Discrete-Time Mixed Proportional Hazard Model

Div Bhagia[a]

[a] *California State University, Fullerton, Department of Economics, SGMH 3361, 2500 Nutwood Ave Fullerton, CA 92831, United States of America*

Abstract

This paper provides a constructive proof of identification for the discrete-time variant of the Mixed Proportional Hazard (MPH) model. The identification result is used to develop a nonparametric estimator based on the Generalized Method of Moments (GMM), which converges at root-n rate and is asymptotically normally distributed. Numerical simulations demonstrate that the estimator also performs well in finite samples.

Keywords: MPH, survival analysis, unobserved heterogeneity, time-to-event, duration models, nonparametric

JEL: C14, C41

1. Introduction

In economics, hazard models are widely employed to analyze time-to-event data, such as in the contexts of exiting unemployment, retirement decisions, and firm failure. It is now well-known that ignoring unobserved heterogeneity in hazard models can cause the estimated hazard to decline more than the true baseline hazard. To account for unobserved heterogeneity, Lancaster (1979) introduced the mixed proportional hazard (MPH) model, a generalization of Cox (1972)'s proportional hazard model. In the MPH model, the hazard rate is specified as the product of three terms: a baseline hazard that varies with time, a regression function that captures the effect of observed covariates, and a random variable that accounts for unobserved heterogeneity.

There is an extensive literature on the identification and estimation of the continuous-time MPH model, where nonparametric identification has been established using variation in the regression function (Elbers and Ridder, 1982; Heckman and Singer, 1984), multiple spell data (Honoré, 1993), and time-varying regressors (Honoré, 1990; Brinch, 2007). However, the discrete-time variant has received less attention. The discrete-time MPH model was first employed by van den Berg and van Ours (1996), who estimated the model using aggregate data, leveraging variation across cohorts for identification. Recently, Alvarez et al. (2021) extended the identification approach using multiple spell data from the continuous to the discrete-time setting.

In this paper, I present the discrete-time counterpart to the identification result that utilizes variation in the regression function. The identification argument is similar to van den Berg and van Ours (1996), but instead of cohort effects, it utilizes variation in an exogenous regressor. My main contribution to the literature is the use of the moment conditions from this model to construct a GMM estimator, which is asymptotically normal and converges at the regular rate of $\sqrt{n}$, where n is the sample size.

Relative to existing alternatives, the discrete-time model and its corresponding GMM estimator outlined in this paper offer several advantages. First, although the continuous-time MPH model is nonparametrically identified, no root-n consistent estimators exist for it.[1] Second, real-world data

Email address: `dbhagia17@gmail.com` (Div Bhagia)

[1] Heckman and Singer (1984) proposed the non-parametric Maximum Likelihood Estimator (NPMLE) for the MPH model, but Baker and Melino (2000) shows that this estimator exhibits significant systematic bias, even in moderately sized samples. Horowitz (1999) proposed an estimator rep-

is often time-aggregated; for example, unemployment durations are typically recorded in weeks, making the discrete-time model particularly relevant. Finally, the estimator is straightforward to implement and allows for regular inference, making it a valuable addition to an applied researcher's toolkit. One notable mention is Hausman and Woutersen (2014), who propose an estimator for the semiparametric continuous time MPH model that accommodates discrete measurement of durations; however, their estimator requires a continuous, time-varying regressor and has a non-smooth objective function, making computation challenging.

2. Identification and Estimation

Let ν represent the unobservable fixed type of each unit. A binary treatment, denoted by the treatment indicator Z, is assigned to each unit, where $Z \in \{0,1\}$.[2] Using potential outcome notation (Rubin, 1974), let T_z denote the realized event duration under treatment z , where T_z takes values in $\{1,2,3,\ldots\}$. The econometrician observes T_z only for units assigned treatment z, so the observed duration T can be written as $T = T_1 Z + T_0(1-Z)$.

Denote the potential probability distribution function and survival probabilities under treatment z by $g_z(\cdot)$ and $S_z(\cdot)$, respectively. To be specific,

$$g_z(t|\cdot) = \Pr(T_z = t|\cdot), \quad S_z(t|\cdot) = \Pr(T_z \geq t|\cdot)$$

with $S_z(1|\cdot) = 1$. Finally, denote the k-th moment of the type distribution by μ_k, such that $\mu_k = \mathbb{E}(\nu^k)$, and define the normalized moments $\tilde{\mu}_k = \mu_k/\mu_1^k$.

Assumption 1. *The hazard at time t for a unit under treatment z is given by the MPH specification as follows:*

$$\Pr(T_z = t \mid T_z \geq t, \nu) = \lambda_t \phi_z \nu,$$

where $\lambda_t > 0$, $\phi_z > 0$, *and* $\nu \in (0, \bar{\nu}]$ *with* $\bar{\nu} < \infty$.

According to the above specification, the hazard, defined as the probability of the event ending at time t given that it hasn't yet ended, is a product of the baseline hazard λ_t capturing the effect of time, the effect of treatment ϕ_z, and the unit-specific unobserved type ν. For instance, if T_z represents unemployment duration, then λ_t would reflect how duration affects the job-finding probability, say due to stigma, skill depreciation, or shifts in job search effort. Meanwhile, the term ν would represent unobserved individual characteristics that influence the likelihood of finding a job. The treatment Z could reflect differences in unemployment benefits, severance pay, or other active labor market policies, with the assumption that these policies proportionally shift the hazard of exiting unemployment by ϕ_z at all durations.

Note that the MPH specification results in the following expressions:

$$g_z(t|\nu) = \lambda_t \phi_z \nu S_z(t|\nu), \quad S_z(t|\nu) = \prod_{s=1}^{t-1} \left[1 - \lambda_s \phi_z \nu\right]$$

Since we assumed that ν is bounded, all its moments will exist. Thus, taking the expectation of the expression for $g_z(t|\nu)$ with respect to ν and expanding it for $t = 1,2,3,\ldots$, we obtain:

$$\begin{aligned}
g_z(1) &= \lambda_1 \phi_z \mu_1 \\
g_z(2) &= \lambda_2 \left[\phi_z \mu_1 - \lambda_1 \phi_z^2 \mu_2\right] \\
g_z(3) &= \lambda_3 \left[\phi_z \mu_1 - (\lambda_1 + \lambda_2)\phi_z^2 \mu_2 + \lambda_1 \lambda_2 \phi_z^3 \mu_3\right] \\
&\vdots
\end{aligned}$$

Or, more compactly,

$$g_z(t) = \lambda_t \sum_{k=1}^{t} \phi_z^k c_k(t) \mu_k$$

where $c_k(t)$ is recursively defined as:

$$c_k(t) = c_k(t-1) - \lambda_{t-1} c_{k-1}(t-1)$$

with $c_1(t) = 1$ for all t and $c_k(t) = 0$ for $k > t$.

resenting the continuous-time MPH model as a transformation model. It requires durations to be measured on a continuous scale and converges at a rate slower than $\sqrt{n}$. For a comprehensive review of estimators for the MPH model, see Hausman and Woutersen (2014).

[2] The identification result and the estimator also apply to cases with more than two treatment levels.

Denote the effect of treatment on hazards by $\gamma = \phi_1/\phi_0$. Note that the baseline hazards can only be identified up to a scale, so we normalize $\phi_0\mu_1 = 1$. Then, by substituting $\phi_1 = \gamma\phi_0$ and $\phi_0 = 1/\mu_1$, we can rewrite the expressions for $g_1(t)$ and $g_0(t)$ as follows:

$$g_1(t) = \lambda_t \sum_{k=1}^{t} \gamma^k c_k(t)\tilde{\mu}_k, \quad g_0(t) = \lambda_t \sum_{k=1}^{t} c_k(t)\tilde{\mu}_k \qquad (1)$$

Given the definition of $c_k(t)$, the above expressions imply that for some integer $\bar{T}$, $g_1(t)$ and $g_0(t)$ for $t = 1,\ldots,\bar{T}$ can be represented as a system of equations involving γ and λ_t and $\tilde{\mu}_t$ for $t = 1,\ldots,\bar{T}$. Since $\tilde{\mu}_1 = 1$ by definition, this results in $2\bar{T}$ unknown parameters and $2\bar{T}$ equations. The following theorem establishes that this system of equations is identified, provided $\phi_1 \neq \phi_0$. The intuition for this result is that the first-period probabilities, $g_1(1)$ and $g_0(1)$, indicate how treatment affects the unit-specific hazard at each duration. This allows us to attribute the remaining differences in later probabilities to composition effects governed by distributional parameters, thereby enabling their identification.

Theorem 1. *Under Assumption 1, the treatment effect $\gamma = \phi_1/\phi_0$ is identified from the first-period potential distributions $g_1(1)$ and $g_0(1)$. Furthermore, provided $\gamma \neq 1$, the baseline hazards $\{\lambda_t\}_{t=1}^{\bar{T}}$ and normalized moments $\{\tilde{\mu}_t\}_{t=2}^{\bar{T}}$ are identified up to a scale from the potential duration distributions $\{g_1(t), g_0(t)\}_{t=1}^{\bar{T}}$.*

Proof. See Appendix A. □

This result implies that if Z is independent of v, we can infer potential duration distributions from the observed duration T and build an estimator for this model using the Generalized Method of Moments (GMM).

Assumption 2. *The following conditions hold:*

(i) Z is independent of v, i.e., $Z \perp v$.

(ii) The sequence $\{T_i, Z_i\}_{i=1}^n$ is independently and identically distributed (i.i.d.).[3]

Denote the set of parameters by $\Theta = \left\{\{\lambda_t\}_{t=1}^{\bar{T}}, \{\tilde{\mu}_t\}_{t=2}^{\bar{T}}, \gamma\right\}$, and let $\tilde{g}_1(t;\Theta)$ and $\tilde{g}_0(t;\Theta)$ represent the potential duration distributions expressed as functions of the parameters, as given by the right-hand side of eq. (1). Define $\pi_z = \Pr(Z = z)$ and consider the following moment condition for unit i:

$$m_{i,t,z}(\Theta) = \frac{\mathbb{I}\{Z_i = z\} \cdot \mathbb{I}\{T_i = t\}}{\pi_z} - \tilde{g}_1(t;\Theta)$$

We can construct a GMM estimator based on the set of moment conditions $m_i(\Theta) = \left\{\{m_{i,t,z}(\Theta)\}_{t=1}^{\bar{T}}\right\}_{z\in\{0,1\}}$ as follows:

$$\hat{\Theta} = \arg\max \hat{m}(\Theta)' \hat{m}(\Theta)$$

where $\hat{m}(\Theta) = \frac{1}{n}\sum_{i=1}^{n} m_i(\Theta)$ is $2\bar{T} \times 1$ vector.

Proposition 1. *Under Assumptions 1 and 2,*

$$\sqrt{n}(\hat{\Theta} - \Theta) \xrightarrow{p} N(0, \hat{M}'M)$$

where $\hat{M} = d\hat{m}(\hat{\Theta})/d\hat{\Theta}$.

Proof. See Appendix B. □

3. Simulations

This section evaluates the finite sample performance of my estimator through Monte Carlo simulations. Each simulation is repeated 5000 times across three sample sizes: $n =$ 1500, 5000, and 10000. The data generation process (DGP) is specified in Table 1. The baseline hazard is modeled using a Weibull function with three different set of parameters to capture a decreasing, constant and increasing hazard, allowing us to test the estimator's performance under different possible realistic scenarios. For example, in the context of unemployment, skill deterioration or stigma would lead to a decreasing hazard. Conversely, individuals searching harder as their savings deplete would result in an increasing hazard.

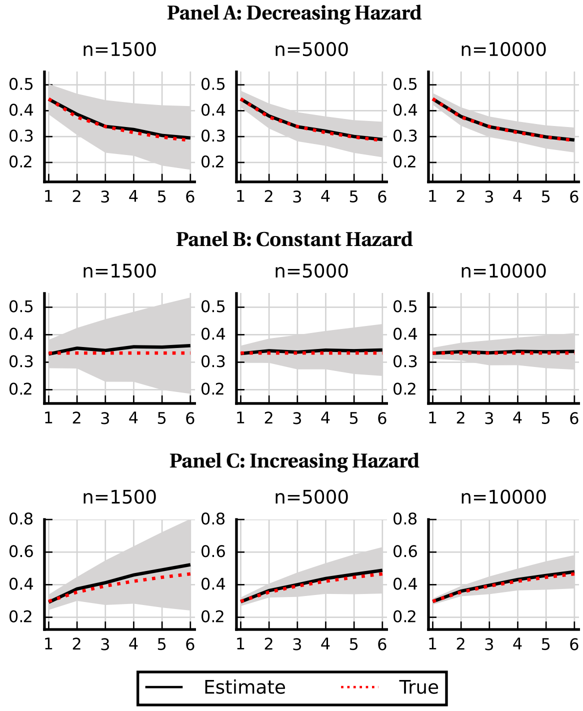
Figure 1 presents the average estimate of the baseline hazard across simulations, plotted against the true baseline hazard for each of the scenarios. The grey shaded areas depict

[3] $\{T_i, Z_i\}$ represents the realization of $\{T, Z\}$ for unit i.

Table 1: Data Generating Process

Parameter	Description	Value
$\bar{T}$	Total number of time periods	6
ϕ_0	Proportional impact of $Z=0$	1
ϕ_1	Proportional impact of $Z=1$	2
$Pr(Z)$	Probability Z equals 1	0.5
$\lambda(t)$	Baseline hazard	$\left(\frac{b}{a}\right)\cdot\left(\frac{t}{a}\right)^{b-1}$
	(a) decreasing ($a=2, b=0.75$)	
	(b) constant ($a=3, b=1$)	
	(c) increasing ($a=3.15, b=1.25$)	

Figure 1: Baseline Hazard Estimates Across Simulations



Note: The plots display the true baseline hazard (dotted red line) alongside the average estimates across simulations (solid black line) for three DGPs with different baseline hazard shapes, each run with sample sizes of 1500, 5000, and 10000. The shaded gray areas represent 95% confidence intervals.

95% confidence intervals, calculated using the standard deviation of the estimates from the simulations. From this figure, we can see that the estimator performs well at capturing the movements of the baseline hazard over time, even in moderately sized samples. Table 2 presents the average bias, standard deviation (SD), and root mean squared error (RMSE) for all parameters, along with the average estimation time per simulation. There are a few things to note: first, estimation is fast, with about 1-4 seconds across different DGPs. Second, the estimates for the normalized moments, specifically higher moments, exhibit some bias in small samples. Though this does not seem to impact the accuracy or precision of the baseline hazard estimates, which are of primary interest in empirical research, making the noisier estimation of higher moments less concerning. However, if the goal is to analyze the distribution of heterogeneity beyond just variance, skewness, or kurtosis, a considerably larger sample size would be necessary.

4. Conclusion

In this paper, I show that the discrete-time MPH model is identified using variation in an exogenous regressor, extending earlier analogous results for the continuous-time model. Moreover, the proof is constructive, resulting in a fully non-parametric GMM-based estimator that is consistent and asymptotically normal. Simulations show that the estimator also performs well in finite samples.

One limitation of the model in the paper from an applied perspective is that the MPH specification is somewhat restrictive, as it assumes the treatment variable affects the hazard uniformly across all durations. However, in practice, the effect of policies can change depending on the time elapsed until the event occurs. For instance, receiving severance pay may reduce the likelihood of exiting unemployment early in the spell but have little impact at later durations. An avenue for future research would be to extend the model to estimate treatment effects that vary with duration.

Disclosure of Interest

The author reports no relevant financial or non-financial competing interests. This study did not receive funding from public, commercial, or non-profit agencies.

References

Alvarez, F. E., K. Borovičková, and R. Shimer (2021, July). Consistent evidence on duration dependence of price changes. Working Paper 29112, National Bureau of Economic Research.

Baker, M. and A. Melino (2000). Duration dependence and nonparametric heterogeneity: A Monte Carlo study. *Journal of Econometrics 96*(2), 357–393.

Brinch, C. N. (2007, April). Nonparametric Identification of the Mixed Hazards Model with Time-Varying Covariates. *Econometric Theory 23*(2), 349–354.

Cox, D. (1972). Regression Models and Life-Tables.

Elbers, C. and G. Ridder (1982, July). True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model. *The Review of Economic Studies 49*(3), 403–409.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica 50*(4), 1029–1054.

Hausman, J. A. and T. Woutersen (2014). Estimating a semi-parametric duration model without specifying heterogeneity. *Journal of Econometrics 178*(P1), 114–131.

Heckman, J. and B. Singer (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica 52*(2), 271–320.

Honoré, B. E. (1990). Identification results for duration models with multiple spells or time–varying covariates.

Honoré, B. E. (1993). Identification Results for Duration Models with Multiple Spells. *The Review of Economic Studies 60*(1), 241–246.

Horowitz, J. L. (1999). Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity. *Econometrica 67*(5), 1001–1028.

Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment. *Econometrica 47*(4), 939–956.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology 66*(5), 688.

van den Berg, G. J. and J. C. van Ours (1996). Unemployment Dynamics and Duration Dependence. *Journal of Labor Economics 14*(1), 100–125.

Table 2: Bias, Standard Error, and RMSE for all Parameters

Panel A: Decreasing Hazard									
Sample size		1500			5000			10000	
Compute time		1.54			1.89			2.01	
Parameter	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
γ	0.017	0.159	0.160	0.006	0.087	0.087	0.003	0.061	0.061
λ_1	-0.002	0.030	0.030	-0.001	0.016	0.016	-0.000	0.012	0.012
λ_2	0.011	0.041	0.042	0.005	0.025	0.025	0.002	0.018	0.018
λ_3	0.000	0.052	0.052	-0.001	0.029	0.029	-0.001	0.020	0.020
λ_4	0.012	0.052	0.053	0.006	0.029	0.030	0.003	0.020	0.021
λ_5	0.006	0.060	0.060	0.002	0.032	0.032	0.000	0.023	0.023
λ_6	0.010	0.063	0.063	0.004	0.035	0.035	0.002	0.025	0.025
$\tilde{\mu}_2$	0.028	0.144	0.146	0.012	0.094	0.094	0.006	0.070	0.070
$\tilde{\mu}_3$	0.223	0.614	0.654	0.095	0.419	0.430	0.053	0.318	0.323
$\tilde{\mu}_4$	1.490	2.562	2.964	0.638	1.526	1.654	0.381	1.158	1.219
$\tilde{\mu}_5$	9.001	14.886	17.396	3.842	6.464	7.519	2.379	4.581	5.162
$\tilde{\mu}_6$	51.255	99.631	112.042	21.285	35.337	41.253	13.420	22.890	26.534

Panel B: Constant Hazard									
Sample size		1500			5000			10000	
Compute time		1.27			1.64			3.58	
Parameter	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
γ	0.037	0.188	0.192	0.016	0.103	0.104	0.009	0.074	0.074
λ_1	-0.004	0.026	0.027	-0.002	0.015	0.015	-0.001	0.010	0.010
λ_2	0.018	0.038	0.042	0.008	0.023	0.024	0.005	0.016	0.017
λ_3	0.009	0.058	0.059	0.003	0.032	0.032	0.001	0.023	0.023
λ_4	0.023	0.065	0.069	0.011	0.036	0.037	0.006	0.026	0.026
λ_5	0.022	0.079	0.082	0.008	0.043	0.044	0.005	0.030	0.031
λ_6	0.027	0.089	0.093	0.011	0.048	0.049	0.006	0.034	0.034
$\tilde{\mu}_2$	0.103	0.236	0.257	0.051	0.158	0.166	0.031	0.122	0.126
$\tilde{\mu}_3$	0.751	1.097	1.329	0.367	0.761	0.844	0.236	0.597	0.642
$\tilde{\mu}_4$	4.476	5.358	6.981	2.136	3.076	3.745	1.413	2.339	2.733
$\tilde{\mu}_5$	24.769	34.429	42.413	11.539	14.679	18.671	7.681	10.200	12.768
$\tilde{\mu}_6$	131.028	232.304	266.709	59.162	83.141	102.042	39.104	54.329	66.939

Panel C: Increasing Hazard									
Sample size		1500			5000			10000	
Compute time		0.98			1.35			1.64	
Parameter	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
γ	0.054	0.195	0.202	0.023	0.107	0.109	0.013	0.075	0.077
λ_1	-0.005	0.024	0.025	-0.003	0.014	0.014	-0.001	0.010	0.010
λ_2	0.020	0.037	0.042	0.010	0.022	0.024	0.006	0.016	0.017
λ_3	0.020	0.070	0.073	0.008	0.038	0.039	0.004	0.028	0.028
λ_4	0.038	0.090	0.098	0.017	0.048	0.051	0.011	0.034	0.036
λ_5	0.046	0.118	0.127	0.018	0.062	0.065	0.011	0.044	0.046
λ_6	0.056	0.143	0.154	0.022	0.073	0.076	0.013	0.052	0.054
$\tilde{\mu}_2$	0.147	0.263	0.301	0.074	0.175	0.190	0.048	0.135	0.143
$\tilde{\mu}_3$	0.953	1.268	1.587	0.488	0.860	0.988	0.323	0.672	0.745
$\tilde{\mu}_4$	4.984	5.756	7.614	2.522	3.414	4.245	1.697	2.647	3.144
$\tilde{\mu}_5$	24.189	31.930	40.058	11.960	14.554	18.838	8.156	10.781	13.519
$\tilde{\mu}_6$	112.368	186.812	218.003	53.898	69.731	88.133	37.034	48.696	61.179

Notes: The table reports average bias, standard deviation, and RMSE for all parameters across nine cases—three DGPs and sample sizes 1500, 5000, and 10000—over 5000 iterations. Compute time represents the average seconds taken for estimation per simulation.

Appendix A. Proof of Theorem 1

Proof. First note that γ is identified as $\gamma = g_1(1)/g_0(1)$. Now, utilizing the expressions for $g_1(t)$ and $g_0(t)$ in eq. (1), we can derive:

$$g_1(t) - \gamma^t g_0(t) = \lambda_t \sum_{k=1}^{t-1} \gamma^k (1-\gamma^{t-k}) c_k(t) \tilde{\mu}_k$$

Note that the summation only extends to $t-1$ instead of t because the last term corresponding to $k = t$ is equal to 0. From the above equation, we can derive the following expression for the baseline hazard:

$$\lambda_t = \frac{g_1(t) - \gamma^t g_0(t)}{\sum_{k=1}^{t-1} \gamma^k (1-\gamma^{t-k}) c_k(t) \tilde{\mu}_k} \tag{1}$$

Similarly, we can also derive an expression for the normalized moments of the type distribution. In particular, taking the ratio of $g_1(t)$ and $g_0(t)$ and expanding the summation to separate out the last term, we get:

$$\frac{g_1(t)}{g_0(t)} = \frac{\gamma^t c_t(t)\tilde{\mu}_t + \sum_{k=1}^{t-1} \gamma^k c_k(t)\tilde{\mu}_k}{c_t(t)\tilde{\mu}_t + \sum_{k=1}^{t-1} c_k(t)\tilde{\mu}_k}$$

Now, from the above expression, we can derive:

$$\tilde{\mu}_t = \frac{g_0(t)\sum_{k=1}^{t-1} \gamma^k c_k(t)\tilde{\mu}_k - g_1(t)\sum_{k=1}^{t-1} c_k(t)\tilde{\mu}_k}{c_t(t)\big[g_1(t) - \gamma^t g_0(t)\big]} \tag{2}$$

The expressions for λ_t and $\tilde{\mu}_t$ in eq. (1) and eq. (2), respectively, are well-defined when $\gamma \neq 1$, which guarantees that $g_1(t) \neq \gamma^t g_0(t)$. Specifically, both the numerator and denominator in these equations are non-zero if $g_1(t) \neq \gamma^t g_0(t)$, because λ_t, $\tilde{\mu}_t$, and $c_t(t)$ (defined as $c_t(t) = (-1)^{t-1}\prod_{s=1}^{t-1}\lambda_s$) are all non-zero by assumption.[4]

To see that $\gamma \neq 1$ implies $g_1(t) \neq \gamma^t g_0(t)$, without loss of generality, assume $\gamma > 1$. In which case, $S_1(t \mid \nu) < S_0(t \mid \nu)$ for all ν and $t > 1$, implying that for $t > 1$, $\mathbb{E}[\nu S_1(t \mid \nu)] < \mathbb{E}[\nu S_0(t \mid \nu)]$. Multiplying both sides by ϕ_0 and λ_t, and noting that $g_z(t) = \lambda_t \phi_z \mathbb{E}[\nu S_z(t \mid \nu)]$, we get that $g_1(t) < \gamma g_0(t)$. Since $\gamma > 1$, it follows that $g_1(t) < \gamma^t g_0(t)$ for $t > 1$. An analogous argument applies when $\gamma < 1$, in which case $g_1(t) > \gamma^t g_0(t)$ for $t > 1$.

The rest of the proof follows by induction. For $\bar{T} = 1$, the statement of the theorem is true as we can identify λ_1 from $g_1(1)$ and $g_0(1)$. Specifically, $\tilde{\mu}_1 = 1$ by definition, and since we normalized $\phi_0\mu_1 = 1$, we have $\lambda_1 = g_0(1)$. Now, assume that the statement is true for $\bar{T} = t-1$; we will argue that then the statement is also true for $\bar{T} = t$. The statement being true for $\bar{T} = t-1$ means that we can identify $\{\lambda_s\}_{s=1}^{t-1}$ and $\{\tilde{\mu}_s\}_{s=2}^{t-1}$ from $\{g_1(s), g_0(s)\}_{s=1}^{t-1}$. Now, if in addition we also have $g_1(t)$ and $g_0(t)$, we can use the expressions given in eq. (1) and eq. (2) to obtain λ_t and $\tilde{\mu}_t$. This is because the expressions for λ_t and μ_t only involve $\{\tilde{\mu}_s\}_{s=2}^{t-1}$ and $\{c_k(t)\}_{k=1}^{t}$, and the latter only depends on $\{\lambda_s\}_{s=1}^{t-1}$. □

Appendix B. Proof of Proposition 1

Proof. From Assumption 2, it follows that:

$$\begin{aligned}\mathbb{E}\left[\frac{\mathbb{I}\{Z=z\}\cdot\mathbb{I}\{T=t\}}{\pi_z}\right] &= \pi_z \cdot \mathbb{E}\left[\frac{\mathbb{I}\{T_z=t\}}{\pi_z}\middle| Z=z\right] \\ &= \mathbb{E}[\mathbb{I}\{T_z=t\}] = g_z(t)\end{aligned}$$

This implies that $\mathbb{E}[m_i(\Theta)] = 0$, and hence, from Theorem 1, Θ is identified from $\mathbb{E}[m_i(\Theta)]$. The rest of the result follows from the asymptotic properties of GMM estimators as established in Hansen (1982). □

[4] Essentially, the denominator in eq. (1), being $\lambda_t > 0$ times the numerator, will also be non-zero if the numerator is non-zero. Additionally, the denominator in eq. (2) is non-zero because $c_t(t) > 0$, and the numerator is non-zero since $\tilde{\mu}_t > 0$.