theta(Y_0, \bar{P}, X^{T-1}) = \mathbb{E} \left[ \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \mid Y_0, \bar{P}, X_1 \right],$$

where the second equality is again a consequence of Lemma D.2. This establishes (ii) which finally yields the desired representation.

### B.3 Supplementary lemma

**Lemma B.1.** *Let  $T \geq 2$  and suppose that  $\varphi_\theta^1(Y^T, X^T), \varphi_\theta^2(Y^T, X^T)$  are two  $L \times 1$  square-integrable moment functions under  $(\theta_0, \omega_0)$  verifying (32) as well as (24) and (25) of Theorem 3.1. Then,  $\Pi(\varphi_{\theta_0}^1(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \omega_0, L}) = \Pi(\varphi_{\theta_0}^2(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \omega_0, L})$ .*

*Proof.* By assumption,  $\phi_\theta(Y^T, X^T) = \varphi_\theta^1(Y^T, X^T) - \varphi_\theta^2(Y^T, X^T) \in \mathcal{T}_{\theta_0, \omega_0, L}^\perp$  since it satisfies the conditions (9)-(10) of Theorem 2.1. Then, we can write

$$\begin{aligned} \varphi_\theta^1(Y^T, X^T) &= \varphi_\theta^2(Y^T, X^T) + \phi_\theta(Y_0, Y^T, X^T) \\ &= \mu(\theta, \omega) + \Pi(\varphi_{\theta_0}^2(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \omega_0, L}) + \Pi(\varphi_{\theta_0}^2(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \omega_0, L}^\perp) + \phi_\theta(Y_0, Y^T, X^T). \end{aligned}$$

By linearity of  $\mathcal{T}_{\theta_0, \omega_0, L}^\perp$ ,  $\Pi(\varphi_{\theta_0}^2(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \omega_0, L}^\perp) + \phi_\theta(Y_0, Y^T, X^T) \in \mathcal{T}_{\theta_0, \omega_0, L}^\perp$ . Therefore, standard properties of Hilbert spaces imply that  $\Pi(\varphi_{\theta_0}^1(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \omega_0, L}) = \Pi(\varphi_{\theta_0}^2(Y^T, X^T) \mid \mathcal{T}_{\theta_0, \omega_0, L})$ .  $\square$

## C Nonlinear regression: derivation of equation (54)

Let  $\psi_\theta(y_0, y_1, x_1, a)$  denote a moment function satisfying (45). Following Corollary 2.1.1, we search for functions  $\phi_\theta$  satisfying equation (13), i.e.,

$$\int_{-\infty}^{+\infty} \phi_\theta(y_0, y_1, y_2, x_1, x_2) \exp\left(-\frac{1}{2\sigma^2}(y_2 - m_\beta(y_1, x_2, a))\right)^2 dy_2 = \psi_\theta(y_0, y_1, x_1, a),$$

which is a convolution equation. Thus, if we let  $\tau = m_\beta(y_1, x_2, a)$ , an application of the Convolution Theorem yields

$$\mathcal{F}[\phi_\theta(y_0, y_1, x_1, x_2)] [s] \exp\left(-\frac{\sigma^2}{2}s^2\right) = \mathcal{F}[\psi_\theta(y_0, y_1, x_1, m_\beta^{-1}(y_1, x_2, \tau))] [s],$$

where  $\mathcal{F}[g] [s] = \int_{-\infty}^{+\infty} g(z) \exp(izs) dz$  denotes the Fourier transform operator, and we have used that the characteristic function of a  $\mathcal{N}(0, \sigma^2)$  is given by  $s \mapsto \exp\left(-\frac{\sigma^2}{2}s^2\right)$ . More explicitly, we have

$$\begin{aligned} & \int_{-\infty}^{+\infty} \phi_\theta(y_0, y_1, y_2, x_1, x_2) \exp(isy_2) dy_2 \\ &= \int_{-\infty}^{+\infty} \psi_\theta(y_0, y_1, x_1, m_\beta^{-1}(y_1, x_2, \tau)) \exp\left(is\tau + \frac{\sigma^2}{2}s^2\right) d\tau \\ &= \int_{-\infty}^{+\infty} \psi_\theta(y_0, y_1, x_1, a) \frac{\partial m_\beta(y_1, x_2, a)}{\partial a} \exp\left(ism_\beta(y_1, x_2, a) + \frac{\sigma^2}{2}s^2\right) da, \end{aligned}$$

where we have used the change in variables  $a \mapsto \tau = m_\beta(y_1, x_2, a)$ .

However, the inverse problem may not have solutions since the inverse Fourier transform

$$\frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-isy_2) \left[ \int_{-\infty}^{+\infty} \psi_\theta(y_0, y_1, x_1, a) \frac{\partial m_\beta(y_1, x_2, a)}{\partial a} \exp\left(ism_\beta(y_1, x_2, a) + \frac{\sigma^2}{2}s^2\right) da \right] ds$$

may not be well-defined. To address this issue, our strategy is to regularize the problem, and compute the regularized inverse

$$\begin{aligned} & \phi_\theta^\lambda(y_0, y_1, y_2, x_1, x_2) \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \lambda \kappa(\lambda s) \exp(-isy_2) \left[ \int_{-\infty}^{+\infty} \psi_\theta(y_0, y_1, x_1, a) \frac{\partial m_\beta(y_1, x_2, a)}{\partial a} \exp\left(ism_\beta(y_1, x_2, a) + \frac{\sigma^2}{2}s^2\right) da \right] ds \\ &= \int_{-\infty}^{+\infty} \psi_\theta(y_0, y_1, x_1, a) \frac{\partial m_\beta(y_1, x_2, a)}{\partial a} \left[ \frac{1}{2\pi} \int_{-\infty}^{+\infty} \lambda \kappa(\lambda s) \exp\left(is(m_\beta(y_1, x_2, a) - y_2) + \frac{1}{2}\sigma^2 s^2\right) ds \right] da \\ &= \int_{-\infty}^{+\infty} \psi_\theta(y_0, y_1, x_1, a) \frac{\partial m_\beta(y_1, x_2, a)}{\partial a} K_\lambda\left(\frac{m_\beta(y_1, x_2, a) - y_2}{\lambda}\right) da, \end{aligned}$$

which is always well-defined (under suitable conditions so that the integrals can be interchanged). This is similar to the strategy used in kernel nonparametric deconvolution (Stefanski and Carroll, 1990).

The expectation of this regularized inverse is

$$\begin{aligned}
& \mathbb{E} [\phi_\theta^\lambda(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, X_1] \\
&= \mathbb{E} \left[ \int_{-\infty}^{+\infty} \psi_\theta(Y_0, Y_1, X_1, a) \frac{\partial m_\beta(Y_1, X_2, a)}{\partial a} K_\lambda \left( \frac{m_\beta(Y_1, X_2, a) - Y_2}{\lambda} \right) da | Y_0, X_1 \right] \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \lambda \kappa(\lambda s) \int_{-\infty}^{+\infty} \mathbb{E} \left[ \psi_\theta(Y_0, Y_1, X_1, a) \frac{\partial m_\beta(Y_1, X_2, a)}{\partial a} \right. \\
&\quad \times \exp \left( \lambda \mathbf{i} s \frac{m_\beta(Y_1, X_2, a) - Y_2}{\lambda} + \frac{1}{2} \sigma^2 s^2 \right) | Y_0, X_1 \Big] da ds \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \lambda \kappa(\lambda s) \int_{-\infty}^{+\infty} \mathbb{E} \left[ \psi_\theta(Y_0, Y_1, X_1, a) \frac{\partial m_\beta(Y_1, X_2, a)}{\partial a} \right. \\
&\quad \times \exp \left( \lambda \mathbf{i} s \frac{m_\beta(Y_1, X_2, a) - \varepsilon_{i2} - m_\beta(Y_1, X_2, A)}{\lambda} + \frac{1}{2} \sigma^2 s^2 \right) | Y_0, X_1 \Big] da ds \\
&= \int_{-\infty}^{+\infty} \mathbb{E} \left[ \psi_\theta(Y_0, Y_1, X_1, a) \frac{\partial m_\beta(Y_1, X_2, a)}{\partial a} \right. \\
&\quad \times \frac{1}{2\pi} \int_{-\infty}^{+\infty} \lambda \kappa(\lambda s) \exp (\mathbf{i} s (m_\beta(Y_1, X_2, a) - m_\beta(Y_1, X_2, A))) ds | Y_0, X_1 \Big] da \\
&= \int_{-\infty}^{+\infty} \mathbb{E} \left[ \psi_\theta(Y_0, Y_1, X_1, a) \frac{\partial m_\beta(Y_1, X_2, a)}{\partial a} \right. \\
&\quad \times \frac{1}{2\pi} \int_{-\infty}^{+\infty} \kappa(u) \exp \left( \mathbf{i} \frac{u}{\lambda} (m_\beta(Y_1, X_2, a) - m_\beta(Y_1, X_2, A)) \right) du | Y_0, X_1 \Big] da.
\end{aligned}$$

Since  $|\kappa(u) \exp (\mathbf{i} \frac{u}{\lambda} (m_\beta(Y_1, X_2, a) - m_\beta(Y_1, X_2, A)))| \leq |\kappa(u)|$  and by assumption  $|\kappa(u)|$  is integrable, the dominated convergence theorem implies

$$\begin{aligned}
& \lim_{\lambda \rightarrow 0} \mathbb{E} [\phi_\theta^\lambda(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, X_1] \\
&= \int_{-\infty}^{+\infty} \mathbb{E} \left[ \psi_\theta(Y_0, Y_1, X_1, a) \frac{\partial m_\beta(Y_1, X_2, a)}{\partial a} \delta(m_\beta(Y_1, X_2, a) - m_\beta(Y_1, X_2, A)) | Y_0, X_1 \right] da,
\end{aligned}$$

where  $\delta(\cdot)$  denotes Dirac's delta. Finally, since by assumption  $a \mapsto m_\beta(y_1, x_2, a)$  is strictly

increasing we have

$$\begin{aligned}
& \lim_{\lambda \rightarrow 0} \mathbb{E} [\phi_\theta^\lambda(Y_0, Y_1, Y_2, X_1, X_2) | Y_0, X_1] \\
&= \int_{-\infty}^{+\infty} \mathbb{E} [\psi_\theta(Y_0, Y_1, X_1, a) \delta(a - A) | Y_0, X_1] da \\
&= \mathbb{E} \left[ \int_{-\infty}^{+\infty} \psi_\theta(Y_0, Y_1, X_1, a) \delta(a - A) da | Y_0, X_1 \right] \\
&= \mathbb{E} [\psi_\theta(Y_0, Y_1, X_1, A) | Y_0, X_1] \\
&= 0,
\end{aligned}$$

where the first line follows from the composition property with the Dirac delta.

# Additional details for the MPH model and the numerical experiments

## D Calculations supporting the MPH results presented in main text

### D.1 Some useful distributional properties of the MPH model

Our derivations of the results presented in the main text for the MPH model exploit a number of its special distributional properties. Variants of these properties feature in prior work, for example that of Hahn (1994) and Ridder and Woutersen (2003). Our analysis requires these known implications of the MPH as well as several (apparently) new ones, specific to the feedback case. This section of the supplemental appendix derives and presents the needed preliminary results.

Recall the definitions  $\rho_\theta(z_t) = \Lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x'_t \beta)$  for  $z_t = (y_{t-1}, y_t, x'_t)'$  and  $P_t = \rho_\theta(Z_t)$  for  $t = 1, \dots, T$  with  $P_t = p_t = \rho_\theta(z_t)$  denoting – when the context is clear – a *specific* value of the random variable  $P_t$ . Unless there is a risk of confusion we omit the dependence of  $P_t$  on  $\theta$ . The period  $t$  conditional survival function equals

$$\begin{aligned} \Pr(Y_t > y_t | y^{t-1}, x^t, a) &= S_\theta(y_t | y^{t-1}, x^t, a) = \exp(-\Lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x'_t \beta + a)) \\ &= \exp(-\rho_\theta(z_t) e^a), \end{aligned} \quad (63)$$

for  $t = 1, \dots, T$ . Using monotonicity of the integrated baseline hazard and (63) we have

$$\begin{aligned} \Pr(P_t > p_t | y^{t-1}, x^t, a) &= \Pr(Y_t > \Lambda_\alpha^{-1}(p_t \exp(-\gamma y_{t-1} - x'_t \beta)) | y^{t-1}, x^t, a) \\ &= \exp(-\Lambda_\alpha(\Lambda_\alpha^{-1}(p_t \exp(-\gamma y_{t-1} - x'_t \beta))) \exp(\gamma y_{t-1} + x'_t \beta + a)) \\ &= \exp(-p_t e^a). \end{aligned} \quad (64)$$

Observe that (64) is the survival function for an exponential random variable with rate parameter  $e^a$ . Since the mapping from  $y_t \rightarrow p_t$  is invertible we therefore have that

$$P_t | P^{t-1}, X^t, A \sim \text{Exponential}(e^A), \quad t = 1, \dots, T,$$

and also that

$$f(x^{2:T}, p^{1:T} | y_0, x_1, a) = \exp\left(Ta - \left(\sum_{t=1}^T p_t\right) e^a\right) \prod_{t=2}^T \tilde{g}(x_t | y_0, p^{1:t-1}, x^{t-1}, a),$$

where  $\tilde{g}(x_t | y_0, p^{1:t-1}, x^{t-1}, a) = g(x_t | y_0, y^{1:t-1}, x^{t-1}, a)$  for  $p_t = \rho_\theta(z_t)$  and  $t = 1, \dots, T$ . Integrating over  $x^{2:T}$  gives the following lemma.

**Lemma D.1.** (EXPONENTIAL STRUCTURE OF THE MPH MODEL) *For the MPH model defined by Example 1 we have, for  $t = 1, \dots, T$ , that (i)  $P_t | P^{t-1}, X^t, A \sim \text{Exponential}(e^A)$  and (ii)  $P_t | Y_0, X_1, A \stackrel{\text{ind}}{\sim} \text{Exponential}(e^A)$ .*

Our characterization of the set of FHR moments for the MPH model uses a particular one-to-one transformation of  $P^T$ . We next derive the properties of this transformation. Let  $P_t \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha_t, \beta)$  for  $t = 1, \dots, T$  and consider the bijective forward orthogonal transformation  $V^T = G(P^T)$ :

$$V_t = \frac{P_t}{\sum_{s=t}^T P_s}, \quad t = 1, \dots, T-1$$

$$V_T = \sum_{t=1}^T P_t,$$

with an inverse given by  $P^T = G^{-1}(V^T)$ :

$$P_1 = V_1 V_T$$

$$P_t = \prod_{s=1}^{t-1} (1 - V_s) V_t V_T, \quad t = 2, \dots, T-1$$

$$P_T = \prod_{s=1}^{T-1} (1 - V_s) V_T.$$

By the change-of-variable formula,

$$f_V(V_1, V_2, \dots, V_T) = f_{P_1}(V_1 V_T) \prod_{t=2}^{T-1} f_{P_t} \left( \prod_{s=1}^{t-1} (1 - V_s) V_t V_T \right) f_{P_T} \left( \prod_{s=1}^{T-1} (1 - V_s) V_T \right) \left| \det \left( \frac{dG^{-1}(V^T)}{dV^T} \right) \right|.$$

We shall show that:

$$V_T \sim \text{Gamma} \left( \sum_{t=1}^T \alpha_t, \beta \right), \quad (65)$$

$$V_t \sim \text{Beta} \left( \alpha_t, \sum_{s=t+1}^T \alpha_s \right), \quad t = 1, \dots, T-1,$$

with  $(V_1, V_2, \dots, V_{T-1}, V_T)$  additionally mutually independent of one another.

To this end, let  $J^{(T)}(V_1, \dots, V_T) = \det \left[ \frac{dG^{-1}(V^T)}{dV^T} \right]$ . In Subsection D.3 we establish that:

$$J^{(T)}(V_1, \dots, V_T) = V_T^{T-1} \prod_{s=1}^{T-2} (1 - V_s)^{T-(s+1)}. \quad (66)$$

Applying the change-of-variables formula then yields, as we show in Subsection D.3,

$$f_V(V_1, V_2, \dots, V_T) = \frac{\beta^{\sum_{t=1}^T \alpha_t}}{\Gamma\left(\sum_{t=1}^T \alpha_t\right)} V_T^{\sum_{t=1}^T \alpha_t - 1} e^{-\beta V_T} \quad (67)$$

$$\times \prod_{t=1}^{T-1} \frac{\Gamma\left(\sum_{s=t}^T \alpha_s\right)}{\Gamma(\alpha_t) \Gamma\left(\sum_{s=t+1}^T \alpha_s\right)} V_t^{\alpha_t - 1} (1 - V_t)^{\sum_{s=t+1}^T \alpha_s - 1}, \quad (68)$$

from which claim (65) immediately follows. This result, in combination with Lemma D.1 above gives, for  $\tilde{P}_t$  for  $t = 1, \dots, T-1$  and  $\bar{P}$  as defined in the main text, the following lemma.

**Lemma D.2.** (“HELMERT TRANSFORMATION” FOR THE MPH MODEL) *For the MPH model defined by Example 1 we have, conditional on  $Y_0, X_1$  and  $A$ , (i)*

$$\begin{aligned} \tilde{P}_t &\sim \text{Beta}(1, T-t), \quad t = 1, \dots, T-1, \\ \bar{P} &\sim \text{Gamma}(T, e^A), \end{aligned}$$

with (ii)  $(\tilde{P}_1, \tilde{P}_2, \dots, \tilde{P}_{T-1}, \bar{P})$  additionally mutually independent of one another.

## D.2 Semiparametric efficiency bounds for the MPH model

Here we present our semiparametric efficiency bound derivations for the MPH model. Our calculations exploit Lemma 2.2 and Theorem 4.1 of the main text. In order to compute the efficient score for the MPH using equation (30) in the main text, the following lemma is useful.

**Lemma D.3.** (MPH PROJECTION) *For any  $L \times 1$  mean-zero random vector  $\varphi_\theta(Y^T, X^T)$  such that  $\mathbb{E} [\|\varphi_\theta(Y^T, X^T)\|^2] < \infty$ , we have*

$$\Pi(\varphi_\theta(Y^T, X^T) | \mathcal{T}_{\theta, \omega, L}^\perp) = \sum_{t=1}^{T-1} \varphi_{\theta, t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t),$$

with

$$\varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t) = \mathbb{E} \left[ \varphi_\theta(Y^T, X^T) | Y_0, \tilde{P}^t, \bar{P}, X^t \right] - \mathbb{E} \left[ \varphi_\theta(Y^T, X^T) | Y_0, \tilde{P}^{t-1}, \bar{P}, X^t \right].$$

*Proof.* First, we note that by iterated expectations,

$$\mathbb{E} \left[ \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t) | Y_0, \tilde{P}^{t-1}, \bar{P}, X^t \right] = 0.$$

Thus, Lemma 2.2 and Theorem 4.1 imply that  $\sum_{t=1}^{T-1} \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t)$  is an element of  $\mathcal{T}_{\theta,\omega,L}^\perp$ . Next fix  $s, t \in \{1, \dots, T-1\}$  and consider any function  $\psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) \in \mathcal{T}_{\theta,\omega,L}^\perp$  such that

$$\mathbb{E} \left[ \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) | Y_0, \tilde{P}^{s-1}, \bar{P}, X^s \right] = 0.$$

Observe that if  $s > t$

$$\begin{aligned} & \mathbb{E} \left[ \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t)' \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) \right] \\ &= \mathbb{E} \left[ \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t)' \mathbb{E} \left[ \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) | Y_0, \tilde{P}^{s-1}, \bar{P}, X^s \right] \right] \\ &= 0. \end{aligned}$$

Symmetrically, if  $t > s$ ,

$$\mathbb{E} \left[ \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t)' \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) \right] = 0,$$

since by construction  $\mathbb{E} \left[ \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t) | Y_0, \tilde{P}^{t-1}, \bar{P}, X^t \right] = 0$ . Using these observations we can show:

$$\begin{aligned} & \mathbb{E} \left[ \left( \varphi_\theta(Y^T, X^T) - \sum_{t=1}^{T-1} \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t) \right)' \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) \right] \\ &= \mathbb{E} \left[ \left( \varphi_\theta(Y^T, X^T) - \varphi_{\theta,s}^\perp(Y_0, \tilde{P}^s, \bar{P}, X^s) \right)' \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) \right] \\ &= \mathbb{E} \left[ \left( \varphi_\theta(Y^T, X^T) - \mathbb{E} \left[ \varphi_\theta(Y^T, X^T) | Y_0, \tilde{P}^s, \bar{P}, X^s \right] \right)' \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) \right] \\ &\quad + \mathbb{E} \left[ \mathbb{E} \left[ \varphi_\theta(Y^T, X^T) | Y_0, \tilde{P}^{s-1}, \bar{P}, X^s \right]' \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) \right] \\ &= 0, \end{aligned}$$

where the second equality uses the definition of  $\varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t)$  given in the statement

of the Lemma and where the last line follows from iterated expectations and the fact that  $\mathbb{E} \left[ \psi_{\theta,s}(Y_0, \tilde{P}^s, \bar{P}, X^s) | Y_0, \tilde{P}^{s-1}, \bar{P}, X^s \right] = 0$ . Since by Lemma 2.2 and Theorem 4.1, any function of  $\mathcal{T}_{\theta,\omega,L}^\perp$  is of the form:

$$\psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) = \sum_{t=1}^{T-1} \psi_{\theta,t}(Y_0, \tilde{P}^t, \bar{P}, X^t),$$

with  $\mathbb{E} \left[ \psi_{\theta,t}(Y_0, \tilde{P}^t, \bar{P}, X^t) | Y_0, \tilde{P}^{t-1}, \bar{P}, X^t \right] = 0$ ,  $t = 1, \dots, T-1$ , these calculations show that

$$\mathbb{E} \left[ \left( \varphi_\theta(Y^T, X^T) - \sum_{t=1}^{T-1} \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t) \right) \psi_\theta(Y_0, \tilde{P}^{T-1}, \bar{P}, X^{T-1}) \right] = 0.$$

By the Projection Theorem (e.g Theorem 11.1 in Van der Vaart (2000)) we conclude that,

$$\Pi(\varphi_\theta(Y_0, Y^T, X^T) | \mathcal{T}_{\theta,\omega,L}^\perp) = \sum_{t=1}^{T-1} \varphi_{\theta,t}^\perp(Y_0, \tilde{P}^t, \bar{P}, X^t),$$

as claimed.  $\square$

### D.2.1 MPH efficient score under feedback

We can use Lemma D.3 to derive an explicit expression of the efficient score for  $\theta = (\alpha, \beta', \gamma)'$ . Given the historical and continued importance of the MPH model, the required calculations, although involved and tedious, are presented in detail here.

Recall the form the integrated likelihood for the semiparametric panel data model with feedback presented in equation (5) of the main text. The form of the parametric period  $t$  panel data model equals, in the MPH setting,

$$f_\theta(y_t | x_t, y_{t-1}, a) = \lambda_\alpha(y_t) \exp(\gamma y_{t-1} + x_t' \beta + a) \exp(-\rho_\theta(z_t) e^a).$$

Further recall the notation:  $P_t = \rho_\theta(Z_t) = \Lambda_\alpha(Y_t) e^{\gamma Y_{t-1} + X_t' \beta}$  (when evaluated at population value  $\theta = \theta_0$ ). The (submodel) score for  $\theta$  equals

$$S^\theta(Y^T, X^T) = \mathbb{E} \left[ \sum_{t=1}^T \frac{\partial \ln f_\theta(Y_t | Y_{t-1}, X_t, A)}{\partial \theta} | Y^T, X^T \right].$$

By Theorem 4.1 and Lemma D.3 the efficient score, when  $T = 2$ , equals

$$\phi_\theta^{\text{eff}}(Y^2, X^2) = \mathbb{E} \left[ S^\theta(Y^2, X^2) \mid Y_0, \tilde{P}_1, \bar{P}, X_1 \right] - \mathbb{E} \left[ S^\theta(Y^2, X^2) \mid Y_0, \bar{P}, X_1 \right], \quad (69)$$

while, in the general  $T \geq 2$  case, it equals

$$\phi_\theta^{\text{eff}}(Y^T, X^T) = \sum_{t=1}^{T-1} \phi_{\theta,t}^{\text{eff},\perp}(Y_0, \tilde{P}^{1:t}, \bar{P}, X^{1:t}),$$

with, for  $t = 1, \dots, T-1$ ,

$$\begin{aligned} & \phi_{\theta,t}^{\text{eff},\perp}(Y_0, \tilde{P}^{1:t}, \bar{P}, X^{1:t}) \\ &= \mathbb{E} \left[ S^\theta(Y^T, X^T) \mid Y_0, \tilde{P}^{1:t}, \bar{P}, X^{1:t} \right] - \mathbb{E} \left[ S^\theta(Y^T, X^T) \mid Y_0, \tilde{P}^{1:t-1}, \bar{P}, X^{1:t} \right]. \end{aligned} \quad (70)$$

In what follows we focus on the  $T = 2$  special case for simplicity (and because it is this case we study in the numerical experiments reported in Section 5 of the main text). The calculations proceed by evaluating – and simplifying as far as appears to be feasible – equation (69) above. We do this separately for  $\beta$ ,  $\gamma$  and  $\alpha$ . Our efficient score expression for  $\beta$  holds for any baseline hazard, those we present for  $\gamma$  and  $\alpha$  assume that the baseline hazard takes the Weibull form:  $\lambda_\alpha(y_t) = \alpha y_t^{\alpha-1}$ . The calculations presented below can be adapted to study semiparametric efficiency bounds under other maintained baseline hazard assumptions.

**The efficient score for  $\beta$ :** The (submodel) score for  $\beta$ , recalling the notation re-established immediately above, equals:

$$\begin{aligned} S^\beta(Y^T, X^T) &= \mathbb{E} \left[ \sum_{t=1}^T \frac{\partial \ln f_\theta(Y_t \mid Y_{t-1}, X_t, A)}{\partial \beta} \mid Y^T, X^T \right] \\ &= \mathbb{E} \left[ \sum_{t=1}^T X_t (1 - P_t e^A) \mid Y^T, X^T \right] \\ &= \sum_{t=1}^T X_t - \sum_{t=1}^T X_t P_t \mathbb{E} [e^A \mid Y^T, X^T], \end{aligned} \quad (71)$$

which, when  $T = 2$ , coincides with

$$S^\beta(Y^2, X^2) = X_1 + X_2 - (X_1 P_1 + X_2 P_2) \mathbb{E} [e^A \mid Y^2, X^2].$$

Reparameterizing in terms of  $\tilde{P}_1$  and  $\bar{P}$  yields the equivalent expression

$$S^\beta(Y^2, X^2) = X_1 + X_2 - \left( X_1 \tilde{P}_1 \bar{P} + X_2 (1 - \tilde{P}_1) \bar{P} \right) \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2].$$

The efficient score for  $\beta$ , from equation (69) above, equals

$$\phi_\theta^{\text{eff},\beta}(Y^2, X^2) = \mathbb{E} [S^\beta(Y^2, X^2) | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [S^\beta(Y^2, X^2) | Y_0, \bar{P}, X_1]. \quad (72)$$

We evaluate the two expectations to the right of the equality above in sequence. First,

$$\begin{aligned} & \mathbb{E} [S^\beta(Y^2, X^2) | Y_0, \tilde{P}_1, \bar{P}, X_1] \\ &= X_1 + \mathbb{E} [X_2 | Y_0, \tilde{P}_1, \bar{P}, X_1] - X_1 \tilde{P}_1 \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] - (1 - \tilde{P}_1) \bar{P} \mathbb{E} [X_2 e^A | Y_0, \tilde{P}_1, \bar{P}, X_1]. \end{aligned}$$

Second, and making use of Lemma D.2 when evaluating expectations over  $\tilde{P}_1$ ,

$$\begin{aligned} & \mathbb{E} [S^\beta(Y^2, X^2) | Y_0, \bar{P}, X_1] \\ &= X_1 + \mathbb{E} [X_2 | Y_0, \bar{P}, X_1] - X_1 \bar{P} \mathbb{E} [\tilde{P}_1 e^A | Y_0, \bar{P}, X_1] - \bar{P} \mathbb{E} [X_2 (1 - \tilde{P}_1) e^A | Y_0, \bar{P}, X_1] \\ &= X_1 + \mathbb{E} [X_2 | Y_0, \bar{P}, X_1] - \frac{1}{2} X_1 \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] - \bar{P} \mathbb{E} [X_2 (1 - \tilde{P}_1) e^A | Y_0, \bar{P}, X_1]. \end{aligned}$$

Collecting terms we get an efficient score for  $\beta$  of

$$\begin{aligned} \phi_\theta^{\text{eff},\beta}(Y^2, X^2) &= -X_1 \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\ &\quad + \mathbb{E} [X_2 | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [X_2 | Y_0, \bar{P}, X_1] \\ &\quad - \left( (1 - \tilde{P}_1) \mathbb{E} [X_2 e^A | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [X_2 (1 - \tilde{P}_1) e^A | Y_0, \bar{P}, X_1] \right) \bar{P}, \end{aligned}$$

or more concisely

$$\begin{aligned} \phi_\theta^{\text{eff},\beta}(Y^2, X^2) &= -X_1 \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\ &\quad + \mathbb{E} \left[ X_2 \left( 1 - (1 - \tilde{P}_1) \bar{P} e^A \right) | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] \\ &\quad - \mathbb{E} \left[ X_2 \left( 1 - (1 - \tilde{P}_1) \bar{P} e^A \right) | Y_0, \bar{P}, X_1 \right], \end{aligned} \quad (73)$$

as claimed in Section 5.

**The efficient score for  $\gamma$ :** The (submodel) score for  $\gamma$  equals:

$$\begin{aligned} S^\gamma(Y^T, X^T) &= \mathbb{E} \left[ \sum_{t=1}^T \frac{\partial \ln f_\theta(Y_t | Y_{t-1}, X_t, A)}{\partial \gamma} | Y^T, X^T \right] \\ &= \mathbb{E} \left[ \sum_{t=1}^T Y_{t-1} (1 - P_t e^A) | Y^T, X^T \right] \\ &= \sum_{t=1}^T Y_{t-1} - \sum_{t=1}^T Y_{t-1} P_t \mathbb{E} [e^A | Y^T, X^T], \end{aligned}$$

which, when  $T = 2$ , coincides with

$$S^\gamma(Y^2, X^2) = Y_0 + Y_1 - (Y_0 P_1 + Y_1 P_2) \mathbb{E} [e^A | Y^2, X^2],$$

or, after reparameterization in terms of  $(\tilde{P}_1, \bar{P})$ , with

$$S^\gamma(Y^2, X^2) = Y_0 + Y_1 - (Y_0 \tilde{P}_1 \bar{P} + Y_1 (1 - \tilde{P}_1) \bar{P}) \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2].$$

The efficient score for  $\gamma$ , from equation (69) above, equals

$$\phi_\theta^{\text{eff}, \gamma}(Y^2, X^2) = \mathbb{E} [S^\gamma(Y^2, X^2) | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [S^\gamma(Y^2, X^2) | Y_0, \bar{P}, X_1], \quad (74)$$

where evaluate the first expectation to the right of equality as

$$\begin{aligned} \mathbb{E} [S^\gamma(Y^2, X^2) | Y_0, \tilde{P}_1, \bar{P}, X_1] &= Y_0 + Y_1 - (Y_0 \tilde{P}_1 \bar{P} + Y_1 (1 - \tilde{P}_1) \bar{P}) \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X_1] \\ &= Y_0 + Y_1 - (Y_0 \tilde{P}_1 \bar{P} + Y_1 (1 - \tilde{P}_1) \bar{P}) \mathbb{E} [e^A | Y_0, \bar{P}, X_1], \end{aligned}$$

using the fact that  $\tilde{P}_1 \perp (A, \bar{P}) | X_0, Y_1$  (see Lemma D.2 above).

In order to evaluate the second expectation entering the efficient score for  $\gamma$  we consider the special case of a Weibull baseline hazard:  $\Lambda_\alpha(y_t) = y_t^\alpha$ ,  $\lambda_\alpha(y_t) = \alpha y_t^{\alpha-1}$ . This hazard implies the convenient relation  $Y_1 = P_1^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} = \tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0}$ . This allows us to use Lemma D.2 to evaluate the expectation:

$$\begin{aligned} \mathbb{E} [Y_1 | Y_0, \bar{P}, X_1] &= e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \mathbb{E} [\tilde{P}_1^{\frac{1}{\alpha}} | Y_0, \bar{P}, X_1] \bar{P}^{\frac{1}{\alpha}} \\ &= \frac{\alpha}{1 + \alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P}^{\frac{1}{\alpha}}, \end{aligned}$$

where we have used the fact that  $\tilde{P}_1 \sim U[0, 1] = \text{Beta}(1, 1)$ . Proceeding in the same general

way we also get

$$\begin{aligned}\mathbb{E} \left[ Y_1(1 - \tilde{P}_1) | Y_0, \bar{P}, X_1 \right] &= e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \mathbb{E} \left[ \tilde{P}_1^{\frac{1}{\alpha}} (1 - \tilde{P}_1) | Y_0, \bar{P}, X_1 \right] \bar{P}^{\frac{1}{\alpha}} \\ &= \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P}^{\frac{1}{\alpha}}.\end{aligned}$$

Using these two results we evaluate the second expectation in equation (74) as:

$$\begin{aligned}\mathbb{E} [S^\gamma(Y^2, X^2) | Y_0, \bar{P}, X_1] &= Y_0 + \mathbb{E} [Y_1 | Y_0, \bar{P}, X_1] \\ &\quad - \left( Y_0 \mathbb{E} [\tilde{P}_1 | Y_0, \bar{P}, X_1] + \mathbb{E} [Y_1(1 - \tilde{P}_1) | Y_0, \bar{P}, X_1] \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\ &= Y_0 + \frac{\alpha}{1 + \alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P}^{\frac{1}{\alpha}} \\ &\quad - \left( Y_0 \frac{1}{2} + \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P}^{\frac{1}{\alpha}} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1].\end{aligned}$$

Collecting terms yields an efficient score for  $\gamma$  of:

$$\begin{aligned}\phi_\theta^{\text{eff}, \gamma}(Y^2, X^2)(\tilde{P}_1, \bar{P}, X_1) &= Y_1 - \frac{\alpha}{1 + \alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P}^{\frac{1}{\alpha}} \\ &\quad - \left( Y_0 \left( \tilde{P}_1 - \frac{1}{2} \right) + \left( Y_1(1 - \tilde{P}_1) - \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P}^{\frac{1}{\alpha}} \right) \right) \\ &\quad \times \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1]. \quad (75)\end{aligned}$$

This expression does not appear in the main text.

**The efficient score for  $\alpha$ :** The (submodel) score for  $\alpha$  equals:

$$\begin{aligned}S^\alpha(Y^T, X^T) &= \mathbb{E} \left[ \sum_{t=1}^T \frac{\partial \ln f_\theta(Y_t | Y_{t-1}, X_t, A)}{\partial \alpha} | Y^T, X^T \right] \\ &= \mathbb{E} \left[ \sum_{t=1}^T \frac{\partial \ln \lambda_\alpha(Y_t)}{\partial \alpha} - \frac{\partial \ln \Lambda_\alpha(Y_t)}{\partial \alpha} P_t e^A | Y^T, X^T \right] \\ &= \sum_{t=1}^T \frac{\partial \ln \lambda_\alpha(Y_t)}{\partial \alpha} - \sum_{t=1}^T \frac{\partial \ln \Lambda_\alpha(Y_t)}{\partial \alpha} P_t \mathbb{E} [e^A | Y^T, X^T],\end{aligned}$$

which, when  $T = 2$ , equals (after reparameterization in terms of  $\tilde{P}_1, \bar{P}$ ):

$$\begin{aligned}S^\alpha(Y^2, X^2) &= \frac{\partial \ln \lambda_\alpha(Y_1)}{\partial \alpha} + \frac{\partial \ln \lambda_\alpha(Y_2)}{\partial \alpha} \\ &\quad - \left( \frac{\partial \ln \Lambda_\alpha(Y_1)}{\partial \alpha} \tilde{P}_1 + \frac{\partial \ln \Lambda_\alpha(Y_2)}{\partial \alpha} \times (1 - \tilde{P}_1) \right) \bar{P} \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X_1, X_2],\end{aligned}$$

and the efficient score for  $\alpha$ , from equation (69) above, equals

$$\phi_{\theta}^{\text{eff},\alpha}(Y^2, X^2) = \mathbb{E} \left[ S^{\alpha}(Y^2, X^2) | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] - \mathbb{E} \left[ S^{\alpha}(Y^2, X^2) | Y_0, \bar{P}, X_1 \right]. \quad (76)$$

To make further progress we again consider the Weibull baseline hazard case; implying  $\ln \Lambda_{\alpha}(Y_t) = \alpha \ln Y_t$  and  $\ln \lambda_{\alpha}(Y_t) = \ln \alpha + (\alpha - 1) \ln Y_t$ . This yields

$$\begin{aligned} \frac{\partial \ln \Lambda_{\alpha}(Y_t)}{\partial \alpha} &= \ln Y_t \\ \frac{\partial \ln \lambda_{\alpha}(Y_t)}{\partial \alpha} &= \frac{1}{\alpha} + \ln Y_t. \end{aligned}$$

In the parameterization  $\tilde{P}_1, \bar{P}$ , with  $Y_1 = P_1^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} = \tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0}$  and  $Y_2 = (1 - \tilde{P}_1)^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0}$  we have

$$\begin{aligned} \frac{\partial \ln \Lambda_{\alpha}(Y_1)}{\partial \alpha} &= \ln Y_1 = -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) + \frac{1}{\alpha} \ln \tilde{P}_1 + \frac{1}{\alpha} \ln \bar{P} \\ \frac{\partial \ln \Lambda_{\alpha}(Y_2)}{\partial \alpha} &= \ln Y_2 = -\frac{1}{\alpha} (X'_2 \beta + \gamma Y_1) + \frac{1}{\alpha} \ln(1 - \tilde{P}_1) + \frac{1}{\alpha} \ln \bar{P} \\ &= -\frac{1}{\alpha} \left( X'_2 \beta + \gamma \tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \right) + \frac{1}{\alpha} \ln(1 - \tilde{P}_1) + \frac{1}{\alpha} \ln \bar{P}, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \ln \lambda_{\alpha}(Y_1)}{\partial \alpha} &= \frac{1}{\alpha} - \frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) + \frac{1}{\alpha} \ln \tilde{P}_1 + \frac{1}{\alpha} \ln \bar{P} \\ \frac{\partial \ln \lambda_{\alpha}(Y_2)}{\partial \alpha} &= \frac{1}{\alpha} - \frac{1}{\alpha} \left( X'_2 \beta + \gamma \tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \right) + \frac{1}{\alpha} \ln(1 - \tilde{P}_1) + \frac{1}{\alpha} \ln \bar{P}. \end{aligned}$$

The following implications of Lemma D.2 are useful for the calculations which follow:  $\tilde{P}_1 \perp \bar{P} | Y_0, X_1$  with  $\tilde{P}_1 \sim U[0, 1]$  and hence, by the properties of the uniform distribution,  $-\ln \tilde{P}_1 \sim \text{Exp}(1)$ . Using these facts we evaluate the following four expectations:

$$\begin{aligned} \mathbb{E} \left[ -\ln(1 - \tilde{P}_1) | Y_0, \bar{P}, X_1 \right] &= \mathbb{E} \left[ -\ln(1 - \tilde{P}_1) | Y_0, X_1 \right] = \mathbb{E} \left[ -\ln \tilde{P}_1 | Y_0, X_1 \right] = +1 \\ \mathbb{E} \left[ \tilde{P}_1^{\frac{1}{\alpha}} | Y_0, \bar{P}, X_1 \right] &= \frac{\alpha}{\alpha + 1} \\ \mathbb{E} \left[ \tilde{P}_1^{\frac{1}{\alpha}} (1 - \tilde{P}_1) | Y_0, \bar{P}, X_1 \right] &= \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) \\ \mathbb{E} \left[ \ln \tilde{P}_1 \tilde{P}_1 | Y_0, \bar{P}, X_1 \right] &= -\frac{1}{4}. \end{aligned}$$

Using these calculations we can then evaluate the various conditional expectations in the

efficient score expression. We start with

$$\mathbb{E} \left[ \frac{\partial \ln \lambda_\alpha(Y_1)}{\partial \alpha} | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] = \frac{1}{\alpha} - \frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) + \frac{1}{\alpha} \ln \tilde{P}_1 + \frac{1}{\alpha} \ln \bar{P},$$

and

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \ln \lambda_\alpha(Y_1)}{\partial \alpha} | Y_0, \bar{P}, X_1 \right] &= \frac{1}{\alpha} - \frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) - \frac{1}{\alpha} \mathbb{E} \left[ -\ln \tilde{P}_1 | Y_0, \bar{P}, X_1 \right] + \frac{1}{\alpha} \ln \bar{P} \\ &= -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) + \frac{1}{\alpha} \ln \bar{P}, \end{aligned}$$

which together give

$$\mathbb{E} \left[ \frac{\partial \ln \lambda_\alpha(Y_1)}{\partial \alpha} | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] - \mathbb{E} \left[ \frac{\partial \ln \lambda_\alpha(Y_1)}{\partial \alpha} | Y_0, \bar{P}, X_1 \right] = \frac{1}{\alpha} + \frac{1}{\alpha} \ln \tilde{P}_1.$$

Second we evaluate

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \ln \lambda_\alpha(Y_2)}{\partial \alpha} | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] &= \frac{1}{\alpha} - \frac{1}{\alpha} \left( \mathbb{E} \left[ X_2 | Y_0, \tilde{P}_1, \bar{P}, X_1 \right]' \beta + \gamma \tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \right) \\ &\quad + \frac{1}{\alpha} \ln (1 - \tilde{P}_1) + \frac{1}{\alpha} \ln \bar{P}, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \ln \lambda_\alpha(Y_2)}{\partial \alpha} | Y_0, \bar{P}, X_1 \right] &= \frac{1}{\alpha} - \frac{1}{\alpha} \left( \mathbb{E} \left[ X_2 | Y_0, \bar{P}, X_1 \right]' \beta + \gamma \mathbb{E} \left[ \tilde{P}_1^{\frac{1}{\alpha}} | Y_0, \bar{P}, X_1 \right] \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \right) \\ &\quad - \frac{1}{\alpha} \mathbb{E} \left[ -\ln (1 - \tilde{P}_1) | Y_0, \bar{P}, X_1 \right] + \frac{1}{\alpha} \ln \bar{P} \\ &= -\frac{1}{\alpha} \left( \mathbb{E} \left[ X_2 | Y_0, \bar{P}, X_1 \right]' \beta + \gamma \frac{\alpha}{1 + \alpha} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \right) + \frac{1}{\alpha} \ln \bar{P}, \end{aligned}$$

which together yield:

$$\begin{aligned} \mathbb{E} \left[ \frac{\partial \ln \lambda_\alpha(Y_2)}{\partial \alpha} | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] - \mathbb{E} \left[ \frac{\partial \ln \lambda_\alpha(Y_2)}{\partial \alpha} | Y_0, \bar{P}, X_1 \right] &= \frac{1}{\alpha} - \frac{1}{\alpha} \left( \mathbb{E} \left[ X_2 | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] - \mathbb{E} \left[ X_2 | Y_0, \bar{P}, X_1 \right] \right)' \beta \\ &\quad - \frac{\gamma}{\alpha} \left( \tilde{P}_1^{\frac{1}{\alpha}} - \frac{\alpha}{1 + \alpha} \right) \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} + \frac{1}{\alpha} \ln (1 - \tilde{P}_1). \end{aligned}$$

Collecting the results so far, we have that the first part of the efficient score, equation (76),

equal to:

$$\begin{aligned} \frac{2}{\alpha} + \frac{1}{\alpha} \ln \tilde{P}_1 + \frac{1}{\alpha} \ln(1 - \tilde{P}_1) - \frac{\gamma}{\alpha} \left( \tilde{P}_1^{\frac{1}{\alpha}} - \frac{\alpha}{1+\alpha} \right) \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \\ - \frac{1}{\alpha} \left( \mathbb{E} [X_2 | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [X_2 | Y_0, \bar{P}, X_1] \right)' \beta. \end{aligned}$$

To find the form of the second part of the efficient score we need to evaluate several additional conditional expectations. We begin with the pair of conditional expectations, using the expression for  $\frac{\partial \ln \Lambda_\alpha(Y_1)}{\partial \alpha}$  presented above,

$$\begin{aligned} & \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_1)}{\partial \alpha} \tilde{P}_1 \bar{P} \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2] | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] \\ &= \mathbb{E} \left[ \left( -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) + \frac{1}{\alpha} \ln \tilde{P}_1 + \frac{1}{\alpha} \ln \bar{P} \right) \tilde{P}_1 \bar{P} \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2] | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] \\ &= \left( -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) \tilde{P}_1 \bar{P} + \frac{1}{\alpha} \ln \tilde{P}_1 \tilde{P}_1 \bar{P} + \frac{1}{\alpha} \tilde{P}_1 \bar{P} \ln \bar{P} \right) \mathbb{E} [e^A | Y_0, \bar{P}, X_1], \end{aligned}$$

and also

$$\begin{aligned} & \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_1)}{\partial \alpha} \tilde{P}_1 \bar{P} \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2] | Y_0, \bar{P}, X_1 \right] \\ &= -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) \mathbb{E} [\tilde{P}_1 e^A | Y_0, \bar{P}, X_1] \bar{P} + \frac{1}{\alpha} \mathbb{E} [\ln \tilde{P}_1 \tilde{P}_1 e^A | Y_0, \bar{P}, X_1] \bar{P} \\ &+ \frac{1}{\alpha} \mathbb{E} [\tilde{P}_1 e^A | Y_0, \bar{P}, X_1] \bar{P} \ln \bar{P} \\ &= \left( -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) \frac{1}{2} \bar{P} - \frac{1}{\alpha} \frac{1}{4} \bar{P} + \frac{1}{\alpha} \frac{1}{2} \bar{P} \ln \bar{P} \right) \mathbb{E} [e^A | Y_0, \bar{P}, X_1]. \end{aligned}$$

Together this yields a difference of expectations equal to

$$\begin{aligned} & \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_1)}{\partial \alpha} \tilde{P}_1 \bar{P} \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2] | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] \\ & - \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_1)}{\partial \alpha} \tilde{P}_1 \bar{P} \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2] | Y_0, \bar{P}, X_1 \right] \\ &= \left( -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) \left( \tilde{P}_1 - \frac{1}{2} \right) + \frac{1}{\alpha} \left( \ln \tilde{P}_1 \tilde{P}_1 + \frac{1}{4} \right) + \frac{1}{\alpha} \left( \tilde{P}_1 - \frac{1}{2} \right) \ln \bar{P} \right) \\ & \quad \times \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1]. \end{aligned}$$

Next we evaluate, using the expression for  $\frac{\partial \ln \Lambda_\alpha(Y_2)}{\partial \alpha}$  given earlier,

$$\begin{aligned}
& \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_2)}{\partial \alpha} \times (1 - \tilde{P}_1) \bar{P} \mathbb{E} \left[ e^A | Y_0, \tilde{P}_1, \bar{P}, X^2 \right] | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] \\
&= \mathbb{E} \left[ \left( -\frac{1}{\alpha} \left( X'_2 \beta + \gamma \tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \right) + \frac{1}{\alpha} \ln(1 - \tilde{P}_1) + \frac{1}{\alpha} \ln \bar{P} \right) (1 - \tilde{P}_1) \bar{P} \right. \\
&\quad \times \mathbb{E} \left[ e^A | Y_0, \tilde{P}_1, \bar{P}, X^2 \right] | Y_0, \tilde{P}_1, \bar{P}, X_1 \left. \right] \\
&= -\frac{1}{\alpha} \mathbb{E} \left[ X_2 e^A | Y_0, \tilde{P}_1, \bar{P}, X_1 \right]' \beta (1 - \tilde{P}_1) \bar{P} \\
&\quad - \frac{\gamma}{\alpha} \tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} (1 - \tilde{P}_1) \bar{P} \mathbb{E} \left[ e^A | Y_0, \bar{P}, X_1 \right] \\
&\quad + \frac{1}{\alpha} \ln(1 - \tilde{P}_1) (1 - \tilde{P}_1) \bar{P} \mathbb{E} \left[ e^A | Y_0, \bar{P}, X_1 \right] \\
&\quad + \frac{1}{\alpha} (1 - \tilde{P}_1) \bar{P} \ln \bar{P} \mathbb{E} \left[ e^A | Y_0, \bar{P}, X_1 \right].
\end{aligned}$$

We also require:

$$\begin{aligned}
& \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_2)}{\partial \alpha} \times (1 - \tilde{P}_1) \bar{P} \mathbb{E} \left[ e^A | Y_0, \tilde{P}_1, \bar{P}, X^2 \right] | Y_0, \bar{P}, X_1 \right] \\
&= -\frac{1}{\alpha} \mathbb{E} \left[ (1 - \tilde{P}_1) X_2 e^A | Y_0, \bar{P}, X_1 \right]' \beta \bar{P} \\
&\quad - \frac{\gamma}{\alpha} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P} \mathbb{E} \left[ \tilde{P}_1^{\frac{1}{\alpha}} (1 - \tilde{P}_1) e^A | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] \\
&\quad + \frac{1}{\alpha} \bar{P} \mathbb{E} \left[ \ln(1 - \tilde{P}_1) (1 - \tilde{P}_1) e^A | Y_0, \bar{P}, X_1 \right] \\
&\quad + \frac{1}{\alpha} \bar{P} \ln \bar{P} \mathbb{E} \left[ (1 - \tilde{P}_1) e^A | Y_0, \bar{P}, X_1 \right],
\end{aligned}$$

which, after further simplification using Lemma D.2, equals

$$\begin{aligned}
& \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_2)}{\partial \alpha} \times (1 - \tilde{P}_1) \bar{P} \mathbb{E} \left[ e^A | Y_0, \tilde{P}_1, \bar{P}, X^2 \right] | Y_0, \bar{P}, X_1 \right] \\
&= -\frac{1}{\alpha} \mathbb{E} \left[ (1 - \tilde{P}_1) X_2 e^A | Y_0, \bar{P}, X_1 \right]' \beta \bar{P} \\
&\quad - \frac{\gamma}{\alpha} \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P} \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) \mathbb{E} \left[ e^A | Y_0, \bar{P}, X_1 \right] \\
&\quad - \frac{1}{\alpha} \frac{1}{4} \bar{P} \mathbb{E} \left[ e^A | Y_0, \bar{P}, X_1 \right] \\
&\quad + \frac{1}{\alpha} \frac{1}{2} \bar{P} \ln \bar{P} \mathbb{E} \left[ e^A | Y_0, \bar{P}, X_1 \right].
\end{aligned}$$

Putting these last two expectation evaluations together yields a difference of:

$$\begin{aligned}
& \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_2)}{\partial \alpha} \times (1 - \tilde{P}_1) \bar{P} \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2] | Y_0, \tilde{P}_1, \bar{P}, X_1 \right] \\
& - \mathbb{E} \left[ \frac{\partial \ln \Lambda_\alpha(Y_2)}{\partial \alpha} \times (1 - \tilde{P}_1) \bar{P} \mathbb{E} [e^A | Y_0, \tilde{P}_1, \bar{P}, X^2] | Y_0, \bar{P}, X_1 \right] \\
= & -\frac{1}{\alpha} \left( (1 - \tilde{P}_1) \mathbb{E} [X_2 e^A | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [(1 - \tilde{P}_1) X_2 e^A | Y_0, \bar{P}, X_1] \right)' \beta \bar{P} \\
& - \frac{\gamma}{\alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \left( \tilde{P}_1^{\frac{1}{\alpha}} (1 - \tilde{P}_1) - \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) \right) \bar{P}^{\frac{1}{\alpha}} \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( \ln(1 - \tilde{P}_1)(1 - \tilde{P}_1) + \frac{1}{4} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( (1 - \tilde{P}_1) - \frac{1}{2} \right) \bar{P} \ln \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1].
\end{aligned}$$

Hence the second component of the efficient score is given by (minus):

$$\begin{aligned}
& -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( \ln \tilde{P}_1 \tilde{P}_1 + \frac{1}{4} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( \tilde{P}_1 - \frac{1}{2} \right) \ln \bar{P} \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( (1 - \tilde{P}_1) - \frac{1}{2} \right) \bar{P} \ln \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( \ln(1 - \tilde{P}_1)(1 - \tilde{P}_1) + \frac{1}{4} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& - \frac{\gamma}{\alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \left( \tilde{P}_1^{\frac{1}{\alpha}} (1 - \tilde{P}_1) - \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) \right) \bar{P}^{\frac{1}{\alpha}} \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& - \frac{1}{\alpha} \left( (1 - \tilde{P}_1) \mathbb{E} [X_2 e^A | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [(1 - \tilde{P}_1) X_2 e^A | Y_0, \bar{P}, X_1] \right)' \beta \bar{P},
\end{aligned}$$

which, after some manipulation, simplifies to

$$\begin{aligned}
& -\frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( \ln \tilde{P}_1 \tilde{P}_1 + \frac{1}{4} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( \ln (1 - \tilde{P}_1) (1 - \tilde{P}_1) + \frac{1}{4} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& - \frac{\gamma}{\alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \left( \tilde{P}_1^{\frac{1}{\alpha}} (1 - \tilde{P}_1) - \left( \frac{\alpha}{1+\alpha} - \frac{\alpha}{1+2\alpha} \right) \right) \bar{P}^{\frac{1}{\alpha}} \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& - \frac{1}{\alpha} \left( (1 - \tilde{P}_1) \mathbb{E} [X_2 e^A | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [(1 - \tilde{P}_1) X_2 e^A | Y_0, \bar{P}, X_1] \right)' \beta \bar{P}.
\end{aligned}$$

We want minus of the former to form the efficient score, so (and also collecting terms)

$$\begin{aligned}
& + \frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& - \frac{1}{\alpha} \left( \ln \tilde{P}_1 \tilde{P}_1 + \frac{1}{4} + \ln (1 - \tilde{P}_1) (1 - \tilde{P}_1) + \frac{1}{4} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{\gamma}{\alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \left( \tilde{P}_1^{\frac{1}{\alpha}} (1 - \tilde{P}_1) - \left( \frac{\alpha}{1+\alpha} - \frac{\alpha}{1+2\alpha} \right) \right) \bar{P}^{\frac{1}{\alpha}} \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( (1 - \tilde{P}_1) \mathbb{E} [X_2 e^A | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [(1 - \tilde{P}_1) X_2 e^A | Y_0, \bar{P}, X_1] \right)' \beta \bar{P}.
\end{aligned}$$

Putting everything together, the efficient score for  $\alpha$  equals:

$$\begin{aligned}
\phi_{\theta}^{\text{eff}, \alpha}(Y^2, X^2) = & \frac{2}{\alpha} + \frac{1}{\alpha} \ln \tilde{P}_1 + \frac{1}{\alpha} \ln (1 - \tilde{P}_1) - \frac{\gamma}{\alpha} \left( \tilde{P}_1^{\frac{1}{\alpha}} - \frac{\alpha}{1+\alpha} \right) \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \\
& - \frac{1}{\alpha} \left( \mathbb{E} [X_2 | Y_0, X_1, \tilde{P}_1, \bar{P}] - \mathbb{E} [X_2 | Y_0, X_1, \bar{P}] \right)' \beta \\
& + \frac{1}{\alpha} (X'_1 \beta + \gamma Y_0) \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& - \frac{1}{\alpha} \left( \ln \tilde{P}_1 \tilde{P}_1 + \frac{1}{4} + \ln (1 - \tilde{P}_1) (1 - \tilde{P}_1) + \frac{1}{4} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{\gamma}{\alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \left( \tilde{P}_1^{\frac{1}{\alpha}} (1 - \tilde{P}_1) - \left( \frac{\alpha}{1+\alpha} - \frac{\alpha}{1+2\alpha} \right) \right) \bar{P}^{\frac{1}{\alpha}} \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\
& + \frac{1}{\alpha} \left( (1 - \tilde{P}_1) \mathbb{E} [X_2 e^A | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E} [(1 - \tilde{P}_1) X_2 e^A | Y_0, \bar{P}, X_1] \right)' \beta \bar{P}.
\end{aligned}$$

Close inspection of the expression above indicates that it includes linear combinations of the efficient scores for  $\beta$  and  $\gamma$  as components. This observation, as well as simplification and

rearrangement, gives a final expression of:

$$\begin{aligned}\phi_{\theta}^{\text{eff},\alpha}(Y^2, X^2) = & \frac{1}{\alpha} \left( 2 + \ln \tilde{P}_1 + \ln(1 - \tilde{P}_1) \right) \\ & - \frac{1}{\alpha} \left( \tilde{P}_1 \ln \tilde{P}_1 + (1 - \tilde{P}_1) \ln(1 - \tilde{P}_1) + \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X_1] \\ & - \phi_{\theta}^{\text{eff},\beta}(Y^2, X^2)' \frac{\beta}{\alpha} - \phi_{\theta}^{\text{eff},\gamma}(Y^2, X^2) \frac{\gamma}{\alpha}.\end{aligned}\quad (77)$$

Like its counterpart for  $\gamma$  this expression does not appear in the main text.

### D.2.2 MPH efficient score without feedback (i.e., under strict exogeneity)

Hahn (1994) derived the SEB for the MPH hazards model with  $T = 2$  and strictly exogenous regressors. His analysis did not include lagged duration dependence, as ours does. For completeness, we sketch the derivation of the efficient scores for  $\alpha$ ,  $\beta$  and  $\gamma$  under strict exogeneity here. Details can be filled in along the lines of our derivation for the scores with feedback and/or by studying the rigorous analysis in Hahn (1994).

By direct analogy, the efficient score when  $T = 2$  under strict exogeneity equals

$$\phi_{\theta}^{\text{eff,SE}}(Y^2, X^2) = \mathbb{E} [S^{\theta}(Y^2, X^2) | Y_0, \tilde{P}_1, \bar{P}, X^2] - \mathbb{E} [S^{\theta}(Y^2, X^2) | Y_0, \bar{P}, X^2].$$

It follows that the expressions for the scores are the same as those derived above under feedback except that each expectation now conditions on all “leads and lags” of the strictly exogenous covariates.

This yields an efficient score for  $\beta$  of

$$\phi_{\theta}^{\text{eff,SE},\beta}(Y^2, X^2) = (X_2 - X_1) \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X^2]. \quad (78)$$

This expression is identical to the one found by Hahn (1994).

Maintaining the Weibull baseline hazard assumption, the efficient score for  $\gamma$  equals:

$$\begin{aligned}\phi_{\theta}^{\text{eff,SE},\gamma}(Y^2, X^2) = & Y_1 - \frac{\alpha}{1 + \alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P}^{\frac{1}{\alpha}} \\ & - \left( Y_0 \left( \tilde{P}_1 - \frac{1}{2} \right) + \left( Y_1(1 - \tilde{P}_1) - \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \bar{P}^{\frac{1}{\alpha}} \right) \right) \\ & \times \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X^2].\end{aligned}\quad (79)$$

The analysis of Hahn (1994) did not accomodate lagged duration dependence. The expression above is therefore new.

Finally, the efficient score for  $\alpha$  equals:

$$\begin{aligned}\phi_{\theta}^{\text{eff}, \text{SE}, \alpha}(Y^2, X^2) &= \frac{1}{\alpha} \left( 2 + \ln \tilde{P}_1 + \ln \left( 1 - \tilde{P}_1 \right) \right) \\ &\quad - \frac{1}{\alpha} \left( \tilde{P}_1 \ln \tilde{P}_1 + (1 - \tilde{P}_1) \ln \left( 1 - \tilde{P}_1 \right) + \frac{1}{2} \right) \bar{P} \mathbb{E} [e^A | Y_0, \bar{P}, X^2] \\ &\quad - \phi_{\theta}^{\text{eff}, \text{SE}, \beta}(Y^2, X^2)' \frac{\beta}{\alpha} - \phi_{\theta}^{\text{eff}, \text{SE}, \gamma}(Y^2, X^2)' \frac{\gamma}{\alpha}.\end{aligned}\tag{80}$$

This expression coincides exactly with the one given in Lemma 3 of Hahn (1994) except that his formulation maintains the additional *a priori* restriction  $\gamma = 0$ . Setting  $\gamma = 0$  in the expression above yields Hahn's expression.

### D.2.3 MPH efficient score for average effects

Lemma D.3 can also be used to derive expressions for the efficient moment function of average effects  $\mu(\theta, \omega)$ .

**Average structural hazard.** Consider first the ASH  $\bar{\lambda}(y_t | x_t, y_{t-1})$  defined in (22). In the main text, we showed that  $\varphi_{\theta}(Y^T, X^T) = \lambda_{\alpha}(y_t) e^{x'_t \beta + \gamma y_{t-1}} \frac{T-1}{\bar{P}}$  is an identifying FHR moment function for  $\bar{\lambda}(y_t | x_t, y_{t-1})$ . Applying Lemma D.3 yields the projection

$$\Pi(\varphi_{\theta}(Y^T, X^T) | \mathcal{T}_{\theta, \omega, L}^{\perp}) = \sum_{t=1}^{T-1} \varphi_{\theta, t}^{\perp}(Y_0, \tilde{P}^t, \bar{P}, X^t) = 0,$$

since, for all  $t \in \{1, \dots, T-1\}$ ,

$$\begin{aligned}\varphi_{\theta, t}^{\perp}(Y_0, \tilde{P}^t, \bar{P}, X^t) &= \mathbb{E} [\varphi_{\theta}(Y^T, X^T) | Y_0, \tilde{P}^t, \bar{P}, X^t] - \mathbb{E} [\varphi_{\theta}(Y^T, X^T) | Y_0, \tilde{P}^{t-1}, \bar{P}, X^t] \\ &= \lambda_{\alpha}(y_t) e^{x'_t \beta + \gamma y_{t-1}} \frac{T-1}{\bar{P}} - \lambda_{\alpha}(y_t) e^{x'_t \beta + \gamma y_{t-1}} \frac{T-1}{\bar{P}} \\ &= 0.\end{aligned}$$

This verifies the claim in Section 5. Recalling from Section 4.1 that the efficient moment function is given by  $\varphi_{\theta, \omega}^{\text{eff}}(Y^T, X^T) = \varphi_{\theta}(Y^T, X^T) - \Pi(\varphi_{\theta}(Y^T, X^T) | \mathcal{T}_{\theta, \omega, L}^{\perp})$ , we conclude that the efficient moment function for the ASH is

$$\varphi_{\theta, \omega}^{\text{eff}}(Y^T, X^T) = \lambda_{\alpha}(y_t) e^{x'_t \beta + \gamma y_{t-1}} \frac{T-1}{\bar{P}}.$$

**Average structural function.** As a second example, consider the ASF  $\mu(x_t, y_{t-1})$  defined

in (23). Under a Weibull baseline hazard, it takes the form

$$\mu(x_t, y_{t-1}) = \Gamma\left(1 + \frac{1}{\alpha}\right) \exp\left(-x'_t \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} y_{t-1}\right) \mathbb{E}\left[\exp\left(-\frac{a}{\alpha}\right)\right].$$

In view of (28), one candidate identifying moment function is

$$\varphi_\theta(Y^T, X^T) = \exp\left(-x'_t \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} y_{t-1}\right) \frac{\Gamma\left(1 + \frac{1}{\alpha}\right) \Gamma(T)}{\Gamma(T + \frac{1}{\alpha})} \bar{P}^{\frac{1}{\alpha}}.$$

Applying Lemma D.3, we again get  $\Pi(\varphi_\theta(Y^T, X^T) | \mathcal{T}_{\theta, \omega, L}^\perp) = 0$ , which implies that the efficient moment function for the ASF is given by

$$\varphi_{\theta, \omega}^{\text{eff}}(Y^T, X^T) = \exp\left(-x'_t \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} y_{t-1}\right) \frac{\Gamma\left(1 + \frac{1}{\alpha}\right) \Gamma(T)}{\Gamma(T + \frac{1}{\alpha})} \bar{P}^{\frac{1}{\alpha}}. \quad (81)$$

Alternatively, based on (17), one could for example consider the FHR moment function

$$\varphi_{\theta_0}^2(Y^T, X^T) = \Gamma\left(1 + \frac{1}{\alpha}\right) \exp\left(-x'_t \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} y_{t-1}\right) P_1^{\frac{1}{\alpha}},$$

noting that  $\mathbb{E}\left[P_1^{\frac{1}{\alpha}} | Y_0, X_1, A\right] = \mathbb{E}\left[\exp\left(-\frac{a}{\alpha}\right)\right]$ . For this choice, the identity  $P_1 = \tilde{P}_1 \bar{P}$  and an application of Lemma D.3 yields

$$\begin{aligned} \Pi(\varphi_\theta^2(Y^T, X^T) | \mathcal{T}_{\theta, \omega, L}^\perp) &= \Gamma\left(1 + \frac{1}{\alpha}\right) \exp\left(-x'_t \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} y_{t-1}\right) \left(\tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} - \mathbb{E}\left[\tilde{P}_1^{\frac{1}{\alpha}} | Y_0, \bar{P}, X^1\right] \bar{P}^{\frac{1}{\alpha}}\right) \\ &= \Gamma\left(1 + \frac{1}{\alpha}\right) \exp\left(-x'_t \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} y_{t-1}\right) \left(\tilde{P}_1^{\frac{1}{\alpha}} \bar{P}^{\frac{1}{\alpha}} - \frac{\Gamma\left(1 + \frac{1}{\alpha}\right) \Gamma(T)}{\Gamma(T + \frac{1}{\alpha})} \bar{P}^{\frac{1}{\alpha}}\right) \\ &= \varphi_\theta^2(Y^T, X^T) - \Gamma\left(1 + \frac{1}{\alpha}\right) \exp\left(-x'_t \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} y_{t-1}\right) \frac{\Gamma\left(1 + \frac{1}{\alpha}\right) \Gamma(T)}{\Gamma(T + \frac{1}{\alpha})} \bar{P}^{\frac{1}{\alpha}}, \end{aligned}$$

where the second equality leverages the distributional properties reported in Lemma (D.2). The projection residual would then deliver the same expression of the efficient moment function as (81). This aligns with the result discussed in Section 4.1 that the efficient moment function for average effects is invariant to the specific choice of  $\varphi_\theta$ .

### D.3 Detailed calculations

**Derivation of equation (66):** Using the inverse mapping defined in Appendix D we can write the (determinant of the) Jacobian of the mapping from  $\mathbf{V}$  back into  $\mathbf{P}$  as:

$$J^{(T)}(V_1, \dots, V_T) = \begin{vmatrix} V_T & 0 & 0 & \cdots & 0 & V_1 \\ -V_2 V_T & (1-V_1)V_T & 0 & \cdots & \cdots & (1-V_1)V_2 \\ -(1-V_2)V_3 V_T & -(1-V_1)V_3 V_T & (1-V_1)(1-V_2)V_T & 0 & \cdots & (1-V_1)(1-V_2)V_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\prod_{s=2}^{T-2}(1-V_s)V_{T-1}V_T & \cdots & \cdots & \cdots & \prod_{s=1}^{T-2}(1-V_s)V_T & \prod_{s=1}^{T-2}(1-V_s)V_{T-1} \\ -\prod_{s=2}^{T-1}(1-V_s)V_T & \cdots & \cdots & -\prod_{s=1}^{T-2}(1-V_s)V_T & \prod_{s=1}^{T-1}(1-V_s) & \prod_{s=1}^{T-1}(1-V_s) \end{vmatrix}$$

A Laplace (co-factor) expansion along the first row gives

$$J^{(T)}(V_1, \dots, V_T) = V_T \begin{vmatrix} (1-V_1)V_T & 0 & 0 & \cdots & (1-V_1)V_2 \\ -(1-V_1)V_3 V_T & (1-V_1)(1-V_2)V_T & 0 & \cdots & (1-V_1)(1-V_2)V_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -(1-V_1)\prod_{s=3}^{T-2}(1-V_s)V_{T-1}V_T & \cdots & \prod_{s=1}^{T-2}(1-V_s)V_T & \prod_{s=1}^{T-2}(1-V_s)V_{T-1} \\ -(1-V_1)\prod_{s=3}^{T-1}(1-V_s)V_T & \cdots & -\prod_{s=1}^{T-2}(1-V_s)V_T & \prod_{s=1}^{T-1}(1-V_s) & \prod_{s=1}^{T-1}(1-V_s) \end{vmatrix} + (-1)^{T+1} V_1 \begin{vmatrix} -V_2 V_T & (1-V_1)V_T & 0 & 0 & V_2 \\ -(1-V_2)V_3 V_T & -(1-V_1)V_3 V_T & (1-V_1)(1-V_2)V_T & 0 & (1-V_2)V_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\prod_{s=2}^{T-2}(1-V_s)V_{T-1}V_T & \cdots & \cdots & \prod_{s=1}^{T-2}(1-V_s)V_T & \prod_{s=1}^{T-2}(1-V_s)V_T \\ -\prod_{s=2}^{T-1}(1-V_s)V_T & \cdots & \cdots & -\prod_{s=1}^{T-2}(1-V_s)V_T & -\prod_{s=1}^{T-2}(1-V_s)V_T \end{vmatrix}$$

and further factorizing common factors across columns yields

$$J^{(T)}(V_1, \dots, V_T) = V_T(1-V_1)^{T-1} \begin{vmatrix} V_T & 0 & 0 & \cdots & V_2 \\ -V_3 V_T & (1-V_2)V_T & 0 & \cdots & (1-V_2)V_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\prod_{s=3}^{T-2}(1-V_s)V_{T-1}V_T & \cdots & \prod_{s=2}^{T-2}(1-V_s)V_T & \prod_{s=2}^{T-2}(1-V_s)V_{T-1} \\ -\prod_{s=3}^{T-1}(1-V_s)V_T & \cdots & -\prod_{s=2}^{T-2}(1-V_s)V_T & \prod_{s=2}^{T-1}(1-V_s) & \prod_{s=2}^{T-1}(1-V_s) \end{vmatrix} + (-1)^T V_1(1-V_1)^{T-2} V_T \begin{vmatrix} V_2 & V_T & 0 & 0 \\ (1-V_2)V_3 & -V_3 V_T & (1-V_2)V_T & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \prod_{s=2}^{T-2}(1-V_s)V_{T-1} & \cdots & \cdots & \prod_{s=2}^{T-2}(1-V_s)V_T \\ \prod_{s=2}^{T-1}(1-V_s) & \cdots & \cdots & -\prod_{s=2}^{T-2}(1-V_s)V_T \end{vmatrix}.$$

Observe that the matrices involved in the first and second determinants have the same set of columns. More specifically, the second matrix can be obtained by swapping  $\lfloor \frac{T}{2} \rfloor$  pairs of columns in the first matrix, implying that the determinant of the second matrix is  $(-1)^{\lfloor \frac{T}{2} \rfloor}$  that of the first matrix.

Next observe that the first determinant has exactly the same structure as  $J^{(T)}(V_1, \dots, V_T)$ , except that it involves  $T-1$  variables instead of  $T$  variables, namely  $(V_2, \dots, V_T)$  instead of  $(V_1, \dots, V_T)$ . More precisely, this determinant coincides with  $J^{(T-1)}(V_2, \dots, V_T)$ .

Together, these observations imply that:

$$\begin{aligned}
J^{(T)}(V_1, \dots, V_T) &= V_T(1 - V_1)^{T-1} J^{(T-1)}(V_2, \dots, V_T) + (-1)^{T+\lfloor \frac{T}{2} \rfloor} V_1(1 - V_1)^{T-2} V_T J^{(T-1)}(V_2, \dots, V_T) \\
&= (1 - V_1)^{T-2} V_T J^{(T-1)}(V_2, \dots, V_T) \\
&= (1 - V_1)^{T-2} V_T(1 - V_2)^{T-3} V_T J^{(T-2)}(V_3, \dots, V_T) \\
&= (1 - V_1)^{T-2} V_T(1 - V_2)^{T-3} V_T \times \dots \times (1 - V_{T-2}) V_T J^{(2)}(V_{T-1}, V_T) \\
&= (1 - V_1)^{T-2} V_T(1 - V_2)^{T-3} V_T \times \dots \times (1 - V_{T-2}) V_T V_T \\
&= V_T^{T-1} \prod_{s=1}^{T-2} (1 - V_s)^{T-(s+1)},
\end{aligned}$$

which coincides with (66) as required.

**Derivation of equation (67)** The change-of-variables formula, in conjunction with the determinant given in (66), yields

$$\begin{aligned}
f_V(V_1, V_2, \dots, V_T) &= f_{P_1}(V_1 V_T) \prod_{t=2}^{T-1} f_{P_t} \left( \prod_{s=1}^{t-1} (1 - V_s) V_t V_T \right) f_{P_T} \left( \prod_{s=1}^{t-1} (1 - V_s) V_T \right) \left| \det \left[ \frac{dG^{-1}(\mathbf{V})}{d\mathbf{V}} \right] \right| \\
&= \frac{\beta^{\alpha_1}}{\Gamma(\alpha_1)} (V_1 V_T)^{\alpha_1-1} e^{-\beta V_1 V_T} \\
&\quad \times \prod_{t=2}^{T-1} \frac{\beta^{\alpha_t}}{\Gamma(\alpha_t)} \left( \prod_{s=1}^{t-1} (1 - V_s) V_t V_T \right)^{\alpha_t-1} e^{-\beta \prod_{s=1}^{t-1} (1 - V_s) V_t V_T} \\
&\quad \times \frac{\beta^{\alpha_T}}{\Gamma(\alpha_T)} \left( \prod_{s=1}^{T-1} (1 - V_s) V_T \right)^{\alpha_T-1} e^{-\beta \prod_{s=1}^{T-1} (1 - V_s) V_T} \\
&\quad \times V_T^{T-1} \prod_{s=1}^{T-2} (1 - V_s)^{T-(s+1)} \\
&= \frac{\beta^{\sum_{t=1}^T \alpha_t}}{\Gamma(\alpha_T)} V_T^{\sum_{t=1}^T \alpha_t-1} e^{-\beta V_T} \\
&\quad \times \frac{1}{\Gamma(\alpha_1)} V_1^{\alpha_1-1} (1 - V_1)^{\sum_{t=2}^T \alpha_t - (T-1) + (T-(1+1))} \\
&\quad \times \frac{1}{\Gamma(\alpha_2)} V_2^{\alpha_2-1} (1 - V_2)^{\sum_{t=3}^T \alpha_t - (T-2) + (T-(2+1))} \\
&\quad \vdots \\
&\quad \times \frac{1}{\Gamma(\alpha_{T-1})} V_{T-1}^{\alpha_{T-1}-1} (1 - V_{T-1})^{\sum_{t=T-1}^T \alpha_t - (T-(T-1))} \\
&= \frac{\beta^{\sum_{t=1}^T \alpha_t}}{\Gamma(\sum_{t=1}^T \alpha_t)} V_T^{\sum_{t=1}^T \alpha_t-1} e^{-\beta V_T} \times \prod_{t=1}^{T-1} \frac{\Gamma(\sum_{s=t}^T \alpha_s)}{\Gamma(\alpha_t) \Gamma(\sum_{s=t+1}^T \alpha_s)} V_t^{\alpha_t-1} (1 - V_t)^{\sum_{s=t+1}^T \alpha_s - 1},
\end{aligned}$$

where the simplification on the penultimate equality follows from:

$$V_T = \sum_{t=1}^T P_t = V_1 V_T + \prod_{s=1}^{t-1} (1 - V_s) V_t V_T + \prod_{s=1}^{T-1} (1 - V_s) V_T.$$

This corresponds to equation (67) as needed.

## E Additional details for the numerical experiments

In this section we again omit the dependence on  $\theta$  in  $P_\theta$  and related quantities.

**Lemma E.1.** *Consider the MPH model defined by Example 1 with  $T = 2$ . If there is no feedback, then (i)  $\bar{P}, X_2, \tilde{P}_1$  are independent conditional on  $Y_0, X_1, A$ , and (ii)  $\tilde{P}_1$  is independent of  $X_2, \bar{P}, A, Y_0, X_1$ . If in addition  $X_2$  is independent of  $A$  conditional on  $Y_0, X_1$ , then (iii)  $\bar{P}, X_2, \tilde{P}_1$  are independent conditional on  $Y_0, X_1$*

*Proof.* If there is no feedback, (6) and Lemma D.2 imply that the joint density of  $(\bar{P}, X_2, \tilde{P}_1)$  conditional on  $(Y_0, X_1, A)$  is given by

$$f(\bar{p}, x_2, \tilde{p}_1 | y_0, x_1, a) = \bar{p} \exp(2a - \bar{p}e^a) \mathbb{1}\{\bar{p} > 0\} g(x_2 | y_0, x_1, a) \mathbb{1}\{\tilde{p}_1 \in (0, 1)\},$$

which establishes (i). This also proves that  $\tilde{P}_1 \perp A, Y_0, X_1$  and hence (ii). If  $X_2$  is independent of  $A$  conditional on  $Y_0, X_1$ ,  $g(x_2 | y_0, x_1, a) = g(x_2 | y_0, x_1)$  whereupon the form of the above joint density implies (iii).  $\square$

**Locally efficient score.** The working model in our numerical experiment is  $\tilde{\omega} = (\tilde{g}, \tilde{\pi})$  where  $\tilde{g}$  is a Bernoulli( $p$ ) and  $\tilde{\pi}(v) = \frac{1}{v} \mathbb{1}\{v > 0\}$  where  $V = \exp(A)$ . Since  $\bar{P} | Y_0, X_1, V \sim \text{Gamma}(T, V)$  by Lemma (D.2), it follows from Bayes rule that  $V | Y_0, \bar{P}, X_1 \sim \text{Gamma}(T, \bar{P})$ . Therefore,  $\mathbb{E}_{\theta, \tilde{\omega}} [V | Y_0, \bar{P}, X_1] = \frac{T}{\bar{P}}$ . In view of (73), the locally efficient score for  $\beta$  under  $\tilde{\omega}$  is

$$\begin{aligned} \tilde{\phi}_{\theta, \tilde{\omega}}^{\text{eff}, \beta}(Y^2, X^2) &= -X_1 \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} \mathbb{E}_{\theta, \tilde{\omega}} [V | Y_0, \bar{P}, X_1] \\ &\quad + \mathbb{E}_{\theta, \tilde{\omega}} [X_2 | Y_0, \tilde{P}_1, \bar{P}, X_1] - \mathbb{E}_{\theta, \tilde{\omega}} [X_2(1 - \tilde{P}_1) \bar{P} V | Y_0, \tilde{P}_1, \bar{P}, X_1] \\ &\quad - \mathbb{E}_{\theta, \tilde{\omega}} [X_2 | Y_0, \bar{P}, X_1] + \mathbb{E}_{\theta, \tilde{\omega}} [X_2(1 - \tilde{P}_1) \bar{P} V | Y_0, \bar{P}, X_1] \end{aligned}$$

$$\begin{aligned}
&= -X_1 \left( \tilde{P}_1 - \frac{1}{2} \right) \overline{P} \mathbb{E}_{\theta, \tilde{\omega}} [V | Y_0, \overline{P}, X_1] \\
&\quad + \mathbb{E}_{\theta, \tilde{\omega}} [X_2 | Y_0, X_1] - (1 - \tilde{P}_1) \overline{P} \mathbb{E}_{\theta, \tilde{\omega}} [X_2 | Y_0, X_1] \mathbb{E}_{\theta, \tilde{\omega}} [V | Y_0, \overline{P}, X_1] \\
&\quad - \mathbb{E}_{\theta, \tilde{\omega}} [X_2 | Y_0, X_1] + \overline{P} \mathbb{E}_{\theta, \tilde{\omega}} [X_2 | Y_0, X_1] \mathbb{E}_{\theta, \tilde{\omega}} \left[ (1 - \tilde{P}_1) \right] \mathbb{E}_{\theta, \tilde{\omega}} [V | Y_0, \overline{P}, X_1] \\
&= (p - X_1) \left( \tilde{P}_1 - \frac{1}{2} \right) T,
\end{aligned} \tag{82}$$

where the second equality follows from the implications of Lemma E.1, and the third equality uses  $\mathbb{E}_{\theta, \tilde{\omega}} [V | Y_0, \overline{P}, X_1] = \frac{T}{\overline{P}}$ ,  $\mathbb{E}_{\theta, \tilde{\omega}} [\tilde{P}_1] = \frac{1}{2}$  by Lemma D.2, and the definition of  $\tilde{g}$ . For the remaining components of the efficient score, we use once more,  $\mathbb{E}_{\theta, \tilde{\omega}} [V | Y_0, \overline{P}, X_1] = \frac{T}{\overline{P}}$  and obtain

$$\begin{aligned}
\tilde{\phi}_{\theta, \tilde{\omega}}^{\text{eff}, \gamma}(Y^2, X^2) &= Y_1 - \frac{\alpha}{1 + \alpha} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \overline{P}^{\frac{1}{\alpha}} \\
&\quad - \left( Y_0 \left( \tilde{P}_1 - \frac{1}{2} \right) + \left( Y_1 (1 - \tilde{P}_1) - \left( \frac{\alpha}{1 + \alpha} - \frac{\alpha}{1 + 2\alpha} \right) e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0} \overline{P}^{\frac{1}{\alpha}} \right) \right) T
\end{aligned} \tag{83}$$

from (75), and finally

$$\begin{aligned}
\tilde{\phi}_{\theta, \tilde{\omega}}^{\text{eff}, \alpha}(Y^2, X^2) &= \frac{1}{\alpha} \left( 2 + \ln \tilde{P}_1 + \ln (1 - \tilde{P}_1) \right) \\
&\quad - \frac{1}{\alpha} \left( \tilde{P}_1 \ln \tilde{P}_1 + (1 - \tilde{P}_1) \ln (1 - \tilde{P}_1) + \frac{1}{2} \right) T \\
&\quad - \tilde{\phi}_{\beta, \tilde{\omega}}^{\text{eff}}(Y^2, X^2)' \frac{\beta}{\alpha} - \tilde{\phi}_{\gamma, \tilde{\omega}}^{\text{eff}}(Y^2, X^2) \frac{\gamma}{\alpha}
\end{aligned} \tag{84}$$

from (77).

**Implementation details.** The asymptotic standard errors reported in Table 1 are obtained via Monte Carlo integration using  $N = 1,000,000$  simulation draws from the DGP of Experiment (A). As a benchmark, we computed the square root of the diagonal elements of  $\mathbb{E} [\phi_{\theta_0}^{\text{eff}, \text{SE}}(Y_i^2, X_i^2) \phi_{\theta_0}^{\text{eff}, \text{SE}}(Y_i^2, X_i^2)' ]^{-1}$ , which corresponds to the semiparametric efficiency bound under strict exogeneity. This quantity serves as a reference for the remaining estimators, which explains the normalization in the first row. Similarly, the second row displays the square root of the (normalized) diagonal elements of  $\mathbb{E} [\phi_{\theta_0}^{\text{eff}}(Y_i^2, X_i^2) \phi_{\theta_0}^{\text{eff}}(Y_i^2, X_i^2)' ]^{-1}$ , i.e the semiparametric efficiency bound with feedback. For the locally efficient estimator and the simple moments approach, we report the (normalized) asymptotic standard errors using the method-of-moments variance formula:  $H^{-1} V H^{-1}$ . In the case of the

locally efficient estimator, the matrices are defined as  $H = \mathbb{E} \left[ \frac{\partial \tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y_i^2, X_i^2)}{\partial \theta} \right]$  and  $V = \mathbb{E} \left[ \tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y_i^2, X_i^2) \tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y_i^2, X_i^2)' \right]$ . For the simple moment-based estimator, we use the same expressions, replacing  $\tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y_i^2, X_i^2)$  with  $\phi_{\theta_0}(Y_i^2, X_i^2)$ .

In Table 2, we follow an analogous procedure with  $N = 1,000,000$  simulation draws from the DGP of Experiment (B) to evaluate population expectations. The main difference is that we now treat the square root of the diagonal elements of  $\mathbb{E} [\phi_{\theta_0}^{\text{eff}}(Y_i^2, X_i^2) \phi_{\theta_0}^{\text{eff}}(Y_i^2, X_i^2)]^{-1}$  as the benchmark for efficiency comparisons. For the ASH, we report the corresponding asymptotic standard errors for the method of moments estimator

$$\hat{\lambda}(y_t | x_t, y_{t-1}; \hat{\theta}) = \frac{1}{N} \sum_{i=1}^N \varphi_{\hat{\theta}}(Y_i^2, X_i^2),$$

where  $\varphi_{\theta}(Y_i^2, X_i^2) = \lambda_{\alpha}(y_t) e^{x_t' \beta + \gamma y_{t-1} \frac{T-1}{P_i}}$ . The benchmark in this case is the square root of the covariance of the efficient influence function for the ASH:

$$\varphi_{\theta_0}(Y_i^2, X_i^2) + \mathbb{E} \left[ \frac{\partial \varphi_{\theta_0}(Y_i^2, X_i^2)}{\partial \theta} \right] \mathbb{E} [\phi_{\theta_0}^{\text{eff}}(Y_i^2, X_i^2) \phi_{\theta_0}^{\text{eff}}(Y_i^2, X_i^2)']^{-1} \phi_{\theta_0}^{\text{eff}}(Y_i^2, X_i^2).$$

We apply the same strategy for the remaining two estimators, modifying the influence function accordingly. Specifically, for the estimator based on working models, we use:

$$\varphi_{\theta_0}(Y_i^2, X_i^2) - \mathbb{E} \left[ \frac{\partial \varphi_{\theta_0}(Y_i^2, X_i^2)}{\partial \theta} \right] \mathbb{E} \left[ \frac{\partial \tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y_i^2, X_i^2)}{\partial \theta} \right]^{-1} \tilde{\phi}_{\theta_0, \tilde{\omega}}^{\text{eff}}(Y_i^2, X_i^2),$$

and for the method based on simple moments:

$$\varphi_{\theta_0}(Y_i^2, X_i^2) - \mathbb{E} \left[ \frac{\partial \varphi_{\theta_0}(Y_i^2, X_i^2)}{\partial \theta} \right] \mathbb{E} \left[ \frac{\partial \phi_{\theta_0}(Y_i^2, X_i^2)}{\partial \theta} \right]^{-1} \phi_{\theta_0}(Y_i^2, X_i^2).$$

Finally, in both experiments, we evaluate  $p = \mathbb{E}[X_2]$  via Monte Carlo integration to compute the locally efficient score for  $\beta$  in equation (82).

**Computation of efficient scores.** In the absence of feedback, Lemma E.1 and Lemma D.2 imply that (73) simplifies to

$$\phi_{\theta}^{\text{eff}, \beta}(Y^2, X^2) = \left( \tilde{P}_1 - \frac{1}{2} \right) \bar{P} (\mathbb{E}[X_2 V | Y_0, \bar{P}, X_1] - X_1 \mathbb{E}[V | Y_0, \bar{P}, X_1]) \quad (85)$$

where  $V = \exp(A)$ . As a result, the efficient score for  $\alpha$  in (77) also simplifies. On the other

hand, the efficient score for  $\gamma$  in (75) remains unchanged. To compute the efficient score, we must evaluate two quantities:  $\mathbb{E}[X_2V|Y_0, \bar{P}, X_1]$  and  $\mathbb{E}[V|Y_0, \bar{P}, X_1]$ . In both numerical experiments  $V \sim \text{Gamma}(\kappa_0, \lambda_0)$  where  $\kappa_0 = \lambda_0 = 5$ . Given that  $\bar{P}|Y_0, X_1, V \sim \text{Gamma}(T, V)$  by Lemma D.2, Bayes rule implies that  $V|Y_0, \bar{P}, X_1 \sim \text{Gamma}(T + \kappa_0, \bar{P} + \lambda_0)$  and hence  $\mathbb{E}[V|Y_0, \bar{P}, X_1] = \frac{T + \kappa_0}{\bar{P} + \lambda_0}$ . Furthermore, the success probability for  $X_2$  in Experiment (A) is  $p(Y_0, Y_1, X_1, V) = 1 - \exp(-\tau_A(Y_0, X_1, Y_1)V)$  with  $\tau_A(Y_0, X_1, Y_1) = Y_0 + X_1$ . Using this specification, we derive:

$$\mathbb{E}[X_2V|Y_0, \bar{P}, X_1] = \frac{T + \kappa_0}{(\lambda_0 + \bar{P})} - (T + \kappa_0) \frac{(\lambda_0 + \bar{P})^{T + \kappa_0}}{(\lambda_0 + \bar{P} + \tau_A(Y_0, X_1, Y_1))^{T + \kappa_0 + 1}},$$

and

$$\mathbb{E}[e^A|Y_0, X_1, X_2, \bar{P}] = X_2 \frac{w_1}{w_1 - w_2} \frac{T + \kappa_0}{(\lambda_0 + \bar{P})} + \left( (1 - X_2) - X_2 \frac{w_2}{w_1 - w_2} \right) \frac{T + \kappa_0}{(\lambda_0 + \bar{P} + \tau_A(Y_0, X_1, Y_1))},$$

for  $w_1 = \frac{\bar{P}^{T-1} \lambda_0^{\kappa_0}}{\Gamma(T)\Gamma(\kappa_0)} \frac{\Gamma(T+\kappa_0)}{(\lambda_0 + \bar{P})^{T+\kappa_0}}$ ,  $w_2 = \frac{\bar{P}^{T-1} \lambda_0^{\kappa_0}}{\Gamma(T)\Gamma(\kappa_0)} \frac{\Gamma(T+\kappa_0)}{(\lambda_0 + \bar{P} + \tau_A(Y_0, X_1, Y_1))^{T+\kappa_0}}$ . This last expression enables us to compute the efficient score under strict exogeneity given in (78)-(79)-(80) for Experiment (A).

Equations (73)–(75)–(77) reveal that, in the presence of feedback, computing the efficient score requires evaluating the following four conditional expectations:  $\mathbb{E}[X_2|Y_0, \tilde{P}_1, \bar{P}, X_1]$ ,  $\mathbb{E}[X_2|Y_0, \bar{P}, X_1]$ ,  $\mathbb{E}[X_2V|Y_0, \tilde{P}_1, \bar{P}, X_1]$ ,  $\mathbb{E}[X_2(1 - \tilde{P}_1)V|Y_0, \tilde{P}_1, \bar{P}, X_1]$ . In Experiment (B), the success probability for  $X_2$  is  $p(Y_0, Y_1, X_1, V) = 1 - \exp(-\tau_B(Y_0, X_1, Y_1)V)$  with  $\tau_B(Y_0, X_1, Y_1) = Y_0 + X_1 + Y_1$ . Under this specification, we obtain the following expressions:

$$\begin{aligned} \mathbb{E}[X_2|Y_0, X_1, \tilde{P}_1, \bar{P}] &= 1 - \frac{(\lambda_0 + \bar{P})^{T + \kappa_0}}{(\lambda_0 + \bar{P} + \tau_B(Y_0, X_1, Y_1))^{T + \kappa_0}} \\ \mathbb{E}[X_2|Y_0, X_1, \bar{P}] &= 1 - (\lambda_0 + \bar{P})^{T + \kappa_0} \times \frac{1}{C_1^{(T + \kappa_0)}} \times {}_2F_1\left(T + \kappa_0, \alpha; 1 + \alpha; -\frac{C_2}{C_1}\right) \\ \mathbb{E}[X_2V|Y_0, \tilde{P}_1, \bar{P}, X_1] &= \frac{T + \kappa_0}{\lambda_0 + \bar{P}} - (T + \kappa_0) \frac{(\lambda_0 + \bar{P})^{T + \kappa_0}}{(\lambda_0 + \bar{P} + \tau_B(Y_0, X_1, Y_1))^{T + \kappa_0 + 1}}, \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E} \left[ X_2(1 - \tilde{P}_1)V \mid Y_0, \bar{P}, X_1 \right] \\ &= \frac{1}{2} \frac{T + \kappa_0}{\lambda_0 + \bar{P}} - (T + \kappa_0) \left( \lambda_0 + \bar{P} \right)^{T+\kappa_0} \times \\ & \quad \frac{1}{C_1^{(T+\kappa_0+1)}} \left( {}_2F_1 \left( T + \kappa_0 + 1, \alpha; 1 + \alpha; -\frac{C_2}{C_1} \right) - \frac{1}{2} \times {}_2F_1 \left( T + \kappa_0 + 1, 2\alpha; 1 + 2\alpha; -\frac{C_2}{C_1} \right) \right), \end{aligned}$$

where we use the shorthands  $C_1 = \lambda_0 + \bar{P} + Y_0 + X_1$ ,  $C_2 = \bar{P}^{\frac{1}{\alpha}} e^{-X'_1 \frac{\beta}{\alpha} - \frac{\gamma}{\alpha} Y_0}$ , and  ${}_2F_1(a, b; c; z)$  denotes the hypergeometric function

$${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-zt)^a} dt.$$