Universidad
Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Duration and Count Data Models

Autor:

Pedro Henrique Costa Gomes de Sant'Anna

Director:

Miguel A. Delgado

DEPARTAMENTO DE ECONOMIA

Getafe, June/2015

Universidad Carlos III de Madrid
www.uc3m.es

TESIS DOCTORAL

Essays on Duration and Count Data Models

Autor: *Pedro Henrique Costa Gomes de Sant'Anna*

Director: Miguel A. Delgado

Firma del Tribunal Calificador:

Firma

Presidente: Winfried Stute

Vocal: Juan Carlos Escanciano

Secretario: Carlos Velasco

Calificación:

Getafe, de de 2015

Essays on Duration and Count Data Models

by

Pedro H. C. Sant'Anna

Supervisor: Miguel A. Delgado

a dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
Economics
Universidad Carlos III de Madrid

June, 2015

Abstract

This thesis is formed by three chapters related to duration and count data models.

In the first chapter, "Testing for Uncorrelated Residuals in Dynamic Count Models with an Application to Corporate Bankruptcy", I propose new model checks for dynamic count models. Both portmanteau and omnibus-type tests for lack of residual autocorrelation are considered, and the resulting test statistics are asymptotically pivotal when innovations are uncorrelated, but possibly exhibiting higher order serial dependence. Moreover, the tests are able to detect local alternatives converging to the null at the parametric rate $T^{-1/2}$, with T the sample size. I examine the finite sample performance of the test statistics by means of a Monte Carlo experiment. Finally, using a dataset on U.S. corporate bankruptcies, I use the new goodness-of-fit tests to check if different risk models are correctly specified.

In the second chapter, "Nonparametric Tests for Conditional Treatment Effects with Duration Outcomes", I propose new nonparametric tests for treatment effects when the outcome of interest, typically a duration, is subjected to right censoring. The new tests are based on Kaplan-Meier integrals, and do not rely on distributional assumptions, shape restrictions, nor on restricting the potential treatment effect heterogeneity across different subpopulations. The proposed tests are consistent against fixed alternatives and can detect nonparametric alternatives converging to the null at the parametric $n^{-1/2}$-rate, n being the sample size. The finite sample properties of the proposed tests are examined by means of a Monte Carlo study. I illustrate the use of the proposed policy evaluation tools by studying the effect of labor market programs on unemployment duration based on experimental and observational datasets.

The third chapter, “A Simple GMM for Randomly Censored Data”, is a joint work with Miguel A. Delgado. This paper proposes a simple yet powerful GMM setup to estimate parametric regression models when the outcome of interest is subjected to right censoring. The estimation procedure is based on Kaplan-Meier integrals, and is suitable for both linear and nonlinear models, with possible non-smooth moment conditions. We derive general conditions for consistency and asymptotic normality of the parameters of interest. Finally, a small scale simulation study demonstrate satisfactory finite sample properties.

Acknowledgements

Six years have passed since I arrived at Carlos III. Nonetheless, my journey towards the Ph.D started earlier, at home. I would like to thank my father, my mother, and my brother who have always supported my decisions. I also would like to thank Claudio Shikida, Fabio Gomes, and Mario Sudano (in memoriam) for fostering my curiosity towards econometrics while I was at Ibmec.

I wish to express my sincere gratitude to my supervisor, Miguel Delgado, who has put a great effort to guide me, always believing in my potential, and continuously encouraging me to work harder and harder. I would like to say that you are a great model of what a supervisor should be, and I am a thankful for that.

I am grateful to Professors Juan Carlos Escanciano, Carlos Velasco and Winfried Stute for always being there when I needed. Your enthusiasm as true scholars motivates me to move forward. Thanks for that!

I also want to thank my colleagues Emre Ergemen, Xiaojun Song, Mehdi Hamidisahneh, Andres Garcia-Suaza, Nikolas Tsakas, Lovleen Kushwah, Omar Rachedi, Christos Mavridis, and Ursula Mello. Hanging out with you has been great!

I could not forget the great Nora FC football team, including Lian Allub, Manuel Ceballos, David Wehrheim, Facundo Mercado, and Rafael Zambrana. I will always carry with me this great experience (and our games against the Italians!). Also, I am in debt with the Getafe group - Silvina Rubio, Mara Sanz, Luciana Orozco, Ab Zareei, Manuel Mosquera, Vanesa Sosa, Juliana Dornelles, and Ginette Lafit - for having so much fun on our Friday nights at El Eskinazo.

Last, and most importantly, I would like to dedicate this thesis to Renata Sant'Anna.

Without any doubt, she is the most important part of this step of my career. I thank her for always supporting me, with her big heart and energetic tone, inspiring me to achieve my best in all possible dimensions. I can not count how many times she read my papers, listen my presentations, help me with my existential doubts, or enlightened me with her love, determination, and passion. Love you!

Contents

Chapter 1

Testing for Uncorrelated Residuals in Dynamic Count Models with an Application to Corporate Bankruptcy

1.1 Introduction

Credit risk affects virtually every financial contract. Due to its importance, the measurement, pricing and management of credit risk have received much attention from economists, bank supervisors, regulators, and financial market practitioners. Among different credit risk measures, the probability of corporate default (PD) is one of the most popular.

In order to analyze PD, an assumption commonly imposed is that default events are conditionally independent, that is, given observable macroeconomic and financial variables, together with firm specific characteristics, defaults are time independent. Despite such assumption being crucial for the validity of many risk management models, recent studies have found evidence against it, see e.g. Das et al. (2007), Koopman et al. (2011, 2012).

To accommodate deviations from conditional independence, richer classes of models have been proposed. For instance, Koopman et al. (2011, 2012) consider that a common frailty effect, modeled as a Gaussian AR(1), drives the excess default counts clustering. However, an important question remain unanswered: Is the AR(1) structure enough to capture all the excess default correlation? Answering this question is appealing for risk management because, as shown by Duffie et al. (2009) and Koopman et al. (2011, 2012), model misspecification may lead to a downward bias when assessing the probability of extreme default losses.

Motivated by this question, this paper proposes a general model check for dynamic count data models, that is well suited to evaluate the correct specification of aggregate default and bankruptcy models. We propose new portmanteau and omnibus tests for lack of autocorrelation of multiplicative residuals from a dynamic count data model, without imposing parametric distributional assumptions, nor relying on innovations being *iid* or martingale difference. Our tests are able to detect local alternatives converging to the null at the parametric rate $T^{-1/2}$, with T the sample size. Such features are in contrast with classical lack of autocorrelation tests, as Box and Pierce (1970) and Ljung and Box

(1978).

Other specifications tests have been proposed for specific classes of count data models. For instance, Davis et al. (2000) consider a residual autocorrelation test for parameter driven Poisson models, Jung and Tremayne (2003) and Sun and Mccabe (2012) proposed score-tests for the integer autoregressive (INAR) class of models, whereas Neumann (2011) and Fokianos and Neumann (2013) propose goodness-of-fit test some Poisson autoregression models. Nonetheless, the aforementioned proposals are not suitable for models with stochastic macroeconomic covariates. Our tests can be applied to a variety of dynamic count data models, including when stochastic covariates may be of primary interest.

Our procedure builds on Delgado and Velasco (2011), who propose asymptotically distribution-free tests based on the residuals of linear parametric models. We extend their proposal to dynamic count models. In order to project out the effect of replacing the true unknown parameters by their estimated counterparts, we propose an asymptotically pivotal transform of the sample autocorrelations of multiplicative residuals in dynamic count data models. Then, we consider a class of tests for the H_0 of zero residual serial correlation expressed as weighted sums of the first s squared transformed autocorrelations. In particular, we consider distribution-free alternatives to the time-honored Box and Pierce (1970) test based on the transformed autocorrelations. To achieve consistency for a broader class of alternatives, it may be desirable to allow s to grow with the sample size. We then extend our tests to this setup. In particular, we propose a natural alternative to the Bartlett's T_p-process.

Eventually, we apply our goodness-of-fit test procedure to the risk management context. Considering a set of observed macroeconomic and financial variables as covariates, we evaluate the specification of different models for US bankruptcy counts for big public firms, using monthly data from 1985 to 2012. First, we test if using only macroeconomic and financial variables is sufficient to entirely capture the linear dynamics of corporate bankruptcies. If this is the case, one should not find evidence of autocorrelation on the multiplicative residuals from the "static" count data model. Using our proposed test statistics, we reject the null of zero residual serial correlation, which provide evidence of a frailty effect in the default count data. Such result confirm the finds of Duffie et al. (2009) and Koopman et al. (2011, 2012).

Once one finds evidence of a frailty effect, it is common practice to augment the "static" models with additional variables, leading to dynamic specifications - see for instance Koopman et al. (2011, 2012). Following this idea, we consider the Davis et al. (2003) Poisson GLARMA model with an AR(1) term, and the Fokianos and Tjostheim (2011) Poisson log-linear autoregression model of order one. To assess if the inclusion of the additional variables suffice to capture the excess default clustering, we again use our test statistics. Although, the evidence against H_0 is now weaker, we still find some evidence that considering only first order autocorrelation might not be enough to capture the linear dynamics of monthly US bankruptcy counts. Once we augmented these dynamic models with a higher order AR term, we fail to reject H_0. To the best of our knowledge, we are the first to formally assess the correct specification of dynamic count models in a risk management framework.

The rest of the paper is organized as follows: the framework of our test is presented in the next section. In Section 3, we introduce the autocorrelation transformation and discuss its asymptotic properties. In Section 4, we apply the transformation to lack of residual autocorrelation testing. In Section 5, we discuss the finite sample properties of the proposed tests via Monte Carlo simulations. Then, we illustrate our tests with an empirical application for big public corporate bankruptcies. Last section concludes.

1.2 Framework

Let $\{Y_t, \mathbf{X}_t\}_{t\in\mathbb{Z}}$ be a stationary time series such that $Y_t \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$, and $\mathbf{X}_t \in \mathbb{R}^d$ is a vector of covariates. Denote $\lambda_t := E\left[Y_t|\mathcal{F}_{t-1}\right]$, where $\mathcal{F}_t$ is the σ-algebra of events generated by $\{Y_s, \mathbf{X}_{s+1} : s \leq t\}$, which summarizes the history of the joint process. Assume that

$$\lambda_t = \lambda\left(Y_{t-1}, \lambda_{t-1}, \mathbf{X}_t, Y_{t-2}, \lambda_{t-2}, \mathbf{X}_{t-1}, \dots; \boldsymbol{\beta}_0\right) \tag{1.1}$$

where $\boldsymbol{\beta}_0 \in \Theta \subset \mathbb{R}^k$ is a vector of unknown parameters and λ is a measurable function such that $0 < \lambda(\cdot) < \infty$ $w.p.1$. This specification covers most commonly used dynamic count data models. Examples include Davis et al. (2003)'s GLARMA specification, the linear Poisson autoregression of Fokianos et al. (2009), the Poisson log-linear autoregression of Fokianos and Tjostheim (2011), and parameter driven count models as in Zeger (1988), Davis et al. (2000) and Davis and Wu (2009).

The goal of this paper is to assess whether λ_t captures the linear dynamics of Y_t, that is, if the putative model (1.1) is correctly specified. To this end, following Engle (2002), denote the multiplicative error of Y_t as

$$\varepsilon_t = \frac{Y_t}{\lambda_t} \tag{1.2}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a stationary process such that $E(\varepsilon_t) = 1$. As discussed by Engle (2002), the multiplicative structure (1.2) is natural whenever a non-negative time series model is used. Moreover, in contrast with additive errors, the centered multiplicative errors $(\varepsilon_t - 1)$ arise as the predictive errors in Davis et al. (2003)'s GLARMA model, or as the "score-type" errors considered by Creal et al. (2013) for dynamic count data models. Hence, we adopt the multiplicative error structure.

Despite the similarities of (1.2) with volatility models, see e.g. Engle (1982), Bollerslev (1986) and Taylor (1986), some remarks are necessary. First, in order to preserve the integer nature of Y_t, ε_t and λ_t are not mutually independent. Second, because Y_t can be zero with positive probability, log linearization is not a feasible alternative in dynamic count models[1].

In this paper, the focus of our attention is the autocorrelation function of the multiplicative errors $\{\varepsilon_t\}_{t\in\mathbb{Z}}$,

$$\rho_\varepsilon(\tau) = \frac{\gamma_\varepsilon(\tau)}{\gamma_\varepsilon(0)}, \tau \in \mathbb{Z},$$

where $\gamma_\varepsilon(\tau) = Cov(\varepsilon_t, \varepsilon_{t-\tau})$, $\tau \in \mathbb{Z}$, denotes the the autocovariance of order τ of ε_t. The

1. We thank a referee for these remarks.

model (1.1) is correctly specified if the null hypothesis

$$H_0 : \rho_\varepsilon(\tau) = 0, \forall \tau \geq 1 \text{ , for some } \beta_0 \in \Theta \tag{1.3}$$

is satisfied.

In order to assess H_0, one has to first estimate $\rho_\varepsilon(\tau)$. For the moment, assume that $\boldsymbol{\beta}_0$ is known, implying one observe the true $\{\varepsilon_t\}_{t\in\mathbb{Z}}$. Thus, given observations $\{Y_t, \boldsymbol{X}_t\}_{t=1}^T$, $\rho_\varepsilon(\tau)$ can be estimated by the sample autocorrelation function

$$\hat{\rho}_\varepsilon(\tau) = \frac{\hat{\gamma}_\varepsilon(\tau)}{\hat{\gamma}_\varepsilon(0)}, \ \tau \in \mathbb{Z} \tag{1.4}$$

where

$$\hat{\gamma}_\varepsilon(\tau) = \frac{1}{T}\sum_{t=1+\tau}^{T} (\varepsilon_t - 1)(\varepsilon_{t-\tau} - 1), \ \tau \in \mathbb{Z}. \tag{1.5}$$

Define the vector containing the first m sample autocorrelations

$$\hat{\boldsymbol{\rho}}_\varepsilon^{(m)} = (\hat{\rho}_\varepsilon(1), \ldots, \hat{\rho}_\varepsilon(m))'.$$

Under H_0, but allowing general high-order dependence on $\{\varepsilon_t\}_{t\in\mathbb{Z}}$,

$$\sqrt{T}\hat{\boldsymbol{\rho}}_\varepsilon^{(m)} \xrightarrow{d} N\left(0, \boldsymbol{A}_\varepsilon^{(m)}\right), \ \ \boldsymbol{A}_\varepsilon^{(m)} = \left[\frac{a_\varepsilon^{(i,j)}}{\gamma_\varepsilon(0)^2}\right]_{i,j=1}^{m} \tag{1.6}$$

where

$$a_\varepsilon^{(i,j)} = \sum_{l=-\infty}^{\infty} \mathbb{E}\left[(\varepsilon_t - 1)(\varepsilon_{t+i} - 1)(\varepsilon_{t+l} - 1)(\varepsilon_{t+l+j} - 1)\right], \ i, j = 1, \ldots, m,$$

see e.g. Romano and Thombs (1996). If one is willing to impose some additional restrictions on $\{\varepsilon_t\}_{t\in\mathbb{Z}}$, $\boldsymbol{A}_\varepsilon^{(m)}$ can greatly simplify, see e.g. Lobato et al. (2002).

To approximate the asymptotic distribution of $\sqrt{T}\hat{\boldsymbol{\rho}}_\varepsilon^{(m)}$ one has mainly two alternatives. First, as suggested by Romano and Thombs (1996), one can use bootstrap techniques. Another alternative is to use the asymptotic approximation after suitable scaling by a consistent estimator of $\boldsymbol{A}_{\varepsilon}^{(m)}$, see e.g. Lobato et al. (2002). In this paper, we follow the second approach, and avoid computationally intensive methods.

Consider $\hat{\boldsymbol{A}}_{\varepsilon}^{(m)}$ a $m \times m$ positive definite matrix of statistics such that $\hat{\boldsymbol{A}}_{\varepsilon}^{(m)} = \boldsymbol{A}_{\varepsilon}^{(m)} + o_p(1)$ under H_0. Additionally, define the vector of re-scaled sample autocorrelations,

$$\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)} \; = \; (\tilde{\rho}_{\varepsilon}^{(m)}(1), \ldots, \tilde{\rho}_{\varepsilon}^{(m)}(m))' \; = \; \hat{\boldsymbol{A}}_{\varepsilon}^{(m)-1/2} \hat{\boldsymbol{\rho}}_{\varepsilon}^{(m)}. \tag{1.7}$$

Thus, under H_0 and some regularity conditions, from (1.6) it is evident that $\sqrt{T}\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)} \xrightarrow{d} N\left(0, \boldsymbol{I}_m\right).$

So far we have seen that, when $\boldsymbol{\beta}_0$ is known, $\sqrt{T}\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ is asymptotically pivotal under H_0, without relying on the true innovations $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ being *iid* nor martingale difference. Nonetheless, in most cases $\boldsymbol{\beta}_0$ is unknown and has to be estimated.

In the following, assume that an estimator $\hat{\boldsymbol{\beta}}$ is available, such that when $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are not autocorrelated,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + O_p(T^{-1/2}) \tag{1.8}$$

and

$$\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = \boldsymbol{A}_{\varepsilon}^{(m)} + o_p(1). \tag{1.9}$$

where

$$\hat{\varepsilon}_t = \frac{Y_t}{\hat{\lambda}_t},$$

with $\hat{\lambda}_t = \lambda\left(Y_{t-1}, \lambda_{t-1}, \boldsymbol{X}_t, \ldots; \hat{\boldsymbol{\beta}}\right).$

Conditions (1.8) and (1.9) are not very restrictive. In order to get $\sqrt{T}$-consistent estimators of $\boldsymbol{\beta}_0$, a general approach is to consider the Poisson (pseudo) maximum likelihood, such that a $\sqrt{T}$-consistent estimators of $\boldsymbol{\beta}_0$ is given by

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} \sum_{t=1}^{T} \left[Y_t \left(\ln \lambda_t \left(Y_{t-1}, \lambda_{t-1}, \boldsymbol{X}_t, \ldots; \boldsymbol{\beta}\right)\right) - \lambda_t \left(Y_{t-1}, \lambda_{t-1}, \boldsymbol{X}_t, \ldots; \boldsymbol{\beta}\right)\right]. \tag{1.10}$$

see e.g. Davis et al. (2000), Davis et al. (2003), Davis et al. (2005), Davis and Wu (2009), Fokianos et al. (2009), Fokianos and Tjostheim (2011), Christou and Fokianos (2014), Personal et al. (2014) and Agosto et al. (2015) for different specifications of λ_t.

With respect to condition (1.9), one can consider the Newey-West type estimator of Lobato et al. (2002), using the centered multiplicative residual $(\hat{\varepsilon}_t - 1)$, that is,

$$\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = g_{\hat{\varepsilon}}^{(m)}(0) + \sum_{j} k\left(\frac{j}{l}\right)\left\{g_{\hat{\varepsilon}}^{(m)}(j) + g_{\hat{\varepsilon}}^{(m)}(j)'\right\}, \tag{1.11}$$

such that $w_{\hat{\varepsilon}t}^{(m)} = (w_{\hat{\varepsilon},1t}, \ldots, w_{\hat{\varepsilon},mt})'$, $w_{\hat{\varepsilon},kt} = (\hat{\varepsilon}_t - 1)(\hat{\varepsilon}_{t-k} - 1)$, and

$$g_{\hat{\varepsilon}}^{(m)}(j) = T^{-1}\sum_{t=1+j}^{T} w_{\hat{\varepsilon}t}^{(m)} w_{\hat{\varepsilon}t-j}^{(m)\prime},$$

l is a bandwidth and k is the kernel or lag window that satisfy mild regularity conditions; see e.g. Lobato et al. (2002) and Appendix B of Delgado and Velasco (2011). Alternatively, if one is willing to impose that $\boldsymbol{A}_{\varepsilon}^{(m)}$ is diagonal, a consistent estimator for $\boldsymbol{A}_{\varepsilon}^{(m)}$ is $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = diag\left\{\hat{a}_{\hat{\varepsilon}}^{(1,1)}, \ldots, \hat{a}_{\hat{\varepsilon}}^{(m,m)}\right\}/\hat{\gamma}_{\hat{\varepsilon}}(0)^2$, with $\hat{a}_{\hat{\varepsilon}}^{(j,j)} = T^{-1}\sum_{t=1+j}^{T}(\hat{\varepsilon}_t - 1)^2(\hat{\varepsilon}_{t-j} - 1)^2$, and $\hat{\gamma}_{\hat{\varepsilon}}(j) = Cov(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-j})$.

Next, define

$$\boldsymbol{\xi}_{\varepsilon}^{(m)} = \boldsymbol{A}_{\varepsilon}^{(m)-1/2}\boldsymbol{\zeta}_{\varepsilon}^{(m)}$$

with $\boldsymbol{\xi}_{\varepsilon}^{(m)} = (\boldsymbol{\xi}(1)', \ldots, \boldsymbol{\xi}_{\varepsilon}(m)')'$ and $\boldsymbol{\zeta}_{\varepsilon}^{(m)} = (\boldsymbol{\zeta}_{\varepsilon}(1)', \ldots, \boldsymbol{\zeta}_{\varepsilon}(m)')'$, such that $\boldsymbol{\zeta}_{\varepsilon}$ is defined by

$$\frac{\partial}{\partial\boldsymbol{\beta}}\hat{\rho}_{\varepsilon}(j) \xrightarrow{p} \boldsymbol{\zeta}_{\varepsilon}(j) \quad \text{each} \quad j \in \mathbb{Z}\setminus\{0\}$$

under $H_{0.}$ Let $\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ be defined as $\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$, but replacing the true ε_t by its estimate $\hat{\varepsilon}_t$.

Next proposition provides an asymptotic expansion for $\sqrt{T}\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$, which implies that under H_0 and Assumptions 1.1-1.3 in the Appendix, $\sqrt{T}\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ converges to a vector of independent normal variables plus a stochastic drift, which depends on the estimation effect, $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$.

Proposition 1.1 *Under H_0, Assumptions 1.1-1.3 in the Appendix,,*

$$\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)} = \tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)} + \boldsymbol{\xi}_{\varepsilon}^{(m)}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(T^{-1/2}) \tag{1.12}$$

Proof See Appendix.

From Proposition 1.1, once can see that replacing $\boldsymbol{\beta}_0$ by $\hat{\boldsymbol{\beta}}$ on the multiplicative error sample autocorrelations does not come with zero cost, since $\boldsymbol{\xi}_{\varepsilon}^{(m)}$ would not be, in general, zero. This is the case even when $\{\varepsilon_t - 1\}_{t\in\mathbb{Z}}$ is a martingale difference. A special case in which $\boldsymbol{\xi}_{\varepsilon}^{(m)} = 0$ is when λ includes only strictly exogenous covariates, that is, when the model does not include lags of Y_t or λ_t, nor stochastic covariates. This is the case considered by Davis et al. (2000), which may not be appealing for economic applications.

To derive the asymptotic distribution of $\sqrt{T}\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ under H_0 without evoking strong exogeneity conditions, one could use, under suitable conditions, the asymptotic joint distribution of $\{\sqrt{T}\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}, \sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}$; see e.g. Francq et al. (2005). Nonetheless, by adopting such procedure, we would not be able to detect Pitman local alternatives, as described in Section 1.4. Furthermore, given that dynamic count data models are highly non-linear, and closed-form solutions for $\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ are not at our disposal, derivations of the joint distribution can be cumbersome. To avoid these drawbacks, we build on Delgado and Velasco (2011)'s proposal, and suggest an asymptotically distribution-free transform of the estimated multiplicative residual sample autocorrelation $\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ by means of recursive least squares projections.

1.3 A martingale transform of the multiplicative residual sample autocorrelation function with estimated parameters

In order to deal with the distribution of the residuals autocorrelation with estimated parameters, Delgado and Velasco (2011) propose a transformation based on the recursive least squares residuals introduced by Brown et al. (1975) for CUSUM tests of parameter instability.

To motivate the transformation, consider the asymptotic decomposition in Proposition

1.1,

$$\tilde{\rho}_{\hat{\varepsilon}}^{(m)}(\tau) = \tilde{V}^{(m)}(\tau) + o_p(T^{-1/2}), \qquad \tau = 1, \ldots, m,$$

with

$$\tilde{V}^{(m)}(\tau) = \tilde{\rho}_{\varepsilon}^{(m)}(\tau) + \boldsymbol{\xi}_{\varepsilon}(\tau)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).$$

Under H_0, one can see that the source of asymptotic autocorrelation in $\left\{\tilde{V}^{(m)}(\tau)\right\}_{\tau=1}^{m}$ is $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Then, the transformation consists in using a linear operator $\mathcal{L}^{(m)}$ such that $\left\{\mathcal{L}^{(m)}\tilde{V}^{(m)}\right\}_{\tau\geq 1}$ are asymptotically uncorrelated when the true multiplicative errors are serially uncorrelated.

Similar to Delgado and Velasco (2011), we consider the operator that transform any sequence $\{\eta(\tau)\}_{\tau=1}^{m}$ in the forward recursive residuals of its least square projection on $\{\boldsymbol{\xi}_{\varepsilon}(\tau)\}_{\tau=1}^{m}$,

$$\mathcal{L}^{(m)}\eta(\tau) = \eta(\tau) - \boldsymbol{\xi}_{\varepsilon}(\tau)\kappa_{\tau+1}^{\eta}$$

where

$$\kappa_{\tau+1}^{\eta} \equiv \left(\sum_{l=\tau+1}^{m} \boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)\right)^{-1} \sum_{l=\tau+1}^{m} \boldsymbol{\xi}_{\varepsilon}(l)'\eta(l).$$

The intuition behind such operator is simple. First, notice that $\kappa_{\tau+1}^{\eta}$ is the OLS coefficient vector from regressing $\eta_{\tau+1}^{(m)}$ on $\boldsymbol{\xi}_{\varepsilon,\tau+1}^{(m)}$, where $\boldsymbol{\xi}_{\varepsilon,\tau+1}^{(m)} = \left(\boldsymbol{\xi}(\tau+1)', \ldots, \boldsymbol{\xi}(m)'\right)'$ is the matrix of last $m - \tau \times k$ standardized scores, and $\eta_{\tau+1}^{(m)} = \left(\eta(\tau+1), \ldots, \eta(m)\right)'$ is defined analogously. Then, one can see that $\mathcal{L}^{(m)}\eta(\tau)$ consist of first computing recursive OLS coefficients, and then, based on these estimates, compute the "one-step-back forecast error" $\eta(\tau) - \boldsymbol{\xi}_{\varepsilon}(\tau)\kappa_{\tau+1}^{\eta}$.

Given the linearity of the operator $\mathcal{L}^{(m)}$, we have that

$$\mathcal{L}^{(m)}\tilde{V}^{(m)}(\tau) = \mathcal{L}^{(m)}\tilde{\rho}_{\varepsilon}^{(m)}(\tau) + \mathcal{L}^{(m)}\boldsymbol{\xi}_{\varepsilon}(\tau)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).$$

Nonetheless,

$$\begin{aligned}\mathcal{L}^{(m)}\boldsymbol{\xi}_{\varepsilon}(\tau)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) &= \boldsymbol{\xi}_{\varepsilon}(\tau)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) - \boldsymbol{\xi}_{\varepsilon}(\tau)\left(\sum_{l=\tau+1}^{m}\boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)\right)^{-1}\sum_{l=\tau+1}^{m}\boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\ &= \boldsymbol{\xi}_{\varepsilon}(\tau)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) - \boldsymbol{\xi}_{\varepsilon}(\tau)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\ &= 0.\end{aligned}$$

Therefore, we have $\mathcal{L}^{(m)}\tilde{V}^{(m)}(\tau) = \mathcal{L}^{(m)}\tilde{\rho}_{\varepsilon}^{(m)}(\tau)$, $\tau = 1,\ldots,m-k$, which does not depend on $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$. Since $\left\{\sqrt{T}\tilde{\rho}_{\varepsilon}^{(m)}(\tau)\right\}_{\tau\geq 1}$ are asymptotically distributed as *iid* standard normal, $\{\sqrt{T}\mathcal{L}^{(m)}\tilde{\rho}_{\varepsilon}^{(m)}(\tau)\}_{\tau\geq 1}$ are asymptotically distributed as independent normal random variables with mean zero and variance

$$\sigma^2(\tau) = 1 + \boldsymbol{\xi}_{\varepsilon}(\tau)\left(\sum_{l=\tau+1}^{m}\boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)\right)^{-1}\boldsymbol{\xi}_{\varepsilon}(\tau)'. \tag{1.13}$$

Thus, under H_0, the vector $\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)} = \left(\bar{\rho}_{\varepsilon}^{(m)}(1),\ldots,\bar{\rho}_{\varepsilon}^{(m)}(m-k)\right)'$, such that

$$\bar{\rho}_{\varepsilon}^{(m)}(\tau) = \frac{\hat{\mathcal{L}}^{(m)}\tilde{\rho}_{\varepsilon}^{(m)}(\tau)}{\sigma(\tau)},$$

is asymptotically distributed as a vector of independent standard normals.

1.3.1 Computation of the transformed multiplicative residual autocorrelations

The aforementioned transformation is unfeasible because, in practice, we do not know $\boldsymbol{\xi}_{\varepsilon}^{(m)}$. Thus, in order to implement a feasible transformation of the multiplicative residual sample autocorrelations, we need an estimator for $\boldsymbol{\xi}_{\varepsilon}^{(m)}$.

From (1.22) and (1.23) in the proof of Proposition 1, it is evident that under, H_0, standardizing by $\hat{\gamma}_{\hat{\varepsilon}}(0)$ in $\hat{\boldsymbol{\rho}}_{\hat{\varepsilon}}$ has no asymptotic effect on $\boldsymbol{\zeta}_{\varepsilon}$ in the expansion (1.12). Therefore, one can estimate $\boldsymbol{\xi}_{\varepsilon}^{(m)}$ by

$$\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}^{(m)} = \hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)-1/2} \hat{\boldsymbol{\zeta}}_{\hat{\varepsilon}}^{(m)}, \tag{1.14}$$

where $\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}^{(m)} = \left(\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(1)', \ldots, \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(m)'\right)'$ and $\hat{\boldsymbol{\zeta}}_{\hat{\varepsilon}}^{(m)} = \left(\hat{\boldsymbol{\zeta}}_{\hat{\varepsilon}}(1)', \ldots, \hat{\boldsymbol{\zeta}}_{\hat{\varepsilon}}(m)'\right)'$, with

$$\hat{\boldsymbol{\zeta}}_{\hat{\varepsilon}}(\tau) = \frac{-1}{T\ \hat{\gamma}_{\hat{\varepsilon}}(0)} \left(\sum_{t=\tau+1}^{T} \frac{\hat{\varepsilon}_t}{\hat{\lambda}_t} \left(\hat{\varepsilon}_{t-\tau} - 1\right) \left.\frac{\partial \lambda_t}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \frac{\hat{\varepsilon}_{t-\tau}}{\hat{\lambda}_{t-\tau}} \left(\hat{\varepsilon}_t - 1\right) \left.\frac{\partial \lambda_{t-\tau}}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right)$$

$$\hat{\gamma}_{\hat{\varepsilon}}(\tau) = \frac{1}{T} \sum_{t=\tau+1}^{T} \left(\hat{\varepsilon}_t - 1\right)\left(\hat{\varepsilon}_{t-\tau} - 1\right).$$

The vector of partial derivatives $\partial\lambda_t/\partial\boldsymbol{\beta}$ can be computed recursively, see e.g. Davis et al. (2005) and Liboschik et al. (2015).

Once $\hat{\boldsymbol{\xi}}^{(m)}$ is defined, one can implement the feasible transformation of $\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ using the following steps:

1. Compute $\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ and $\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}^{(m)}$ using (1.7) and (1.14), respectively.

2. For $\tau = 1, \ldots, m-k$, compute

$$\hat{\kappa}_{\tau+1} = \left(\sum_{l=\tau+1}^{m} \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(l)' \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(l) \right)^{-1} \sum_{l=\tau+1}^{m} \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(l)' \tilde{\rho}_{\hat{\varepsilon}}^{(m)}(l).$$

3. For $\tau = 1, \ldots, m-k$, compute the feasible recursive least squares residuals

$$\hat{\mathcal{L}}^{(m)} \tilde{\rho}_{\hat{\varepsilon}}^{(m)}(\tau) = \tilde{\rho}_{\hat{\varepsilon}}^{(m)}(\tau) - \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(\tau) \hat{\kappa}_{\tau+1}.$$

With $\hat{\mathcal{L}}^{(m)} \tilde{\rho}_{\hat{\varepsilon}}^{(m)}(\tau)$ at hands, the transformed multiplicative residuals sample autocorrelations is given by

$$\bar{\rho}_{\hat{\varepsilon}}^{(m)}(\tau) = \frac{\hat{\mathcal{L}}^{(m)} \tilde{\rho}_{\hat{\varepsilon}}^{(m)}(\tau)}{\hat{\sigma}(\tau)}, \tau = 1 \ldots, m-k, \tag{1.15}$$

where $\hat{\sigma}^2(\tau) = 1 + \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(\tau) \left(\sum_{l=\tau+1}^{m} \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(l)' \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(l) \right)^{-1} \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(\tau)'$ is the estimator of $\sigma^2(\tau)$. Notice than we can only transform the first $m-k$ sample autocorrelations, because, giving a scaling matrix $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)}$, there are no more degrees of freedom available when k parameters are estimated.

Next, we must prove that, under H_0, $\bar{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)} = \left(\bar{\rho}_{\hat{\varepsilon}}^{(m)}(1), \ldots, \bar{\rho}_{\hat{\varepsilon}}^{(m)}(m-k)\right)'$ and $\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ are asymptotically equivalent, and $\sqrt{T}\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ is asymptotically distributed as a vector of independent standard normals.

Theorem 1.1 *Under H_0, $m > k$ and Assumptions 1.1-1.4 in the Appendix,*

$$\bar{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)} = \bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)} + o_p(T^{-1/2})$$

and

$$\sqrt{T}\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)} \xrightarrow{d} N_{m-k}(0, \boldsymbol{I}_{m-k}).$$

Proof See Appendix.

Theorem 1.1 shows that, without relying on $\{\varepsilon_t - 1\}$ being *iid* or martingale difference, $\sqrt{T}\bar{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ is asymptotically distribution-free under H_0. This result forms the basis for implementing our test of lack of autocorrelation described in the next section.

1.4 Testing lack of autocorrelation on the multiplicative residuals with estimated parameters

We consider the class of tests for H_0 expressed as weighted sums of the squared transform autocorrelations. That is, for some $1 \leq s \leq m-k$, our test statistics are of the form

$$W^{(m)}(s) = T\sum_{\tau=1}^{s} w_T(j)\, \bar{\rho}_{\hat{\varepsilon}}^{(m)}(\tau)^2, \tag{1.16}$$

where w_T is a summable weight function such that $w_T : \mathbb{N} \to \mathbb{R}^+$ and $w(j) = \lim_{T\to\infty} w_T(j)$ is bounded. We consider portmanteau-type tests, where s is fixed, and also omnibus tests, where we allow s to grow with the sample size.

To discuss the power of our proposed tests, consider Pitman local alternatives of the type

$$H_{1T}\colon\ \rho_\varepsilon(\tau) = \frac{r(\tau)}{\sqrt{T}} + \frac{j_T(\tau)}{T}\ \forall\tau = 1, 2, \ldots, \tag{1.17}$$

where r and j_T are square summable such that ρ_ε is positive definite sequence for all T.

In order to proceed with the power discussion, we must first derive the asymptotic distribution of $\bar{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ under H_{1T}. To this end, define the vector of projected and standardized autocorrelation drifts as $\check{\boldsymbol{h}}_\varepsilon^{(m)} = \left(\check{h}_\varepsilon^{(m)}(1),\ \ldots,\ \check{h}_\varepsilon^{(m)}(m-k)\right)'$, where

$$\check{h}_\varepsilon^{(m)}(\tau) = h_\varepsilon^{(m)}(\tau) - \boldsymbol{\xi}_\varepsilon(\tau)\left(\sum_{l=\tau+1}^{m}\boldsymbol{\xi}_\varepsilon(l)'\boldsymbol{\xi}_\varepsilon(l)\right)^{-1}\sum_{l=\tau+1}^{m}\boldsymbol{\xi}_\varepsilon(l)'h_\varepsilon^{(m)}(\tau) \tag{1.18}$$

$\tau = 1, \ldots, m-k$,

$$h_\varepsilon^{(m)}(\tau) = \sum_{i=1}^{m}\left[\boldsymbol{A}_\varepsilon^{(m)-1/2}\right]_{(\tau,i)} r(i). \tag{1.19}$$

and let $\bar{\boldsymbol{h}}_\varepsilon^{(m)} = \left(\bar{h}_\varepsilon^{(m)}(1),\ \ldots,\ \bar{h}_\varepsilon^{(m)}(m-k)\right)'$,where $\bar{h}_\varepsilon^{(m)}(\tau) = \check{h}_\varepsilon^{(m)}(\tau)/\sigma(\tau)$, with $\sigma(\tau)$ as defined in (1.13).

Theorem 1.2 *Under H_{1T}, $m > k$, Assumptions 1.2-1.4 in the Appendix,*

$$\bar{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)} = \bar{\boldsymbol{\rho}}_\varepsilon^{(m)} + o_p(T^{-1/2})$$

and

$$\sqrt{T}\bar{\boldsymbol{\rho}}_\varepsilon^{(m)} \xrightarrow{d} N_{m-k}(\bar{\boldsymbol{h}}_\varepsilon^{(m)}, \boldsymbol{I}_{m-k}).$$

Proof See Appendix.

Next, we discuss the asymptotic properties of our portmanteau and omnibus type tests

1.4.1 Portmanteau-type tests

In this sub-section we consider tests where the number of autocorrelations s in $W^{(m)}(s)$ is fixed.

It follows from Theorem 1.1 that, under H_0,

$$W^{(m)}(s) \xrightarrow{d} \sum_{\tau=1}^{s} w\left(j\right) Z_\tau^2$$

where $\{Z_\tau\}_{\tau\in\mathbb{N}}$ are *iid* standard normals. On the other hand, from Theorem 1.2, one can see that under H_{1T},

$$W^{(m)}(s) \xrightarrow{d} \sum_{\tau=1}^{s} w\left(j\right) \left(Z_\tau + \bar{h}_\varepsilon^{(m)}(\tau)\right)^2 .$$

Hence, one can easily conclude that our tests $W^{(m)}(s)$ are asymptotically distribution-free, and are able to detect nonparametric local alternatives like H_{1T}, which converges to the null hypothesis at the parametric rate.

Next, we discuss some particular choices of w_T. For instance, consider the uniform weights $w_T\left(j\right) = 1\left\{j \le s\right\}$, $1 \le s \le m-k$, for each $j \in \mathbb{N}$, which correspond to the test statistic

$$\bar{Q}_{BP}^{(m)}(s) = T\sum_{\tau=1}^{s} \bar{\rho}_{\hat{\varepsilon}}^{(m)}(\tau)^2, \tag{1.20}$$

which is the transformed version of the time honored Box and Pierce (1970) statistic (henceforth BP), $\hat{Q}_{BP}(s)$, with

$$\hat{Q}_{BP}(s) = T\sum_{\tau=1}^{s} \hat{\rho}_{\hat{\varepsilon}}(\tau)^2. \tag{1.21}$$

Alternatively, setting $w_T\left(j\right) = 1\left\{j \le s\right\}\left(T+2\right)/\left(T-j\right)$, $1 \le s \le m-k$, for each $j \in \mathbb{N}$, we get

$$\bar{Q}_{LB}^{(m)}(s) = T\left(T+2\right)\sum_{\tau=1}^{s} \frac{\bar{\rho}_{\hat{\varepsilon}}(\tau)^2}{T-\tau},$$

leading to a natural alternative to Ljung and Box (1978) statistic

$$\hat{Q}_{LB}(s) = T\left(T+2\right)\sum_{\tau=1}^{s} \frac{\hat{\rho}_{\hat{\varepsilon}}(\tau)^2}{T-\tau}.$$

Next corollary summarizes the asymptotic properties of $\bar{Q}_{BP}^{(m)}(s)$. The results for $\bar{Q}_{LB}^{(m)}(s)$ are analogous.

Corollary 1.1 *(i) Under H_0 and the conditions stated on Theorem 1.1,*

$$\bar{Q}_{BP}^{(m)}(s) \xrightarrow{d} \chi^2_{(s)},$$

$1 \leq s \leq m-k$, *where* $\chi^2_{(s)}$ *is a chi-square distribution with* s *degrees of freedom.*

(ii)Under H_{1T} and the conditions stated on Theorem 1.2,

$$\bar{Q}_{BP}^{(m)}(s) \xrightarrow{d} \chi^2_{(s)}(\varphi),$$

$1 \leq s \leq m-k$, *where* $\chi^2_{(s)}\left(\varphi^{(m)}\right)$ *is a non-central chi-squared with non-centrality parameter* $\varphi^{(m)} = \sum_{j=1}^{s}\left(\bar{h}_{\varepsilon}^{(m)}(j)\right)^2$.

Our results in Corollary 1.1 are in sharp contrast with the ones of Box and Pierce (1970) and Ljung and Box (1978). Assuming that $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are *iid* and letting s increase with sample size T, in particular $s = o\left(T^{1/2}\right)$, $\hat{Q}_{BP}(s)$ is asymptotically distributed as $\chi^2_{(s-k)}$. However, if $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ are not *iid* or s remains fixed, $\hat{Q}_{BP}(s)$ has a limiting null distribution that depends on the true parameters $\boldsymbol{\beta}_0$ and other unknown features of the underlying data generating process. Additionally, as shown by Hong (1996), $\hat{Q}_{BP}(s)$ and $\hat{Q}_{LB}(s)$ are not able to detect nonparametric local alternatives like H_{1T}. Our test statistics $\bar{Q}_{BP}^{(m)}(s)$, on the other hand, is asymptotically distributed as $\chi^2_{(s)}$, for any $s \leq m-k$, and have non-trivial power against H_{1T}, without relying on $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ being *iid* or martingale difference. These are the main advantages of our proposal.

So far, we have not discussed how one can choose m, the number of scores included to compute the projection. We recommend to set $m = T-1$, that is, to use all the information available in the data. This choice leads to test statistics with good size and power properties, as illustrated in the simulation results of Section 1.5. Moreover, Delgado and Velasco (2011) show that, in a Gaussian context, letting $m \to \infty$ as $T \to \infty$ leads to optimal tests. Hence, the aforementioned choice of m seems appropriate. Nonetheless, we notice that setting $m = T-1$ is not feasible if one does not impose additional restrictions on $\boldsymbol{A}_{\varepsilon}^{(m)}$. When this is the case, and the Newey-West type estimator (1.11) is used, we set $m = s + k + \sqrt{T}$. With this choice, we allow m to grow with the sample size, and at the same time, avoid potential singularity issues that may arise when m is too large.

1.4.2 Omnibus-type tests

The portmanteau class of tests discussed in the previous sub-section hold s fixed. Hence, they are unable to detect serial correlation appearing at lags larger than s. In order to circumvent this issue, and achieve consistency for a broader class of alternatives, it may be desirable to allow s to grow with the sample size.

In the following, we derive the asymptotic properties of tests of the form (1.16), but allowing s to grow with the sample size.

Proposition 1.2 *Consider test statistics of the form (1.16) where the sequence of weights* $\{w(j)\}_{j=1}^{\infty}$ *satisfies* $\sum_{j=1}^{\infty} w(j) < \infty$. *Then, for any* $s \to \infty$ *as* $T \to \infty$, $s(T) = m(T) - k < T$, *it follows that:*

(i) Under H_0, *and the conditions stated on Theorem 1.1*

$$W^{(m)}(s) \xrightarrow{d} \sum_{\tau=1}^{\infty} w(\tau) Z_{\tau}^2,$$

where $\{Z_\tau\}_{\tau=1}^{\infty}$ *is a sequence of independent standard normal distributions.*

(ii) Under H_{1T} *and the conditions stated on Theorem 1.2,*

$$W^{(m)}(s) \xrightarrow{d} \sum_{\tau=1}^{\infty} w(\tau) \left(Z_\tau + \bar{h}_{\varepsilon}^{(m)}(\tau)\right)^2.$$

Proof See Appendix.

From the results in Proposition 1.2, one can see that our tests are distribution-free and are able to detect local alternatives of the form (1.17).

Notice that setting $w_T(j) = 1/(\pi j)^2$ leads to the test statistic

$$\bar{T}^{(m)} = T \sum_{\tau=1}^{s(T)} \frac{\bar{\rho}_{\hat{\varepsilon}}(\tau)^2}{(\pi\tau)^2}$$

which resemble the spectral representation of the $T_p - process$ test based on the Cramér-von Mises criterion

$$\hat{T}^{(m)} = T \sum_{\tau=1}^{T-1} \frac{\hat{\rho}_{\hat{\varepsilon}}(\tau)^2}{(\pi\tau)^2}.$$

See, for example, Anderson (1993) and Delgado et al. (2005). Asymptotic 10%, 5% and 1% critical values for $\bar{T}^{(m)}$ are 0.347, 0.461 and 0.743, respectively, see e.g. Shorack and Wellner (1986).

Regarding the choice of $s(T)$, when one assume that $\boldsymbol{A}_{\varepsilon}^{(m)}$ is diagonal (or the identity matrix), one can set $s(T) = m(T) = T - 1 - k$, where k is the number of parameters in the model. In cases where $\boldsymbol{A}_{\varepsilon}^{(m)}$ is unrestricted, a choice of $s(T)$ (and $m(T)$) can be more delicate due to potential singularity issues. A full discussion about it is beyond the scope of this paper, and we leave this for future research.

Remark 1.1 Although the omnibus-type tests are consistent against a broader class of alternatives when compared to the portmanteau tests, there are situations in which the later is more powerful than the first. In fact, Delgado and Velasco (2010) show that portmanteau type tests as $W^{(m)}(s)$ fall into the class of Neyman's smooth test, which are optimal to detect local alternatives of the type (1.17). Moreover, Escanciano and Lobato (2009) show that, when one does not have an alternative r in mind, Box-Pierce type tests are optimally adaptive to the unknown local alternative. Hence, one should see the portmanteau and omnibus type tests as complements rather than substitutes.

1.5 Monte Carlo Simulations

In this section we present a comparative study of the significance level and power of the statistics $\hat{Q}_{BP}(s)$ and $\bar{Q}_{BP}^{(m)}(s)$ given by (1.21) and (1.20). The critical values of $\hat{Q}_{BP}(s)$ and $\bar{Q}_{BP}^{(m)}(s)$ have been obtained using percentiles of the $\chi^2_{(s-k)}$ and $\chi^2_{(s)}$ distributions, respectively, where k is the number of regressors included in the model. We consider sample sizes $T = 100$ and 300, and $10,000$ replications in each experiment. For the sake of comparison, we use values of s equal to 1, 2, 5 and 10 when $T = 100$, and 1, 2, 5, 10 and 17 when $T = 300$. The choice of 17 is based on the rule of thumb of setting $s = \sqrt{T}$. All models are estimated using the Poisson quasi-maximum likelihood (1.10).

The nominal level of all tests is 5%.

To compute $\bar{Q}_{BP}^{(m)}(s)$, we consider three estimates of $\boldsymbol{A}_{\varepsilon}^{(m)}$: (i) $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = \boldsymbol{I}_m$, (ii) $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = diag\left\{\hat{a}_{\hat{\varepsilon}}^{(1,1)}, \ldots, \hat{a}_{\hat{\varepsilon}}^{(m,m)}\right\} / \hat{\gamma}_{\hat{\varepsilon}}(0)^2$, with $\hat{a}_{\hat{\varepsilon}}^{(j,j)} = T^{-1}\sum_{t=1+j}^{T} \hat{\varepsilon}_t^2 \hat{\varepsilon}_{t-j}^2$, and (iii) the Newey-West-type unrestricted estimator of $\boldsymbol{A}_{\varepsilon}^{(m)}$ in (1.11), with preliminary bandwidth $n = 2\left(T/100\right)^{1/3}$, no prewhitening and Barlett's kernel.

Regarding m, we follow the recommendation discussed in the previous section and set $m = T-1$ if $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)}$ is diagonal, or $m = s + k + \sqrt{T}$ if the Newey-West-type estimator of $\boldsymbol{A}_{\varepsilon}^{(m)}$ is used.

We concentrate on exponential mean models, since this is the canonical functional form in count data models. For $t = 1, \ldots, T$, we consider the following models under H_0:

$$(i)\ Y_t \sim Poisson(\lambda_t),$$
$$(ii)\ Y_t \sim Neg.Binomial\ \left(2, \lambda_t\right),$$
$$(iii)\ Y_t \sim Poisson(\lambda_t \exp\left(v_t\right)),$$

where

$$\lambda_t = \exp\left(1 + 0.5X_t\right),$$
$$X_t = 0.5X_{t-1} + u_t,$$

and $\{u_t\}_{t\in\mathbb{Z}}$ follows an *iid* standard normal distribution, and $\{v_t\}_{t\in\mathbb{Z}}$ follows and *iid* normal distribution with mean 0 and variance 0.5. This way, on specification (i) we have a standard Poisson model, on (ii) a Negative Binomial model, and on (c) we introduce a multiplicative latent process as first considered by Zeger (1988). Note that models (ii) and (iii) leads to overdispersed count models, a common feature in applications.

Tables 1.1 and 1.2 report the simulated empirical size of the tests with $T = 100$ and $T = 300$, respectively. We can see that, for the BP test $\hat{Q}_{BP}(s)$, the type I error is out of control for any sample size when s is smaller than $\sqrt{T}$. Given that the chi-squared approximation to $\hat{Q}_{BP}(s)$ is only justified when s is large, this result is expected. Our projected test statistics, in general, control the type I error well, even when s is small.

When the Newey-West type estimator of $\boldsymbol{A}_{\varepsilon}^{(m)}$ is coupled with large values of s, our test $\bar{Q}_{BP}^{(m)}(s)$ tends to be oversized when $T = 100$, and undersized when $T = 300$, perhaps due to the need of inverting a matrix of large dimension. Diagonal restrictions on $\boldsymbol{A}_{\varepsilon}^{(m)}$ seems to be a good strategy to control size across the different DGP considered.

Table 1.1: Empirical size at 5 % significance. T=100

DGP	s	$\hat{Q}_{BP}$	$\bar{Q}_{BP}^{(m)}(I)$	$\bar{Q}_{BP}^{(m)}(D)$	$\bar{Q}_{BP}^{(m)}(N)$
	1	NA	5.41	4.61	5.96
(i)	2	15.6	4.79	4.25	7.06
	5	8.86	4.20	4.62	7.84
	10	6.35	3.70	5.18	7.64
	1	NA	4.23	5.30	5.87
(ii)	2	13.28	3.83	4.79	6.79
	5	7.28	3.65	5.15	7.97
	10	5.64	3.25	5.72	7.91
	1	NA	4.54	5.04	6.18
(iii)	2	13.86	3.99	4.50	7.05
	5	7.42	3.37	4.56	7.98
	10	5.58	3.36	5.50	7.64

Note: $\bar{Q}_{BP}^{(m)}(W)$ denotes $\bar{Q}_{BP}^{(m)}(s)$ using $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m} = W$. $W = I$ means $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m} = \boldsymbol{I}_m$, $W = D$ means $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m}$ is diagonal, and $W = N$ means the Newey-West type estimator of $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m}$ is used.

In order to analyze the power of our tests, we consider the following specifications under H_1 :

$$(iv)\ Y_t \sim Poisson(\omega_t),\ \omega_t = \exp\left(1 + 0.5X_t + c\frac{Y_{t-1} - \omega_{t-1}}{\omega_{t-1}}\right),$$

$$(v)\ Y_t \sim Neg.Binomial\ (2, \omega_t),\ \omega_t = \exp\left(1 + 0.5X_t + c\frac{Y_{t-1} - \omega_{t-1}}{\omega_{t-1}}\right),$$

$$(vi)\ Y_t \sim Poisson(\lambda_t \exp(z_t)),\ z_t = cz_{t-1} + v_t,$$

$$(vii)\ Y_t \sim Poisson(\omega_t),\ \omega_t = \exp(1 + 0.5X_t + c\log(1 + Y_{t-1})),$$

$$(viii)\ Y_t \sim Neg.Binomial\ (2, \omega_t),\ \omega_t = \exp(1 + 0.5X_t + c\log(1 + Y_{t-1})),$$

$$(ix)\ Y_t = \mathcal{B}(Y_{t-1}, c) + A_t,\ A_t \sim Poisson(\lambda_t),$$

$$(x)\ Y_t = \mathcal{B}(Y_{t-1}, c) + A_t,\ A_t \sim Neg.Binomial\ (2, \lambda_t),$$

Table 1.2: Empirical size at 5 % significance. T=300

DGP	s	$\hat{Q}_{BP}$	$\bar{Q}_{BP}^{(m)}(I)$	$\bar{Q}_{BP}^{(m)}(D)$	$\bar{Q}_{BP}^{(m)}(N)$
	1	NA	6.16	4.7	3.47
	2	17.26	6.07	4.53	3.45
(i)	5	10.24	5.85	5.17	3.14
	10	7.98	5.29	5.49	2.46
	17	6.31	4.31	5.24	1.59
	1	NA	5.21	5.13	3.48
	2	14.48	4.81	4.88	3.05
(ii)	5	8.98	4.7	5.59	3.33
	10	6.83	4.14	5.5	2.62
	17	5.63	3.82	6.16	1.77
	1	NA	5.24	5.00	3.66
	2	14.78	4.90	4.97	3.63
(iii)	5	8.77	4.56	4.9	3.10
	10	7.11	4.62	5.33	2.65
	17	6.37	4.43	6.01	1.90

Note: $\bar{Q}_{BP}^{(m)}(W)$ denotes $\bar{Q}_{BP}^{(m)}(s)$ using $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m} = W$. $W = I$ means $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m} = \boldsymbol{I}_m$, $W = D$ means $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m}$ is diagonal, and $W = N$ means the Newey-West type estimator of $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m}$ is used.

where $\mathcal{B}(Y_{t-1}, c)$ denotes the binomial distribution with Y_{t-1} trials and c, $c \in (0,1)$(with the convention $\mathcal{B}(Y_{t-1}, c) = 0$ when $Y_{t-1} = 0$). Specifications (iv) and (v) falls in the class of GLARMA models considered by of Davis et al. (2003). Specification (v) is the parameter driven specification introduced by Zeger (1988). Models (vi) and (vii) are Log-linear Poisson and Negative Binomial autoregression, as considered by Fokianos and Tjostheim (2011). Finally, specifications (ix) and (x) leads to Poisson and Negative Binomial INAR(1) models. Hence, our designs under H_1 cover a wide range of alternatives, including both observation and parameter driven models.

For the power analysis, we only report the rejections with $T = 100$. Table 1.3 present the empirical power our the test statistics for $c = \{-0.5, -0.3, 0.3, 0.5\}$[2]. For $c < 0$, specifications (ix) and (x) are not well defined, and hence, we omit them. Because $\bar{Q}_{BP}^{(m)}(s)$ tend to be oversized when the Newey-West type estimator of $\boldsymbol{A}_{\varepsilon}^{(m)}$ is used with $T = 100$ (but not when $T = 300$), we omit it in the power analysis below. Although

2. Results for other values of c and sample sizes are similar and are available from the author.

$\hat{Q}_{BP}(s)$ also tends to be oversized, we include it in the power comparisons, so one can address the relative power of our proposal.

In general, one can see that the power of our tests decrease with s. $\bar{Q}_{BP}^{(m)}(s)$ is usually more powerful than $\hat{Q}_{BP}(10)$, when s is small (up to 5). The gain in power can be substantial when one compare $\bar{Q}_{BP}^{(m)}(1)$ with $\hat{Q}_{BP}(10)$. On the other hand, when $s = 10$, $\bar{Q}_{BP}^{(m)}(10)$ tends to be slightly less powerful than $\hat{Q}_{BP}(10)$. As expected, as we increase $|c|$, all tests have higher empirical power. When $c < 0$, our tests $\bar{Q}_{BP}^{(m)}(s)$ with $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)}$ being diagonal tends to be more powerful, whereas when $c > 0$, the tests with $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = \boldsymbol{I}_m$ present higher empirical power in the DGP's analyzed. Alternatives of the type (vi) tends to be the hardest to detect, whereas (ix) and (x) tends to be the easiest.

Overall, our simulations show that our proposed tests have very good size and power properties.

1.6 Risk Management and U.S. Corporate Bankruptcies

In order to illustrate the appealing of our proposed test statistics in applied settings, we analyze different specifications of credit risk models.

In a seminal paper, Das et al. (2007) analyze if observable macroeconomic and firm-specific variables are sufficient to explain the default time correlation of U.S. non-financial corporations. Using a test statistic based on the count of defaults in a given period, Das et al. (2007) reject the hypothesis of defaults being conditional independent, suggesting some evidence of excess default clustering. This finding has important implications for practitioners because many popular default risk models rely on the assumption of conditionally independent defaults. Moreover, as shown by Duffie et al. (2009), ignoring such default clustering leads to substantial downward bias on extreme default losses probabilities.

Table 1.3: Empirical power at 5 % significance. T=100.

	c	$\hat{Q}_{BP}^{(m)}(10)$	$\bar{Q}_{BP}^{(m)}(1,I)$	$\bar{Q}_{BP}^{(m)}(2,I)$	$\bar{Q}_{BP}^{(m)}(5,I)$	$\bar{Q}_{BP}^{(m)}(10,I)$	$\bar{Q}_{BP}^{(m)}(1,D)$	$\bar{Q}_{BP}^{(m)}(2,D)$	$\bar{Q}_{BP}^{(m)}(5,D)$	$\bar{Q}_{BP}^{(m)}(10,D)$
	-0.5	78.04	97.23	95.98	85.02	68.54	96.33	93.88	80.35	65.68
(iv)	-0.3	41.18	74.93	63.44	44.02	31.49	72.72	60.18	41.73	33.20
	0.3	45.33	75.85	67.85	48.97	34.67	62.90	50.25	34.34	29.22
	0.5	78.53	95.88	95.40	84.66	67.63	87.74	83.03	64.99	53.31
	-0.5	53.86	93.33	85.79	62.20	42.74	95.83	92.71	77.37	61.03
(v)	-0.3	27.10	63.90	48.38	28.61	19.09	72.96	61.82	42.37	32.62
	0.3	40.88	72.69	63.78	43.84	30.32	49.33	36.86	26.15	24.19
	0.5	55.22	85.37	79.96	61.21	44.01	69.88	60.30	43.37	38.71
	-0.5	33.02	54.24	48.12	35.53	25.62	61.39	53.52	42.71	34.04
(vi)	-0.3	11.66	20.26	15.06	10.24	7.80	25.81	18.78	13.78	12.04
	0.3	11.66	18.00	14.50	9.84	7.46	14.09	10.32	8.53	8.98
	0.5	31.60	50.17	44.14	32.84	24.91	41.19	34.53	25.50	22.55
	-0.5	53.69	83.74	75.04	59.08	44.34	83.51	74.36	59.55	48.17
(vii)	-0.3	23.01	47.18	36.21	24.02	17.07	46.38	35.52	24.27	19.93
	0.3	27.03	50.00	39.53	27.33	20.23	48.25	36.71	26.35	22.32
	0.5	84.95	96.64	94.24	86.16	75.91	96.66	93.58	84.35	75.46
	-0.5	34.41	70.81	56.79	37.92	26.60	78.67	68.11	52.07	41.31
$(viii)$	-0.3	16.56	36.54	25.83	15.95	11.35	45.87	35.49	24.71	19.59
	0.3	23.89	46.07	35.66	24.37	16.92	39.19	28.17	20.06	18.86
	0.5	70.15	90.57	84.85	73.11	61.21	85.59	76.23	60.83	53.00
(ix)	0.3	48.06	76.56	66.44	51.06	39.45	74.15	62.25	46.48	38.41
	0.5	95.45	99.67	99.16	96.84	92.50	99.59	98.92	94.76	88.92
(x)	0.3	52.36	80.36	71.09	56.04	43.55	76.51	64.15	47.51	40.23
	0.5	96.28	99.73	99.34	97.57	94.40	99.47	98.54	94.11	88.28

Note: $\bar{Q}_{BP}^{(m)}(s,W)$ denotes $\bar{Q}_{BP}^{(m)}(s)$ using $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m} = W$. $W = I$ means $\hat{A}_{\hat{\varepsilon}}^{m} = \boldsymbol{I}_m$, $W = D$ means $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{m}$ is diagonal.

In order to overcome such consequences, Duffie et al. (2009) propose to add a common dynamic "frailty" effect on the firms default hazard, that is, an unobserved correlated latent process common to all firms. As an alternative to the duration model of Duffie et al. (2009), Koopman et al. (2011, 2012), using time series count data panel models, propose new estimators for the measurement and forecasting of default probabilities when excess default clustering is present, allowing for a large number of macroeconomic and financial variables, an industry fixed effects and a common frailty effect. Differently than Duffie et al. (2009), which model the frailty effect as continuous-time process, Koopman et al. (2011, 2012) rely on a state space specification, such that the frailty effect is modeled as a Gaussian AR(1). Their results confirms the findings of Das et al. (2007) in the sense that there is some evidence of a correlated frailty effect. However, an important question remain unanswered: Is the AR(1) latent process structure enough to capture all the excess default? In other words, is there any evidence of residuals serial correlation, after including this additional parameter?

Our test for lack of autocorrelation is a valuable tool in order to assess if the proposed model for bankruptcy counts is correctly specified. Within our approach, we are able to test both if there is evidence of excess correlation, and, in case there is, if the usual assumption that considering only an additional AR(1) term is enough to capture the excess of default/bankruptcy correlation. This second hypothesis, to the best of our knowledge, has not been verified so far. This is an important model check since, as pointed out by Koopman et al. (2011), model misspecification can lead to underestimation of corporate risk.

When the interest is on determining adequate economic capital buffers, the focus of the analysis is on aggregate default or bankruptcies rather than on firm specific default. Therefore, a modeling strategy that deals directly with aggregate default counts is a natural alternative from the procedure of Duffie et al. (2009) and Koopman et al. (2011, 2012), in which they first estimate the firms default probability and then aggregate. Such strategy has been adopted by Keenan et al. (1999), Giesecke et al. (2011), Azizpour et al. (2014) and Personal et al. (2014), who directly model the economy-wide default counts using macroeconomic and financial covariates.

With this in mind, using monthly data on bankruptcy filed in the United States Bankruptcy Courts from January 1985 until October 2012, available from UCLA-LoPucki Bankruptcy Research Database [3], we analyze different model specifications for bankruptcy counts. Although bankruptcy data is available since October 1979, we only use data from 1985 onwards, that is, only the period after the "Great Moderation". We do it in order to avoid the presence of well documented structural breaks.

UCLA-LoPucki Bankruptcy Research Database contains data on all large, public company bankruptcy cases filed in the United States Bankruptcy Courts. By large firms, they consider firms which have declared assets of more than \$100 million, measured in 1980 dollars, the year before the firm filled the bankruptcy case. A firm is considered public if they report to the Securities Exchange Commission in the last three years prior to bankruptcy. Following Compustat, although only 22% of the public firms has higher market value than \$100 Million in 2011, these firms represent 70% of total assets and sales of all firms listed, and hence represent an important category of firms. Monthly bankruptcy counts are considered in terms of the month the bankruptcy file was filed.

Macroeconomic and financial monthly data are obtained from the St. Louis Fed online database FRED, see Table 1.4 for a listing of macroeconomic and financial data. These involve business cycle measurements, labor market conditions, interest rate and credit spread and are typically used in macro stress test - see Tarullo (2010) for instance. Variables are expressed as yearly growth rates (INDPRO, PERMIT, PPIFGS and PPIENG) or as yearly differences (UNRATE, BAA, FEDFUNDS, GS10, SP500RET and SP500VOL). We also consider a time dummy which takes value one after September 2005, in order to capture the effect of a major bankruptcy law reform, the Bankruptcy Abuse Prevention and Consumer Protection Act (BAPCPA), signed in October 2005. Keenan et al. (1999), Das et al. (2007), Duffie et al. (2009), Giesecke et al. (2011) and Koopman et al. (2011, 2012) have used similar covariates as ours, in related contexts. We denote this vector of covariates by $\boldsymbol{X}_t$.

3. Available at http://lopucki.law.ucla.edu/

Table 1.4: Macroeconomic Time Series data

Variable	Shortname
Industrial production index	INDPRO
New housing permits	PERMIT
Civilian unemployment rate	UNRTAE
Moody's Seasoned Baa Corporate Bond Yield	BAA
10-Year Treasury Constant Maturity Rate	GS10
Federal Funds Rate	FEDFUNDS
Producer Price Index: Finished Goods	PPIFGS
Producer Price Index: Fuels and related energy	PPIENG
S&P 500 yearly returns	SP500RET
S&P 500 return volatility	SP500VOL
2005 Bankruptcy Act	DUMMY2005

All models are estimated using the Poisson quasi-likelihood in (1.10). For each specification considered, AIC and BIC values are provided. For checking the fit of the models, we use the Box-Pierce tests based on the transformed multiplicative residuals autocorrelation, $\bar{Q}_{BP}^{(m)}(s)$, with s equal to 1, 2, 5, 10, 15 and 20. These choices include all the usual lag choices in similar applications supported by our simulations, given that $T = 324$. We report the analysis with $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)}$ being the identity matrix and diagonal, and m is set to $T - 1$, its maximum possible value. As illustrated by our simulations, this choice leads to tests with good size and power properties. In addition, we also use our omnibus test $\bar{T}^{(m)}$. The results of the test statistic are presented on Table 1.5. Estimated parameters for the different models are reported on Table 1.6.

In order to assess if including macroeconomic and financial covariates is enough to capture the linear dynamics in the bankruptcy data, we first consider the "static" model

$$\lambda_t = \exp\left(\boldsymbol{X}_t'\boldsymbol{\beta}\right),$$

which includes only covariates (Model (1) on Tables 1.5 and 1.6). From the specification tests presented on Table 1.5 one can conclude that this simple static Poisson model is strongly rejected using the recursive Portmanteau test statistic $\bar{Q}_{BP}^{(m)}(s)$ for any choice of s, and by the omnibus test $\bar{T}^{(m)}$. Although we use an economy-wide rather than a firm-level approach as Duffie et al. (2009) and Koopman et al. (2011, 2012), our first conclusion points out to the same directions as theirs: there is evidence of a bankruptcy cluster beyond the one implied by the macroeconomic variables. As pointed out by these authors, ignoring such excess autocorrelation can lead to mismeasures on risk management, specially underestimation of extreme loss given default.

To understand better the source of the rejection of the null for the "static" model, we consider the analysis of individual projected residuals autocorrelations for lags up to 20, with $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)}$ being diagonal, and $m = T - 1$. Recall that transformed autocorrelations can be correctly compared with the usual $\pm 2/\sqrt{T}$ confidence band, as when working with raw data. In Figure 1.1, we have plotted the autocorrelograms of the transformed residuals of the model only with covariates. In this plot we can identify the source of the rejection, since transformed autocorrelations provide evidence on serial correlation of the underlying innovation from the very first lag onwards.

Once the simple "static" model is reject, we consider two popular richer classes of dynamic count data models: the Davis et al. (2003)'s GLARMA and Fokianos and Tjøstheim (2011)'s Poisson log-linear autoregression models. The GLARMA(p, q) specification is given by

$$\begin{aligned}
\lambda_t &= \exp\left(\boldsymbol{X}_t' \boldsymbol{\beta} + Z_t\right), \\
Z_t &= \sum_{i=1}^{p} \phi_i \left(Z_{t-i} + \left(\frac{Y_{t-i} - \lambda_{t-i}}{\lambda_{t-i}} \right) \right) + \sum_{i=1}^{q} \theta_i \left(\frac{Y_{t-i} - \lambda_{t-i}}{\lambda_{t-i}} \right),
\end{aligned}$$

that is, Z_t follows an ARMA(p, q) structure. On the other hand, the Poisson log-linear autoregression (p, q) specification (henceforth, Ploglin(p, q)) is given by

$$\begin{aligned}
\lambda_t &= \exp\left(v_t\right) \\
v_t &= \boldsymbol{X}_t' \boldsymbol{\beta} + \sum_{i=1}^{p} \phi_i \ln\left(1 + Y_{t-i}\right) + \sum_{i=1}^{q} \theta_i v_{t-i}.
\end{aligned}$$

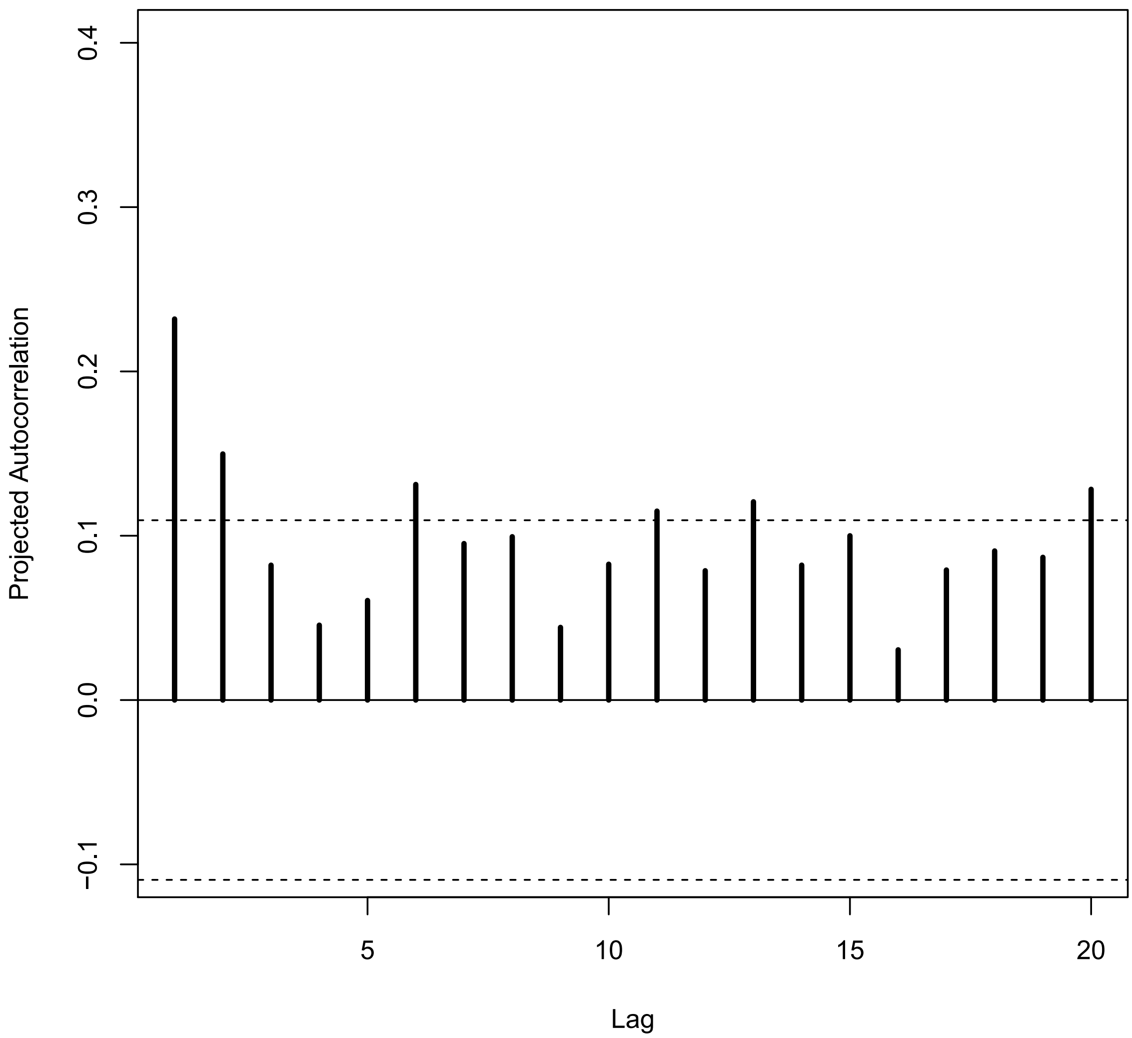


Figure 1.1: Projected residual sample autocorrelations from Poisson regression model only with macroeconomic covariates, with diagonal $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}$, and $m = T - 1 = 333$. Dashed lines show 95% confidence bands.

Table 1.5: Goodness-of-Fit Analysis for U.S.A. Bankruptcy Counts based on differen model specifications

Models	$\hat{\boldsymbol{A}}_t^m$	$\bar{Q}_{BP}^{(m)}(1)$	$\bar{Q}_{BP}^{(m)}(2)$	$\bar{Q}_{BP}^{(m)}(5)$	$\bar{Q}_{BP}^{(m)}(10)$	$\bar{Q}_{BP}^{(m)}(15)$	$\bar{Q}_{BP}^{(m)}(20)$	$\bar{T}^{(m)}$
Model (1)	I_m	26.872***	34.438***	39.675***	57.958***	75.670***	87.041***	3.018***
	$diag$	17.963***	25.455***	29.624***	44.637***	61.576***	74.744***	2.095***
Model (2)	I_m	0.000	0.558	0.727	17.833*	28.817**	33.424**	0.072
	$diag$	0.000	0.762	0.888	11.788	24.016*	29.778*	0.062
Model (3)	I_m	0.963	1.179	2.001	15.782	24.052*	26.573	0.151
	$diag$	1.134	1.365	2.228	12.060	20.240	23.445	0.161
Model (4)	I_m	0.054	0.935	1.094	4.172	14.721	17.391	0.045
	$diag$	0.077	1.291	1.440	4.538	15.474	18.700	0.057
Model (5)	I_m	0.980	1.119	3.190	8.495	16.851	19.255	0.137
	$diag$	1.236	1.396	3.517	7.309	15.209	18.064	0.162

Note: ** and ***denotes significant at 5 and 1% level. $m = T - 1$. Model (1) is the model only with covariates, Model (2) is the GLARMA (1,0), Model (3) is the Poisson log-linear autoregression (Ploglin) of order 1, Model (4) is the Glarma ((1,6),0) and Model (5) is the Ploglin of order 1 and 6.

Table 1.6: Estimated parameters of Poisson models for U.S.A. Bankruptcy Counts

COVARIATES	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
INTERCEPT	1.263***	1.281***	0.408***	1.246***	0.145
Dummy2005	-0.303*	-0.242	-0.139**	-0.216	-0.130**
BAA	0.270***	0.173*	0.141***	0.167	0.186***
FEDFUNDS	-0.015	-0.034	-0.029	-0.018	-0.013
GS10	-0.11	-0.047	-0.067**	-0.073	-0.111***
INDPRO	-0.063***	-0.084***	-0.034***	-0.090***	-0.029***
PERMIT	0.017***	0.012***	0.008***	0.010**	0.007***
PPIENG	0.028***	0.026***	0.018***	0.017***	0.012***
PPIFGS	-0.112***	-0.113***	-0.074***	-0.079**	-0.44**
SP500RET	-0.003	-0.004	-0.002*	-0.003	-0.001
SP500VOL	-0.001	0.000	-0.000	0.000	0.000
UNRATE	0.341***	0.171	0.148***	0.123	0.111***
AR.1		0.334***	0.572***	0.348***	0.526***
AR.6				0.172***	0.227***
BIC	1361.759	1306.185	1292.792	1294.211	1285.451
AIC	1316.025	1256.679	1243.246	1244.666	1232.095

Note: *,**,*** denote significant at 10%, 5% and 1% level. Model (1) is the model only with covariates, Model (2) is the GLARMA (1,0), Model (3) is the Poisson log-linear autoregression (Ploglin) of order 1, Model (4) is the Glarma ((1,6),0) and Model (5) is the Ploglin of order 1 . and 6.

Notice that in the GLARMA(p, q) model, the additional dynamic is modeled via Z_t, whereas in the Ploglin(p, q) model it is modeled via lags of $\ln(1 + Y_t)$ and/or v_t. A potential advantage of using a Ploglin model is that the coefficients associated with $\ln(1 + Y_t)$ may be interpreted as a "contagion effect". On the other hand, the coefficients in the GLARMA model have a less clear interpretation. Despite this interpretability issue, these two classes of models are non-nested and have their own merits. Given that our goal is to check if a given parametric model is able to capture the dynamics in the bankruptcy data, it is worth to evaluate the performance of both GLARMA and Ploglin models.

To assess if only first order dynamics is enough to capture the excess of bankruptcy correlation, we consider GLARMA and Ploglin specifications with p equal to one[4] (Model (2) and (3), respectively, on Tables 1.5 and 1.6). Once we apply our tests $\bar{Q}_{BP}^{(m)}(s)$ and $T^{(m)}$ to the residuals of these dynamic models, the evidence of residual serial correlation

4. The result for the GLARMA model with $q = 1$ is very similar and therefore we ommit it. For the Ploglin, the specification with $q = 1$ and $p = 0$ is numerically unstable.

is mixed.

First, we do not find any evidence against H_0 when $T^{(m)}$ or $\bar{Q}_{BP}^{(m)}(s)$ is used with s smaller or equal than 5. When s is grater than 5, we find some evidence of residual serial autocorrelation from GLARMA residuals, specially when $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = \boldsymbol{I}_m$. When $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)}$ is diagonal, the evidence against H_0 is weaker. For the residuals from Ploglin model of order 1, we only find evidence against H_0 when one uses $\bar{Q}_{BP}^{(m)}(15)$ with $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = \boldsymbol{I}_m$.

To better understand this potentially conflicting findings, Figure 1.2 presents the plot the autocorrelograms of the transformed residuals of both dynamic models. From the plot, one can see that the residual autocorrelation at the sixth lag is statistically significant, corroborating the evidence against H_0. Once we fit a GLARMA or a Poisson log-linear autoregressive model with p equal to 1 and 6 (Model (4) and (5), respectively, on Tables 1.5 and 1.6), we fail to reject H_0 using any of our proposed tests.

Overall, our results provide some evidence that, within the class of dynamic count data models we have analyzed, it may be necessary to couple macroeconomic and financial variables with higher order autocorrelation terms in order to entirely capture the bankruptcy clustering present in the data.

1.7 Conclusion

In this paper, we have proposed new distribution-free tests for lack of autocorrelation in count data models in the presence of estimated parameters, under relatively weak assumptions. Both portmanteau and omnibus type tests are considered, but contrary to classical tests, our test statistics are able to detect local alternatives converging to the null at the parametric rate. Our tests present satisfactory finite sample properties as demonstrated via Monte Carlo simulations. Once our proposal is applied to bankruptcy count models, we rejected the specification of a model with only macroeconomic covariates. When we include an additional AR(1) term, the evidence against H_0 is weaker, but we still find that considering higher order lags might be more appropriate when modelling bankruptcy data.

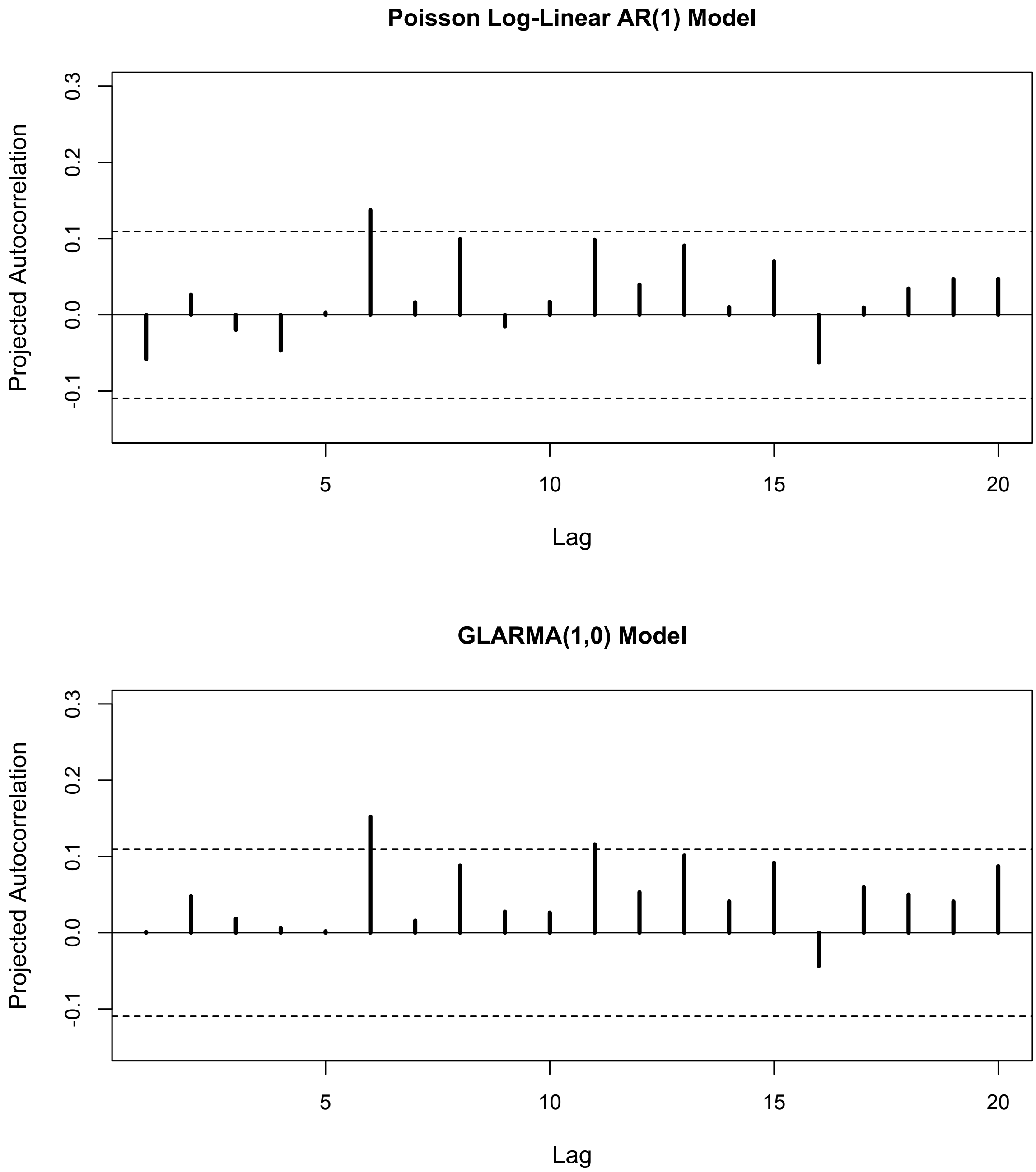


Figure 1.2: Projected residual sample autocorrelations from dynamic models with diagonal $\hat{\boldsymbol{A}}_{\hat{\varepsilon}}$, and $m = T - 1 = 333$. Dashed lines show 95% confidence bands. Top panel: Residual from the Poisson log-linear AR(1) model. Botton panel: Residuals from GLARMA(1,0) model.

Our empirical results have important implications for risk management. Many industry credit risk models, such as CreditMetrics, Moody's KMV and CreditRisk+ rely on the assumption that default and bankruptcies are (conditionally) time independent, see e.g. Keenan et al. (1999), Gordy (2000, 2003), Frey and McNeil (2002), Das et al. (2007)

and Duffie et al. (2009). However, from the results of our specification tests, we conclude that there is evidence of an excess bankruptcy clustering when only macroeconomic and financial variables are included in the model. The presence of residual autocorrelation may increase bankruptcy rate volatility, and as result it may shift probability mass of an portfolio credit loss distribution toward more extreme values. This would increase capital buffers prescribed by the risk models. Hence, if one ignores the presence of a frailty and/or contagion effect, portfolio credit risk models will tend to be wrong. On the other hand, if one considers dynamic count models with appropriate lag structure, such as the GLARMA and Poisson log-linear autoregression models with $AR(1, 6)$ terms as we have analyzed in this paper, it seems that there is no evidence of model misspecification. This way, we argue that this richer classes of models could be more appropriate to model bankruptcies, and adjusting the credit risk models to comply with these specifications could not only be relevant for internal risk assessment, but also for external supervision of financial institutions.

Although we have focused on dynamic count models, our specification tests can be applied to other situations in which a multiplicative error model is suitable. For example, one can check the correct specification of ACD models, GARCH and stochastic volatility models. Also, our tests can be easily adapted to assess if there is evidence of serial autocorrelation on the *Pearson* residuals, instead of the multiplicative ones we have considered.

Finally, our theoretical results can be extended to other interesting setups. For instance, regarding the choice of the number of lags s included in the portmanteau test statistic, one can adopt a data-driven procedure based on an AIC/BIC criterion in the lines of Escanciano and Lobato (2009) and Escanciano et al. (2013), at the cost of not being able to detect the kind of local alternatives beyond the first lag, as considered here. Coupling the results in this paper with those of Escanciano and Lobato (2009) and Escanciano et al. (2013), one may be able to proposing a data-driven distribution-free portmanteau test, without requiring the residuals to be martingale difference.

Another interesting extension is to relax the assumption that $\sum_{j=1}^{\infty} w(j) < \infty$ used in our omnibus tests. This way, one could use the weights in Box-Pierce and Ljung-Box tests, for instance, but allowing s to be large. In cases where the weights are not summable, it becomes necessary to normalize the test statistics to obtain convergence in distribution, see e.g. Hong (1996). By following similar steps as in Hong (1996), one may provide larger classes of distribution-free omnibus tests. We leave these extensions for future research.

1.8 Appendix

In this appendix, we present sufficient assumptions for the proof of our results.

In the following, we assume the fairly mild regularity conditions:

Assumption 1.1 *Assume that an estimator $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$ is available, such that*

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + O_p(T^{-1/2})$$

and

$$\hat{\boldsymbol{A}}_{\hat{\varepsilon}}^{(m)} = \boldsymbol{A}_{\varepsilon}^{(m)} + o_p(1).$$

Assumption 1.2 *$(Y_t, \mathbf{X}_t', \varepsilon_t)'$ is strictly stationary, ε_t has mean 1, $E\left[\varepsilon_t^{4+2\delta}\right] < \infty$, for some $\delta > 0$, and $(Y_t, \boldsymbol{X}_t', \varepsilon_t)'$ is strong mixing with coefficients α_j satisfying $\sum_{j=1}^{\infty} \alpha_j^{\delta/(2+\delta)} < \infty$, where,*

$$\alpha_j = \sup_{A,B} |\Pr(AB) - P(A)P(B)|$$

and A and B vary over events in the σ-fields generated by $\{(Y_t, \mathbf{X}_t', \varepsilon_t)', t \leq 0\}$ and $\{(Y_t, \mathbf{X}_t', \varepsilon_t)', \, t \geq j\}$.

Assumption 1.3 *(i) The function $\lambda_t(\cdot) = \lambda(Y_{t-1}, \lambda_{t-1}, \mathbf{X}_t, \ldots; \boldsymbol{\beta})$ is twice continuously differentiable with respect to $\boldsymbol{\beta} \in \boldsymbol{\Theta}$ a.s., with $E||\frac{\varepsilon_t}{\lambda_t}\frac{\partial \lambda_t}{\partial \boldsymbol{\beta}}||^{4+2\delta} < \infty$, for some $\delta > 0$.*

(ii) Let Θ_0 be a small convex neighborhood of $\boldsymbol{\beta}_0$, and then

$$E \sup_{\beta \in \Theta_0} \left\| \frac{\varepsilon_t}{\lambda_t^2} \frac{\partial \lambda_t}{\partial \boldsymbol{\beta}} \frac{\partial \lambda_t}{\partial \boldsymbol{\beta}'} \right\| + E \sup_{\beta \in \Theta_0} \left\| \frac{\varepsilon_t}{\lambda_t} \frac{\partial^2 \lambda_t}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right\| < \infty.$$

Assumption 1.4 *For $m > k$,*

$$\sum_{l=1+m-k}^{m} \boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)$$

is positive definite.

Assumption 1.1 is a re-statement of conditions (1.8) and (1.9) discussed in the main text. Assumptions 1.2 and 1.3 are standard in the literature, see e.g. Romano and Thombs (1996), Lobato et al. (2002), Francq et al. (2005) and Delgado and Velasco (2011) for similar assumptions. Assumption 1.2 is about the data generating process, where we assume a mixing condition to justify a central limit theorem for the autocovariances of the multiplicative errors. If Y_t follows standard count data distributions such as Poisson and Negative Binomial, the existence of all moments is guaranteed. In this case, the assumption of having finite $4 + 2\delta$ moments would not be restrictive. Assumption 1.4 is a technical assumption needed for computing the recursive residuals.

PROOF OF PROPOSITION 1

It follows from Taylor expansion around $\boldsymbol{\beta}_0$, element by element. For each $j = 1, \ldots, m$, we write

$$\hat{\rho}_{\hat{\varepsilon}}(j) - \hat{\rho}_{\varepsilon}(j) = \frac{\partial \hat{\rho}_{\varepsilon}(j)}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + D_T(j)$$

where $D_T(j)$ is

$$D_T(j) = \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' \ddot{\boldsymbol{\rho}}_{\varepsilon^*}(j) \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right),$$

$$\ddot{\boldsymbol{\rho}}_{\varepsilon*}(j) = \left.\frac{\partial^2 \hat{\rho}_{\varepsilon}(j)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_j^*}$$

and $\boldsymbol{\beta}_j^*$ are such that $||\boldsymbol{\beta}_j^* - \boldsymbol{\beta}_0|| \leq ||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||$, $\forall j = 1, \ldots, m$.

Then, for each $j = 1, \ldots, m$,

$$\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\rho}_{\varepsilon}(j) = \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\gamma}_{\varepsilon}(j)}{\hat{\gamma}_{\varepsilon}(0)} - \frac{\hat{\gamma}_{\varepsilon}(j)}{\hat{\gamma}_{\varepsilon}(0)} \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\gamma}_{\varepsilon}(0)}{\hat{\gamma}_{\varepsilon}(0)}. \tag{1.22}$$

Using that $\hat{\gamma}_\varepsilon(j) = \gamma_\varepsilon(j) + o_p(1)$, in particular $\gamma_\varepsilon(j) = 0$ for $j \neq 0$ under H_0 and that

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} \gamma_\varepsilon(0) &= \frac{-2}{T} \left(\sum_{t=1}^{T} \frac{\varepsilon_t}{\lambda_t} (\varepsilon_t - 1) \frac{\partial \lambda_t}{\partial \boldsymbol{\beta}} \right) \\
&= O_P(1) \qquad (1.23)
\end{aligned}
$$

under Assumptions (1.2) and (1.3) and Law of Large Numbers, we conclude that the normalization of $\hat{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}$ has no asymptotic effect under H_0, so that

$$
\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\boldsymbol{\rho}}_\varepsilon(j) = \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \gamma_\varepsilon(j)}{\gamma_\varepsilon(0)} + o_p(1).
$$

Without loss of generality, assume that $\gamma_\varepsilon(0) = 1$. Writing now

$$
\begin{aligned}
\frac{\partial \gamma_\varepsilon(j)}{\partial \boldsymbol{\beta}} &= \frac{-1}{T} \left(\sum_{t=\tau+1}^{T} \frac{\varepsilon_t}{\lambda_t} (\varepsilon_{t-\tau} - 1) \frac{\partial \lambda_t}{\partial \boldsymbol{\beta}} + \frac{\varepsilon_{t-\tau}}{\lambda_{t-\tau}} (\varepsilon_t - 1) \frac{\partial \lambda_{t-\tau}}{\partial \boldsymbol{\beta}} \right) \\
&= -\mathbf{A}_{T,1} - \mathbf{A}_{T,2}.
\end{aligned}
$$

Setting $\boldsymbol{\zeta}_\varepsilon^{(i)}(j) := \lim_{T\to\infty} E[\mathbf{A}_{T.i}(j)], i = 1, 2$, we wish to show that $\mathbf{A}_{T,i}(j) = \boldsymbol{\zeta}_\varepsilon^{(i)}(j) + o_p(1), i = 1, 2;\ j = 1, 2, 3, \ldots$.

It suffices to show that $E \left\| \mathbf{A}_{T,i}(j) - E\left[\mathbf{A}_{T,i}(j)\right] \right\|^2$ is $o(1)$, $i = 1, 2$. First consider $i = 1$,

$$
E \left\| \boldsymbol{A}_{T,i}(j) - E\left[\boldsymbol{A}_{T,i}(j)\right] \right\|^2 = \frac{1}{T^2} \sum_{t=j+1}^{T} \sum_{r=j+1}^{T} E\left[e(t, t-j)' e(r, r-j)\right] = o(1)
$$

where $e(t, t-j) = (\varepsilon_t/\lambda_t)(\varepsilon_{t-\tau} - 1)\, \partial\lambda_t/\partial\boldsymbol{\beta} - E\left[(\varepsilon_t/\lambda_t)(\varepsilon_{t-\tau} - 1)\, \partial\lambda_t/\partial\boldsymbol{\beta}\right]$ and, henceforth we omit dependence on $\boldsymbol{\beta}_0$ in the notation.

For some $n > 0$ fixed with $T, E \left\| \mathbf{A}_{T,1}(j) - E\left[\mathbf{A}_{T,1}(j)\right] \right\|^2$ is

$$
\begin{aligned}
&\frac{1}{T^2} \sum_{t=j+1}^{T} E\left[e(t, t-j)' e(t, t-j)\right] + \frac{2}{T^2} \sum_{t=j+1}^{T} \sum_{t-n-j \le r < t}^{T} E\left[e(t, t-j)' e(r, r-j)\right] \qquad (1.24) \\
&\qquad + \frac{2}{T^2} \sum_{t=j+1}^{T} \sum_{j+1 \le r < t-n-j}^{T} E\left[e(t, t-j)' e(r, r-j)\right].
\end{aligned}
$$

The first two terms of (1.24) are $O(T^{-1}) = o(1)$ since it involves a maximum of $T + n$

terms with bounded absolute expectation, since by Assumptions 1.2-1.3 and Cauchy-Schwarz inequalities,

$$\begin{aligned}
& E\left\|\frac{\varepsilon_t}{\lambda_t}\left(\varepsilon_{t-\tau}-1\right)\frac{\partial\lambda_t}{\partial\boldsymbol{\beta}}-E\left[\frac{\varepsilon_t}{\lambda_t}\left(\varepsilon_{t-\tau}-1\right)\frac{\partial\lambda_t}{\partial\boldsymbol{\beta}}\right]\right\|^2 \\
& \leq 2E\left\|\frac{\varepsilon_t}{\lambda_t}\left(\varepsilon_{t-\tau}-1\right)\frac{\partial\lambda_t}{\partial\boldsymbol{\beta}}\right\|^2 \\
& \leq 2E\left\|\frac{\varepsilon_t}{\lambda_t}\frac{\partial\lambda_t}{\partial\boldsymbol{\beta}}\right\|^4 E\left|(\varepsilon_t-1)\right|^4<\infty.
\end{aligned}$$

In order to show that the third term of (1.24) is bounded, notice that $e(r,r-j)$ is $\mathcal{F}_1^r$ measurable and that $e(t,t-j)$ is $\mathcal{F}_t^\infty$ measurable. Given Assumption 1.3, $E\left\|e(t,t-j)\right\|^{2+\delta}<\infty$, $E\left\|e(r,r-j)\right\|^{2+\delta}<\infty$, we can use Roussas and Ioannides (1987) moment inequality to show that the third term of (1.24) is bounded in absolute value by

$$\frac{C}{T^2}\left(E\left\|e(t,t-j)\right\|^{2+\delta}E\left\|e(r,r-j)\right\|^{2+\delta}\right)^{2+\delta}\sum_{t=j+1}^{T}\sum_{j+1\leq r<t-n-j}^{T}\alpha_{t-j-r}^{\frac{\delta}{2+\delta}}=O(T^{-1})=o(1).$$

Using exactly the same procedure, we can show that $E\left\|\mathbf{A}_{T,2}(j)-E\left[\mathbf{A}_{T,2}(j)\right]\right\|^2$ is $o(1)$. Then, we have that, under H_0,

$$\frac{\partial\gamma_\varepsilon(j)}{\partial\boldsymbol{\beta}}=-E\left(\frac{\varepsilon_t}{\lambda_t}\left(\varepsilon_{t-\tau}-1\right)\frac{\partial\lambda_t}{\partial\boldsymbol{\beta}}\right)-E\left(\frac{\varepsilon_{t-\tau}}{\lambda_{t-\tau}}\left(\varepsilon_t-1\right)\frac{\partial\lambda_{t-\tau}}{\partial\boldsymbol{\beta}}\right)+o_p\left(1\right). \tag{1.25}$$

Now, we just need to show that the second order term on the expansion is $o_p(T^{-1/2})$. In order to do that, we just need to show that $\ddot{\boldsymbol{\rho}}_{\varepsilon^*}(j)=\left(\partial^2\hat{\rho}_\varepsilon(j)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'\right)|_{\boldsymbol{\beta}=\boldsymbol{\beta}_j^*}$ is $O_p(1)$. For $j=1,\ldots,m$ we have

$$\ddot{\rho}_{\varepsilon^*}(j)=\frac{\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\hat{\gamma}_{\varepsilon^*}(j)}{\hat{\gamma}_{\varepsilon^*}(0)}-\frac{\frac{\partial}{\partial\boldsymbol{\beta}}\hat{\gamma}_{\varepsilon^*}(j)}{\hat{\gamma}_{\varepsilon^*}(0)}\frac{\frac{\partial}{\partial\boldsymbol{\beta}'}\hat{\gamma}_{\varepsilon^*}(j)}{\hat{\gamma}_{\varepsilon^*}(0)}-\frac{\partial}{\partial\boldsymbol{\beta}'}\hat{\rho}_{\varepsilon^*}(j)\frac{\frac{\partial}{\partial\boldsymbol{\beta}}\hat{\gamma}_{\varepsilon^*}(0)}{\hat{\gamma}_{\varepsilon^*}(0)}$$

$$-\hat{\rho}_{\varepsilon^*}(j)\left(\frac{\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\hat{\gamma}_{\varepsilon^*}(0)}{\hat{\gamma}_{\varepsilon^*}(0)}-\frac{\frac{\partial}{\partial\boldsymbol{\beta}}\hat{\gamma}_{\varepsilon^*}(0)}{\hat{\gamma}_{\varepsilon^*}(0)}\frac{\frac{\partial}{\partial\boldsymbol{\beta}'}\hat{\gamma}_{\varepsilon^*}(0)}{\hat{\gamma}_{\varepsilon^*}(0)}\right),$$

where

$$\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\hat{\gamma}_{\varepsilon^*}(j) = -\frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{A}_{T,1}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} - \frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{A}_{T,2}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*}$$

and

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{A}_{T,1}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} &= -\frac{1}{T}\sum_{t=1+\tau}^{T} 2\frac{\varepsilon_t^*}{\lambda_t^{*2}}\left(\varepsilon_{t-\tau}^* - 1\right)\frac{\partial\lambda_t^*}{\partial\boldsymbol{\beta}}\frac{\partial\lambda_t^*}{\partial\boldsymbol{\beta}'} \\ &- \frac{1}{T}\sum_{t=1+\tau}^{T}\frac{\varepsilon_t^*}{\lambda_t^*}\frac{\varepsilon_{t-\tau}^*}{\lambda_{t-\tau}^*}\frac{\partial\lambda_t^*}{\partial\boldsymbol{\beta}}\frac{\partial\lambda_{t-\tau}^*}{\partial\boldsymbol{\beta}'} \\ &+ \frac{1}{T}\sum_{t=1+\tau}^{T}\frac{\varepsilon_t^*}{\lambda_t^*}\left(\varepsilon_{t-\tau}^* - 1\right)\frac{\partial^2\lambda_t^*}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\beta}'}\boldsymbol{A}_{T,2}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}^*} &= -\frac{1}{T}\sum_{t=1+\tau}^{T} 2\frac{\varepsilon_{t-\tau}^*}{\lambda_{t-\tau}^{*2}}\left(\varepsilon_t^* - 1\right)\frac{\partial\lambda_{t-\tau}^*}{\partial\boldsymbol{\beta}}\frac{\partial\lambda_{t-\tau}^*}{\partial\boldsymbol{\beta}'} \\ &- \frac{1}{T}\sum_{t=1+\tau}^{T}\frac{\varepsilon_t^*}{\lambda_t^*}\frac{\varepsilon_{t-\tau}^*}{\lambda_{t-\tau}^*}\frac{\partial\lambda_{t-\tau}^*}{\partial\boldsymbol{\beta}}\frac{\partial\lambda_t^*}{\partial\boldsymbol{\beta}'} \\ &+ \frac{1}{T}\sum_{t=1+\tau}^{T}\frac{\varepsilon_{t-\tau}^*}{\lambda_{t-\tau}^*}\left(\varepsilon_t^* - 1\right)\frac{\partial^2\lambda_{t-\tau}^*}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\end{aligned}$$

and $\lambda_s^* = \lambda\left(Y_{t-1}, \lambda_{t-1}, \mathbf{X}_t, \ldots; \boldsymbol{\beta}_\tau^*\right)$, and $\varepsilon_t^* = Y_t/\lambda_t^*$. Then, using Assumptions (1.2) and (1.3), we find that $E\sup_{\boldsymbol{\beta}}\|\ddot{\boldsymbol{\rho}}_{\varepsilon^*}(j)\| < \infty$, so that $\ddot{\boldsymbol{\rho}}_{\varepsilon^*}(j) = O_p(1)$. ■

PROOF OF THEOREM 1

Using algebra and Proposition 1.1, we find that $\hat{\mathcal{L}}^{(m)}\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)} = \hat{\mathcal{L}}^{(m)}\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)} + o_p(T^{-1/2})$, because from Assumption 1.4,

$$\hat{\kappa}_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}\right] = \hat{\kappa}_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right] + \left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\right) + o_p(T^{-1/2}),$$

$\tau = 1, \ldots, m-k$, such that $\hat{\kappa}_{\tau+1}[\tilde{\boldsymbol{\rho}}] = \left(\sum_{l=\tau+1}^m \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(l)'\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(l)\right)^{-1}\sum_{l=\tau+1}^m \hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(l)'\rho(l)$ and $\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(\tau) \to_p \boldsymbol{\xi}_{\varepsilon}(\tau)$, which can be proved using the same methods used in the proof of Proposition 1.1.

Similar, we can show that $\hat{\mathcal{L}}^{(m)}\tilde{\rho}_{\varepsilon}^{(m)}(\tau) = \tilde{\rho}_{\varepsilon}^{(m)}(\tau) - \boldsymbol{\xi}_{\varepsilon}(\tau)\kappa_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right] + o_p(T^{-1/2})$, where $\kappa_{\tau+1}[\boldsymbol{\rho}] = \left(\sum_{l=\tau+1}^m \boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)\right)^{-1}\sum_{l=\tau+1}^m \boldsymbol{\xi}_{\varepsilon}(l)'\rho(l)$, $\tau = 1, \ldots, m-k$.

The CLT for $\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ follows from the CLT for $\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ under Assumptions 1.2, condition (1.4), H_0 and from the fact that $\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ are standardized by construction if $\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ is already standardized.

Under H_0, $\tilde{\rho}_{\varepsilon}^{(m)}(\tau)=0$ for all $\tau=1,2\ldots$,and hence $\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ has asymptotic mean equal to 0. In order to show the asymptotic variance of $\bar{\boldsymbol{\rho}}_{\varepsilon}^{(m)}$ is equal to $\boldsymbol{I}_{m-k}$, notice that $AVar\left(T^{1/2}\hat{\mathcal{L}}^{(m)}\tilde{\rho}_{\varepsilon}^{(m)}(\tau)\right)$ is equal to

$$\begin{aligned}AVAR\left(T^{1/2}\left(\tilde{\rho}_{\varepsilon}^{(m)}(\tau)-\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(\tau)\hat{\kappa}_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}\right]\right)\right)&=AVAR\left(T^{1/2}\left(\tilde{\rho}_{\varepsilon}^{(m)}(\tau)-\boldsymbol{\xi}_{\varepsilon}(\tau)\kappa_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right]\right)\right)\\&=1+\boldsymbol{\xi}_{\varepsilon}(\tau)\left(\sum_{l=\tau+1}^{m}\boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)\right)^{-1}\boldsymbol{\xi}_{\varepsilon}(\tau)',\end{aligned}$$

while for $1\leq\tau<q\leq m-k$, $ACov\left(T^{1/2}\hat{\mathcal{L}}^{(m)}\tilde{\rho}_{\varepsilon}^{(m)}(\tau),T^{1/2}\hat{\mathcal{L}}^{(m)}\tilde{\rho}_{\varepsilon}^{(m)}(q)\right)$ is given by

$$\begin{aligned}&ACov\left(T^{1/2}\left(\tilde{\rho}_{\varepsilon}^{(m)}(\tau)-\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(\tau)\hat{\kappa}_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}\right]\right),T^{1/2}\left(\tilde{\rho}_{\varepsilon}^{(m)}(q)-\hat{\boldsymbol{\xi}}_{\hat{\varepsilon}}(q)\hat{\kappa}_{q+1}\left[\tilde{\boldsymbol{\rho}}_{\hat{\varepsilon}}^{(m)}\right]\right)\right)\\&=ACov\left(T^{1/2}\left(\tilde{\rho}_{\varepsilon}^{(m)}(\tau)-\boldsymbol{\xi}_{\varepsilon}(\tau)\kappa_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right]\right),T^{1/2}\left(\tilde{\rho}_{\varepsilon}^{(m)}(q)-\boldsymbol{\xi}_{\varepsilon}(q)\kappa_{q+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right]\right)\right)\\&=ACov\left(T^{1/2}\tilde{\rho}_{\varepsilon}^{(m)}(\tau),T^{1/2}\tilde{\rho}_{\varepsilon}^{(m)}(q)\right)-ACov\left(T^{1/2}\tilde{\rho}_{\varepsilon}^{(m)}(\tau),T^{1/2}\boldsymbol{\xi}_{\varepsilon}(q)\kappa_{q+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right]\right)\\&-ACov\left(T^{1/2}\boldsymbol{\xi}_{\varepsilon}(\tau)\kappa_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right],T^{1/2}\tilde{\rho}_{\varepsilon}^{(m)}(q)\right)\\&+ACov\left(T^{1/2}\boldsymbol{\xi}_{\varepsilon}(\tau)\kappa_{\tau+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right],T^{1/2}\boldsymbol{\xi}_{\varepsilon}(q)\kappa_{q+1}\left[\tilde{\boldsymbol{\rho}}_{\varepsilon}^{(m)}\right]\right)\end{aligned}$$

where the terms are respectively equal to $0,0,$ $-\boldsymbol{\xi}_{\varepsilon}(\tau)\left(\sum_{l=\tau+1}^{m}\boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)\right)^{-1}\boldsymbol{\xi}_{\varepsilon}(q)'$ and $\boldsymbol{\xi}_{\varepsilon}(\tau)\left(\sum_{l=\tau+1}^{m}\boldsymbol{\xi}_{\varepsilon}(l)'\boldsymbol{\xi}_{\varepsilon}(l)\right)^{-1}\boldsymbol{\xi}_{\varepsilon}(q)'$. Hence, the asymptotic covariance of the projection is 0. ■

PROOF OF THEOREM 2

The result follows from noticing that under H_{1T} Proposition 1 is still valid, because for each $j = 1, \ldots, m$, we have

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\rho}_{\varepsilon}(j) &= \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\gamma}_{\varepsilon}(j)}{\hat{\gamma}_{\varepsilon}(0)} - \frac{\hat{\gamma}_{\varepsilon}(j)}{\hat{\gamma}_{\varepsilon}(0)} \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\gamma}_{\varepsilon}(0)}{\hat{\gamma}_{\varepsilon}(0)} \\
&= \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\gamma}_{\varepsilon}(j)}{\hat{\gamma}_{\varepsilon}(0)} - \frac{r(j)}{\sqrt{T}} O_p(1) + O_p(T^{-1}) \\
&= \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\gamma}_{\varepsilon}(j)}{\hat{\gamma}_{\varepsilon}(0)} + O_p(T^{-1/2}),
\end{aligned}$$

and then, for each $j = 1, \ldots, m$, we have

$$\begin{aligned}
\hat{\rho}^{(m)}_{\hat{\varepsilon}}(j) - \hat{\rho}^{(m)}_{\varepsilon}(j) &= \frac{\frac{\partial}{\partial \boldsymbol{\beta}} \hat{\gamma}_{\varepsilon}(j)}{\hat{\gamma}_{\varepsilon}(0)} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + O_p(T^{-1/2})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(T^{-1/2}) \\
&= \frac{\frac{\partial}{\partial \beta} \hat{\gamma}_{\varepsilon}(j)}{\hat{\gamma}_{\varepsilon}(0)} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(T^{-1/2}).
\end{aligned}$$

Hence, from Theorem 1, we have that

$$\begin{aligned}
\hat{\boldsymbol{\mathcal{L}}}^{(m)} \tilde{\rho}^{(m)}_{\hat{\varepsilon}}(\tau) &= \hat{\boldsymbol{\mathcal{L}}}^{(m)} \tilde{\rho}^{(m)}_{\varepsilon}(\tau) + o_p(T^{-1/2}) \\
&= \tilde{\rho}^{(m)}_{\varepsilon}(\tau) - \boldsymbol{\xi}_{\varepsilon}(\tau) \kappa_{\tau+1} \left[\tilde{\boldsymbol{\rho}}^{(m)}_{\varepsilon} \right] + o_p(T^{-1/2}),
\end{aligned}$$

$\tau = 1, \ldots, m-k$, also under H_{1T}.

We have seem in Theorem 1 that, under Assumptions 1.2 - 1.4, the CLT for $\bar{\boldsymbol{\rho}}^{(m)}_{\varepsilon}$ follows from the CLT for $\tilde{\boldsymbol{\rho}}^{(m)}_{\varepsilon}$. Since under H_{1T} $\tilde{\boldsymbol{\rho}}^{(m)}_{\varepsilon}$ has asymptotic mean equal to $\tilde{\boldsymbol{h}}^{(m)}_{\varepsilon} = \left(h^{(m)}_{\varepsilon}(1), \ldots, h^{(m)}_{\varepsilon}(m) \right)'$, with $h^{(m)}_{\varepsilon}(\tau)$ as in (1.19), it is clear that, for $\tau = 1, \ldots, m-k$, $\hat{\boldsymbol{\mathcal{L}}}^{(m)} \tilde{\rho}^{(m)}_{\varepsilon}(\tau)$ is asymptotically normal, with an asymptotic drift equal to $\check{h}^{(m)}_{\varepsilon}(\tau)$, defined in (1.18), and asymptotic variance equal to $1 + \boldsymbol{\xi}_{\varepsilon}(\tau) \left(\sum_{l=\tau+1}^{m} \boldsymbol{\xi}_{\varepsilon}(l)' \boldsymbol{\xi}_{\varepsilon}(l) \right)^{-1} \boldsymbol{\xi}_{\varepsilon}(\tau)'$.

Since $\hat{\boldsymbol{\mathcal{L}}}^{(m)} \tilde{\rho}^{(m)}_{\varepsilon}(\tau)$ is asymptotic independent of $\hat{\boldsymbol{\mathcal{L}}}^{(m)} \tilde{\rho}^{(m)}_{\varepsilon}(q)$,for $1 \leq \tau < q \leq m-k$, as shown in Theorem 1, the result follows. ■

PROOF OF PROPOSITION 2

From the results of Theorem 2, we have that, under H_{1T} for $s = m-k < T$ fixed, $W^{(m)}(s) \xrightarrow{d} \sum_{\tau=1}^{m-k} w(\tau) \left(Z_{\tau} + \bar{h}^{(m)}_{\varepsilon}(\tau) \right)^2$ as $T \to \infty$. Next, from Theorem 3.2 of Billingsley (1999), it suffices to show that

$$\lim_{s\to\infty}\lim_{T\to\infty}\sup P\left(\left|W^{(m)}(s)-W^{(\infty)}(\infty)\right|>\epsilon\right)=0$$

for any $\epsilon>0$. This follows from the proof of Proposition 1 and Markov inequality, since for each fixed s,

$$E\left(T\sum_{\tau=s+1}^{s(T)}w\left(\tau\right)\bar{\rho}_{\hat{\varepsilon}}(\tau)^{2}\right)\leq C\sum_{\tau=s+1}^{s(T)}w\left(\tau\right)\leq C\sum_{\tau=s+1}^{\infty}w\left(\tau\right),$$

and the right-hand side converges to zero as $s\to\infty$. This concludes the proof of part (ii). For part (i), the results follows by noticing that under H_0, $\bar{h}_{\varepsilon}^{(m)}(\tau)=0\ \forall\tau\geq 1$.

Chapter 2

Nonparametric Tests for Conditional Treatment Effects with Duration Outcomes

2.1 Introduction

Assessing whether a policy has any effect on an outcome of interest has been one of the main concerns in empirical research. As summarized in Imbens (2004), Heckman and Vytlacil (2007), and Imbens and Wooldridge (2009), the focus of the policy evaluation literature has been mainly confined to situations where the realized outcome of interest is completely observed for the treated and the control groups. However, when the outcome variable is subjected to censoring, such inference procedures may provide misleading conclusions on the effect of the proposed policy. Assessing if labor market programs affect the length of unemployment, if correctional programs affect recidivism of criminal activities, or whether the survival time is affected by a new clinical therapy are just few examples where the outcome of interest is usually subjected to censoring mechanisms, and hence, standard policy evaluation procedures are not suitable. This article remedy this by proposing new nonparametric tests for conditional treatment effects when the outcome of interest, typically a duration, is subjected to right censoring.

Our test statistics are suitable functionals of empirical processes whose limiting distributions under the null can be estimated using a multiplicative-type bootstrap, which is proved to be valid. Our proposed tests are consistent against both one and two-sided alternative fixed alternatives and can detect nonparametric alternatives converging to the null at the parametric $n^{-1/2}$-rate, n being the sample size. Since our test proposal does not rely on continuity assumptions regarding the duration outcome, our policy evaluation tools are suitable for both discrete and continuous censored data. Moreover, our tests can be used not only for unconfounded treatment assignments, but also for the local treatment effect setup of Imbens and Angrist (1994) and Angrist et al. (1996), and for the case of dynamic treatment allocations as described in Sianesi (2004). Overall, this paper offers a unifying approach to derive uniformly valid nonparametric tests for treatment effects with censored outcomes. Although our focus is on hypotheses testing, estimators for unconditional treatment effects naturally arises as a by-product of the testing procedure.

To achieve the aforementioned properties, this paper relies on three components.

First, our tests are based on inverse probability weighting (IPW) estimators of the relevant treatment effect measures, in which the propensity score is estimated by nonparametric methods. In particular, we consider the series logit estimator proposed by Hirano et al. (2003), but other estimators are possible. Second, because the focus of this paper is on testing for conditional treatment effects, our hypotheses of interest are based on conditional moment restrictions. To avoid the use of smooth estimates, we adopt an integrated moment approach, reducing the conditional moment restrictions to an infinite number of unconditional orthogonality restrictions, as others have adopted in different contexts, see e.g. Delgado (1993), Stute (1997), Stute et al. (1998) and Delgado and González-Manteiga (2001). In a setup without censoring, we would be able to estimate the integrated moments by their empirical analogue. However, this is not feasible when the outcome of interest is subjected to right censoring. To handle this issue, we characterize the integrated moments as Kaplan-Meier (KM) integrals, see e.g. Stute and Wang (1993*a*,*b*), Stute (1993, 1995, 1996*a*), and Sellero et al. (2005). However, because the treatment effect measures depend on the propensity score, our integrand is unknown, which is in contrast to the literature on KM integrals. To accommodate this issue, we present new results for Kaplan-Meier integrals indexed by unknown, possibly infinite-dimensional nuisance parameters.

This paper is directly connected to the treatment effects literature. For recent reviews of this huge literature, see e.g. Imbens (2004), Heckman and Vytlacil (2007), and Imbens and Wooldridge (2009), among others. In cases where the outcome is subjected to censoring, few estimation procedures have been considered, see e.g. Ham and Lalonde (1996), Eberwein et al. (1997), Hubbard et al. (2000), Abbring and van den Berg (2003), Abbring and van den Berg (2005), Crépon et al. (2009), and Frandsen (2014), among others. Nonetheless, the aforementioned papers have not devoted attention to nonparametric tests. In fact, the literature on nonparametric tests for treatment effects is scarce, Abadie (2002), Crump et al. (2008), Lee and Whang (2009), Delgado and Escanciano (2013), and Hsu (2013) being exceptions when censoring is not an issue. In the presence of censoring, Lee (2009) developed a nonparametric test of the null hypothesis of no distributional treatment effect. However, the "two sample" setup adopted by Lee (2009) greatly differs from ours.

To illustrate the relevance of our new policy evaluation tools, we apply the proposed tests to evaluate labor market programs using two different sets of applications. First, as in Woodbury and Spiegelman (1987), we analyze the Illinois Reemployment Bonus Experiments that was carried out in the 1980's. Then, as in Lee (2009), we use observational female job training data from the Department of Labor in South Korea to test if receiving job training instead of unemployment insurance affects the unemployment duration. With these applications we show that introducing ad hoc parametric assumptions or ignoring treatment effect heterogeneity may lead to spurious conclusions about the policy effectiveness.

The remainder of the paper is organized as follows. We first describe the basic setup and the concentrate on testing the null of zero conditional distributional treatment effects. In Section 3, we derive the asymptotic distribution for the baseline tests and introduce a bootstrap method to approximate their critical values. A Monte Carlo study in Section 4 investigates the finite sample properties of the test proposals. In Section 5, we present some applications of our basic setup, i.e. we consider the null of zero conditional average treatment effects and show that our test procedure is also suitable when treatment allocation is endogenous or dynamic. In Section 6, we apply the policy evaluation tests to different datasets. Finally, we offer concluding remarks and suggest extensions for future research. Mathematical proofs are gathered in an appendix at the end of the article.

2.2 Testing for zero conditional treatment effects with censored outcomes

2.2.1 Basic setup

We consider a set of individuals flowing into a state of interest, and the time these individuals spend in that state is our outcome of interest. Upon inflow, an individual is assigned to a treatment or to a control group. The goal of this paper is to assess different hypotheses related to the causal effect of the treatment on the time spent in this state of interest. Henceforth, all random variables are defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Let D be an indicator of participation in the program, i.e. $D = 1$ if the unit participates in the treatment and $D = 0$ otherwise. Define Y_0 and Y_1 as the potential outcomes under the control and treatment groups, respectively. Additionally, let $X \in \mathbb{R}^k$ be vector of pre-treatment variables, and $\chi_{Y,X} \subseteq \mathbb{R} \times \mathbb{R}^k$ denote the support of $Y \times X$.

In this paper, the treatment effect measure of main interest is the conditional distributional treatment effect, that is, the difference between the conditional cumulative distribution function (CDF) of the potential outcome under treatment and control:

$$\Upsilon(t, x) = \mathbb{E}\left[1\{Y_1 \leq t\} - 1\{Y_0 \leq t\} | X = x\right].$$

Our main focus is on testing the hypothesis that the distributional treatment effect (DTE) is equal to zero for every subpopulation defined by covariates, that is,

$$H_0 : \Upsilon(t, x) = 0 \ \forall (x, t) \in \mathcal{W}, \tag{2.1}$$

where $\mathcal{W} \subseteq \chi_{Y,X}$. Under the null hypothesis H_0, the conditional distribution of Y is not affected by the treatment at $\mathcal{W}$, and the alternative hypothesis H_1 is the negation of H_0.

An important feature of the hypothesis in (2.1) is its focus on distributional treatment effects, and not only on the average treatment effects. By doing so, one can assess if the treatment has affected any feature of the distribution of the outcome, and not necessarily just the mean. In fact, by looking at the outcome distribution, one is able to perform welfare analysis under mild assumptions about social preferences, see e.g. Abadie (2002). Such analysis would not be possible if the focus were only at average treatment effects.

Another distinguishing characteristic of (2.1) is its focus on *conditional* treatment effects, and not only on the *unconditional* treatment effects. That is, in this paper we are concerned about the ubiquitous and commonly ignored feature that treatment effects may vary across different subpopulations. Although heterogeneity in the effect of a policy is generally allowed, *unconditional* measures of treatment effects may neglect some important differences in policy evaluations. For instance, a labor market program that does not affect the unemployment duration for the overall population might still be effective for a subgroup of individuals with specific observable characteristics. As illustrated by Bitler et al. (2006, 2008, 2014) and Crump et al. (2008), being able to assess if the treatment has affected any subpopulation is a crucial element of policy evaluations.

Next, we describe our setup. In order to model the treatment effect, we adopt the potential outcome notation popularized by Rubin (1974). Let D, Y_0, Y_1 and X be defined as before, and let $p(x) \equiv \mathbb{P}(D = 1|X = x)$ be the propensity score, i.e. the conditional probability of receiving treatment. Although our interest is on Y_0 and Y_1, one can only observe $Q \equiv DQ_1 + (1 - D)Q_0$, where $Q_0 = \min\{Y_0, C_0\}$, $Q_1 = \min\{Y_1, C_1\}$, C_0 and C_1 being potential censoring random variables under the control and treatment groups, respectively. Censoring might appear for different reasons such as the end of a follow-up or drop out. In addition to Q, one also observe the censoring indicator $\delta \equiv D\delta_1 + (1 - D)\,\delta_0$, where, for $j \in \{0, 1\}$, $\delta_j = 1\{Y_j \leq C_j\}$.

Assumption 2.1 $\{(Q_i, \delta_i, D_i, X_i)\}_{i=1}^{n}$ *are independent and identically distributed observations of* (Q, δ, D, X).

Assumption 2.2 $(Y_0, Y_1, C_0, C_1) \perp\!\!\!\perp D|X$ *a.s.*

Assumption 2.3 *For all* $x \in \mathcal{W}$ *and some* $\varepsilon > 0$, $\varepsilon \leq p(x) \leq 1 - \varepsilon$.

Assumption 2.4 *Assume that*

(i) $(Y_0, Y_1) \perp\!\!\!\perp (C_0, C_1)$

(ii) For $j \in \{0, 1\}$, $\mathbb{P}(\delta_j = 1|X, Y_j) = \mathbb{P}(\delta_j = 1|Y_j)$.

Assumption 2.5 *The distributions of* Y_j *and* C_j, $j \in \{0, 1\}$, *has no common jumps*

Assumptions 2.1-2.3 are standard in the treatment effects literature. Assumption 2.2 was introduced by Rosenbaum and Rubin (1983), and states that, conditional on observables, treatment assignment is independent of potential outcomes and censoring. Assumption 2.3 states that there is overlap in the covariate distributions. As shown by Khan and Tamer (2010), Assumption 2.3 is crucial in determining the convergence rate of inverse probability weighted estimators.

In the absence of censoring, Rosenbaum and Rubin (1983) show that Assumptions 2.2 and 2.3 would suffice to identify different treatment effects measures, in particular $\Upsilon(t,x)$. Nonetheless, it is important to notice that censoring introduces an additional identification challenge because the probability of censoring is related to potential outcomes, that is, censoring occurs only if $Y_j > C_j$, $j \in \{0,1\}$. Ignoring the censoring problem or analyzing only the uncensored outcomes would therefore introduce another source of confounding. To overcome such issue, Assumption 2.4 imposes additional structure on the censoring mechanism.

Assumption 2.4 states that, given the "time of death" Y_j, the covariates do not provide any further information whether censoring will take place, that is, δ_j and X are conditionally independent given the potential outcome Y_j. A particular case in which it holds is when C_j is independent of (Y_j, X), as assumed in Honore et al. (2002), Lee and Lee (2005) and Frandsen (2014), for example. Nonetheless, Assumption 2.4 is more general and allows censoring to depend on the covariates through the potential outcome Y_j. We notice that similar assumptions have been used in different contexts, see e.g. Chen (2001), Tang et al. (2003), D'Haultfoeuille (2010) and Breunig et al. (2014). An alternative to Assumption 2.4 is $(Y_0, Y_1) \perp\!\!\!\perp (C_0, C_1)\,|X$. In this case the use of smoothing techniques and trimming procedures are required, see Akritas (1994), González-Manteiga and Cadarso-Suárez (1994), and Iglesias Pérez and González-Manteiga (1999) for examples in different contexts. With Assumption 2.4, the use of smothers and trimming is avoided.

Assumption 2.5 is a regularity condition that does not exclude discontinuities of $F_{Y_j}(\cdot) \equiv \mathbb{P}(Y_j \leq \cdot)$ and $G_j(\cdot) \equiv \mathbb{P}(C_j \leq \cdot)$ at distinct points, that is, we do not impose that F_{Y_j} and G_j must be absolutely continuous. Therefore, we allow for both discrete and continuous potential outcomes.

With the aforementioned assumptions, the next proposition shows that we can point identify $\Upsilon(t,x)$ from the (Q,δ,D,X). For $j \in \{0,1\}$, let $\tau_{C_j} = \sup\{t : G_j(t) < 1\}$. For simplicity, assume that $\tau_{C_0} = \tau_{C_1} = \tau_C$.

Proposition 2.1 *Under Assumptions 2.2-2.4, for* $(t,x) \in (-\infty, \tau_C) \times \mathbb{R}^k$,

$$\Upsilon(t,x) = \mathbb{E}\left[\left(\frac{D\delta 1\{Q \le t\}}{(1-G_1(Q-))\,p(X)} - \frac{(1-D)\,\delta 1\{Q \le t\}}{(1-p(X))\,(1-G_0(Q-))}\right)\middle| X = x\right]. \tag{2.2}$$

Some remarks are necessary. From Proposition 2.1, one can see that nonparametric point identification of the distributional treatment effect over the entire outcome support may not be feasible. This is intuitive because, due to right censoring mechanisms, potential outcomes beyond τ_C are never observed. Given that one may not point identify the whole distributional treatment effect, the point identification of traditional measures such as the average treatment effect $\mathbb{E}[Y_1 - Y_0]$ is also at stake[1]. Nonetheless, (2.2) has considerable identification power. That is, by focusing on $\mathcal{W} \subseteq (-\infty, \tau_C) \times \mathbb{R}^k$, one can still point identify the distributional treatment effects measure of interest and test the hypothesis (2.1) within this portion of the CDF. This is feasible because τ_C is usually known in applications.

Another important feature of (2.2) is that the potentially restrictive condition that the censoring distribution is the same under both treatment regimes is not necessary for identification. Such result is in contrast with the one in Frandsen (2014), for example. Indeed, if one assumes that the censoring distribution is the same but this condition is not fulfilled, treatment effects measures may suffer from severe bias and tests based on this assumption may have large size distortions; see Section 2.4.

1. In Section 2.5, we show how one can identify a related measure, the trimmed ATE.

2.2.2 Characterization of the null hypothesis

Given that $\Upsilon(t,x)$ is identified from the data, we are able to characterize the null hypothesis (2.1) in terms of observables. In fact, based on the representation in (2.2) and using Assumption 2.3 guaranteeing that $p(\cdot) \in (0,1)$, we have,

$$\Delta(t,x) = 0 \ \forall (x,t) \in \mathcal{W} \Leftrightarrow \Upsilon(t,x) = 0 \ \forall (x,t) \in \mathcal{W}$$

where

$$\begin{aligned}\Delta(t,x) &= \mathbb{E}\left[\left(\frac{D(1-p(X))}{(1-G_1(Q-))} - \frac{(1-D)\,p(X)}{(1-G_0(Q-))}\right)\delta 1\{Q \leq t\} \,\middle|\, X = x\right] \\ &= \Upsilon(t,x)\, p(x)(1-p(x)).\end{aligned}$$

That is, in order to test the null hypothesis (2.1), it suffices to check if $\Delta = 0$. The main advantage of focusing on $\Delta(\cdot,\cdot)$ instead of $\Upsilon(\cdot,\cdot)$ is that random denominators due to the propensity score are avoided.

In order to assess if $\Delta(\cdot,\cdot) = 0$, there are two main approaches. The first one consists of using nonparametric smooth estimates of Δ. An important limitation of this local approach arises when X is multivariate due to the "curse of dimensionality". Moreover, tests in this category are not able to detect local alternatives converging to the null at the parametric rate $n^{-1/2}$. Instead, we adopt an integrated moment approach, avoiding the use of smoothers by means of reducing the conditional moment restriction to an infinite number of unconditional orthogonality restrictions, i.e., we characterize the null hypothesis (2.1) as

$$H_0 : I(t,x) = 0 \ \forall (t,x) \in \mathcal{W}, \tag{2.3}$$

where

$$I(t,x) = \mathbb{E}\left[\left(\frac{D(1-p(X))}{(1-G_1(Q-))} - \frac{(1-D)\,p(X)}{(1-G_0(Q-))}\right)\delta 1\{Q \leq t\}\, 1\{X \leq x\}\right]. \tag{2.4}$$

This integrated approach has been used in different contexts, see e.g. Delgado (1993), Stute (1997), Stute et al. (1998), Koul and Stute (1999) and Delgado and González-Manteiga (2001). Although other characterizations of H_0 are feasible (see Bierens and Ploberger (1997), Stinchcombe and White (1998) and Escanciano (2006*a*,*b*)), we do not pursue these possibilities in this paper.

2.2.3 Kaplan-Meier integrals and test statistics

The characterization of the null hypothesis in (2.3) suggests using functionals of an estimator of $I(\cdot,\cdot)$ as test statistics. Therefore, we must first estimate $I(\cdot,\cdot)$ using the sample $\{(Q_i, \delta_i, D_i, X_i)\}_{i=1}^n$. From (2.4), the challenge of estimating $I(\cdot,\cdot)$ is reduced to estimating $p(\cdot)$, $G_1(\cdot)$, $G_0(\cdot)$, and then applying the plug-in principle.

The task of nonparametrically estimate $p(\cdot)$ is relatively standard. Following Hirano et al. (2003), we can nonparametrically estimate $p(\cdot)$ using the Series Logit Estimator (SLE) based on power series. Although other nonparametric estimators could be used - see e.g. Ichimura and Linton (2005) and Li et al. (2009) - we do not exploit these possibilities in this paper.

To define the SLE, let $\lambda = (\lambda_1, \ldots, \lambda_r)'$ be a r-dimensional vector of non-negative integers with norm $|\lambda| = \sum_{j=1}^r \lambda_j$. Let $\{\lambda(l)\}_{l=1}^\infty$ be a sequence including all distinct multi-indices λ such that $|\lambda(l)|$ is non-decreasing in l and let $x^\lambda = \prod_{j=1}^r x_j^{\lambda_j}$. For any integer L, define $R^L(x) = \left(x^{\lambda(1)}, \ldots, x^{\lambda(L)}\right)'$ as a vector of power functions. Let $\mathcal{L}(a) = \exp(a)/(1+\exp(a))$ be the logistic CDF. The SLE for $p(x)$ is defined as $\hat{p}(x) = \mathcal{L}\left(R^L(x)'\hat{\pi}_L\right)$, where

$$\hat{\pi}_L = \arg\max_{\pi_L} \frac{1}{n}\sum_{i=1}^n D_i \log\left(\mathcal{L}\left(R^L(X_i)'\pi_L\right)\right) + (1-D_i)\log\left(1-\mathcal{L}\left(R^L(X_i)'\pi_L\right)\right).$$

Next, instead of directly considering estimators for $G_1(\cdot)$ and $G_0(\cdot)$, we show that, similarly to Stute (1993, 1996*a*), we can estimate $I(\cdot,\cdot)$ by means of empirical Kaplan-Meier integrals. To fix ideas, suppose we could fully observe (Y, X, D), implying that

$G_1(\cdot) = G_0(\cdot) = 0$ a.s.. For a given $(t,x) \in \mathcal{W}$, define

$$\xi_1(\bar{y},\bar{x},\bar{z};t,x) = \bar{z}(1-p(\bar{x}))1\{\bar{y}\leq t\}1\{\bar{x}\leq x\}, \tag{2.5}$$

$$\xi_0(\bar{y},\bar{x},\bar{z};t,x) = (1-\bar{z})p(\bar{x})1\{\bar{y}\leq t\}1\{\bar{x}\leq x\}, \tag{2.6}$$

and notice that, in the absence of censoring,

$$\begin{aligned} I(t,x) &= \mathbb{E}[\xi_1(Y,X,D;t,x)] - \mathbb{E}[\xi_0(Y,X,D;t,x)] \\ &= \int \xi_1(\bar{y},\bar{x},\bar{z};t,x)F_1(d\bar{y},d\bar{x}) - \int \xi_0(\bar{y},\bar{x},\bar{z};t,x)F_0(d\bar{y},d\bar{x}), \end{aligned}$$

where $F_j(t,x) \equiv \mathbb{P}(Y\leq t, X\leq x, D=j)$, $j\in\{0,1\}$.

From the above representation, and with the SLE $\hat{p}(\cdot)$ at our disposal, one could estimate $I(\cdot,\cdot)$ by its sample analogue

$$\begin{aligned} &\int \hat{\xi}_1(\bar{y},\bar{x},\bar{z};t,x)\hat{F}_1(d\bar{y},d\bar{x}) - \int \hat{\xi}_0(\bar{y},\bar{x},\bar{z};t,x)\hat{F}_0(d\bar{y},d\bar{x}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\xi}_1(Y_i,X_i,D_i;t,x) - \hat{\xi}_0(Y_i,X_i,D_i;t,x)\right] \end{aligned} \tag{2.7}$$

where $\hat{F}_j(t,x)$ denotes the empirical analog of $F_j(t,x)$, and

$$\hat{\xi}_1(\bar{y},\bar{x},\bar{z};t,x) = \bar{z}(1-\hat{p}(\bar{x}))1\{\bar{y}\leq t\}1\{\bar{x}\leq x\}, \tag{2.8}$$

$$\hat{\xi}_0(\bar{y},\bar{x},\bar{z};t,x) = (1-\bar{z})\hat{p}(\bar{x})1\{\bar{y}\leq t\}1\{\bar{x}\leq x\}, \tag{2.9}$$

the analogous of (2.5) and (2.6), but with the true $p(\cdot)$ replaced by the SLE $\hat{p}(\cdot)$. Unfortunately, due to the censoring problem, $\hat{F}_j(\cdot,\cdot)$ is not at our disposal and therefore, the above procedure is infeasible. Nonetheless, we can exploit other possibilities. Since the Kaplan and Meier (1958) estimator is the analogous to the empirical CDF when the outcome is subjected to right censoring, a convenient way to proceed involves using some multivariate Kaplan-Meier (KM) estimator of $F_j(\cdot,\cdot)$, which would use only the information available at the sample $\{(Q_i,\delta_i,D_i,X_i)\}_{i=1}^n$.

To define the KM estimator of $F_j(t,x)$, $j=0,1$, let n_1 and n_0 be the total number of individuals in the treated and control subsamples, $Q_{j,1:n_j} \leq \cdots \leq Q_{j,n_j:n_j}$ be the ordered Q values for the subsamples with $D = j \in \{0,1\}$, where ties within Y or within C are ordered arbitrarily and ties among Y and C are treated as if the former precedes the later, and let $\delta_{j,[i:n_j]}$ and $X_{j,[i:n_j]}$ be the concomitant of the *ith* order statistics of the subsample with $D = j$, i.e. the δ and X paired with $Q_{j,i:n_j}$. Similarly to Stute (1993, 1996*a*), the multivariate Kaplan-Meier estimator of $F_j(t,x)$ is given by

$$\hat{F}_j^{KM}(t,x) = \sum_{i=1}^{n_j} W_{j,i:n_j} 1\left\{Q_{j,i:n_j} \leq t\right\} 1\left\{X_{j,[i:n_j]} \leq x\right\},$$

where

$$W_{j,k:n_j} = \frac{n_j}{n} \frac{\delta_{j,[k:n_j]}}{n_j - k + 1} \prod_{l=1}^{k-1} \left(\frac{n_j - l}{n_j - l + 1}\right)^{\delta_{j,[l:n_j]}}$$

denotes its "jump" at observation k. It is important to notice that, because we do not impose that the censoring variables C_1 and C_0 follow the same distribution, the KM jump differ depending on whether D is equal to 0 or 1. This is the reason why we must consider different KM estimators for $F_0(\cdot,\cdot)$ and $F_1(\cdot,\cdot)$.

With the SLE $\hat{p}(\cdot)$ and the KM estimators $\hat{F}_1^{KM}(\cdot,\cdot)$ and $\hat{F}_0^{KM}(\cdot,\cdot)$ at hands, one can follow the same steps as in (2.7), and estimate $I(\cdot,\cdot)$ by

$$\begin{aligned}
\hat{I}(t,x) &= \int \hat{\xi}_1(\bar{y},\bar{x},\bar{z};t,x)\,\hat{F}_1^{KM}(d\bar{y},d\bar{x}) - \int \hat{\xi}_0(\bar{y},\bar{x},\bar{z};t,x)\,\hat{F}_0^{KM}(d\bar{y},d\bar{x}) \\
&= \Bigg[\sum_{i=1}^{n_1} W_{1,i:n_1}\hat{\xi}_1\left(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1,[i:n_1]}; t,x\right) \\
&\qquad\qquad - \sum_{l=1}^{n_0} W_{0,l:n_0}\hat{\xi}_0\left(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t,x\right)\Bigg] \qquad (2.10)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\xi}_1(Y_i,X_i,D_i;t,x)\,\delta_i}{1-\hat{G}_1^{KM}(Q_i-)} - \frac{\hat{\xi}_0(Y_i,X_i,D_i;t,x)\,\delta_i}{1-\hat{G}_0^{KM}(Q_i-)}\right),
\end{aligned}$$

where $\hat{G}_j^{KM}(\cdot)$ is the Kaplan and Meier (1958) estimator of $G_j(\cdot)$, $j = 0,1$, and the last equality follows from the results of Satten and Datta (2001).

From the above representation of $\hat{I}(\cdot,\cdot)$, one can clearly see that indeed the task of estimating $I(\cdot,\cdot)$ is reduced to estimate $p(\cdot)$, use KM estimators for $G_1(\cdot)$, $G_0(\cdot)$, and then applying plug-in principle. Moreover, in the absence of censoring, for $i = 1,\ldots,n$,

$Q_i = Y_i$, $\delta_i = 1$ and $W_{1,i:n_1} = W_{0,i:n0} = n^{-1}$ $a.s.$.Therefore, (2.10) naturally reduces to (2.7). Hence, one can clearly see that our procedure is suitable regardless of the presence of censoring.

With $\hat{I}(\cdot,\cdot)$ at hand, we are able to test the null hypothesis (2.1). Our test statistics are based on distances from $\sqrt{n}\hat{I}(\cdot,\cdot)$ to zero. We consider the usual *sup* and L_2 norms, leading to the Kolmogorov-Smirnov (KS), and Cramér-von Mises (CvM) test statistics

$$KS_n = \sqrt{n} \sup_{(t,x)\in\mathcal{W}} \left|\hat{I}(t,x)\right|, \tag{2.11}$$

$$CvM_n = n \int_{\mathcal{W}} \left|\hat{I}\left(t,x\right)\right|^2 \hat{H}\left(dt,dx\right), \tag{2.12}$$

respectively, where $\hat{H}\left(t,x\right)$ denotes the sample analog of $H\left(t,x\right) = \mathbb{P}\left(Q \leq t, X \leq x\right)$. Obviously, different test statistics could be developed by applying other distances, but for ease of exposition, we concentrate of KS_n and CvM_n.

Notice that, as a by-product of the our testing procedure, for $t \in (-\infty, \tau_C)$, one can estimate the unconditional distributional treatment effects (DTE)

$$\Upsilon\left(t\right) = \mathbb{E}\left[\frac{D\delta 1\left\{Q \leq t\right\}}{\left(1 - G_1\left(Q-\right)\right)p(X)} - \frac{\left(1-D\right)\delta 1\left\{Q \leq t\right\}}{\left(1-p(X)\right)\left(1-G_0\left(Q-\right)\right)}\right] \tag{2.13}$$

by

$$\hat{\Upsilon}\left(t\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i\delta_i 1\left\{Q_i < \tau\right\}}{1 - \hat{G}_1^{KM}\left(Q_i-\right)\hat{p}\left(X_i\right)} - \frac{\left(1-D_i\right)\delta_i 1\left\{Q_i < \tau\right\}}{\left(1 - \hat{G}_0^{KM}\left(Q_i-\right)\right)\left(1-\hat{p}\left(X_i\right)\right)}\right).$$

Hubbard et al. (2000) proposes a similar estimator, but relying on parametric methods, whereas Abbring and van den Berg (2005) consider a related estimator in a context without covariates. A detailed comparison between these estimators is beyond the scope of this paper. Furthermore, by using test statistics similar to (2.11) and (2.12), one can test for the presence of overall treatment effects. To avoid repetition of arguments, we focus on the conditional tests.

2.3 Asymptotic Theory

2.3.1 Asymptotic linear representation

We now discuss the asymptotic theory for our test statistics KS_n and CvM_n, using the following notation. For a generic set $\mathcal{G}$, let $l^{\infty}(\mathcal{G})$ be the Banach space of all uniformly bounded real functions on $\mathcal{G}$ equipped with the uniform metric $\|f\|_{\mathcal{G}} \equiv \sup_{z\in\mathcal{G}}|f(z)|$. We study the weak convergence of $\sqrt{n}\left(\hat{I}-I\right)(\cdot,\cdot)$ and related processes as elements of $l^{\infty}(\mathcal{W})$. Let $\Rightarrow$ denote weak convergence on $(l^{\infty}(\mathcal{W}),\mathcal{B}_{\infty})$ in the sense of J. Hoffmann-Jϕrgensen, where $\mathcal{B}_{\infty}$ denotes the corresponding Borel σ-algebra - see e.g. van der Vaart and Wellner (1996).

As shown in Section 2.2.3, $\hat{I}(\cdot,\cdot)$ is the difference of two empirical Kaplan-Meier integrals. However, because our KM integrals depend on a nonparametric estimate for the propensity score $p(\cdot)$, the results available in the literature cannot be straightforwardly applied, see e.g. Stute and Wang (1993*b*), Stute (1993, 1995, 1996*a*), and Sellero et al. (2005). To accommodate this issue, we must present new results for our Kaplan-Meier integrals indexed by unknown, infinite-dimensional nuisance parameters. In short, we show that, due the propensity score estimation effect, an additional term in the asymptotic representation of $\sqrt{n}\left(\hat{I}-I\right)(t,x)$ must be considered.

In order to proceed with the asymptotic analysis, let us introduce some additional notation. For $j\in\{0,1\}$, let $F_j(t|\bar{x})=\mathbb{E}[1\{Y_j\le t\}|X=\bar{x}]$, $H_j(t)=\mathbb{P}(Q\le t,D=j)$, $H_{j,0}(y)=\mathbb{P}(Q\le t,\delta=0,D=j)$, and $H_{j,11}(t,x)=\mathbb{P}(Q\le t,X\le x,D=j,\delta=1)$. Note that H_j, $H_{j,0}$ and $H_{j,11}$ may be consistently estimated from the observed data.

For $j\in\{0,1\}$ define

$$\gamma_{j,0}(\bar{t})=\exp\left\{\int_0^{\bar{t}-}\frac{H_{j,0}(d\bar{w})}{1-H_j(\bar{w})}\right\}.$$

Let

$$\gamma_{j,1}(\bar{t})=\frac{1}{1-H_j(t)}\int 1\{\bar{t}<\bar{w}\}\,\xi_j(\bar{w},\bar{x},\bar{z};t,x)\,\gamma_{j,0}(\bar{w})\,H_{j,11}(d\bar{w},d\bar{x})$$

and

$$\gamma_{j,2}\left(\bar{t}\right)=\int\int\frac{1\left\{\bar{v}<\bar{t},\bar{v}<\bar{w}\right\}\xi_{j}\left(\bar{w},\bar{x},\bar{z};t,x\right)}{\left[1-H_{j}\left(\bar{v}\right)\right]^{2}}\gamma_{j,0}\left(\bar{w}\right)H_{j,0}\left(d\bar{v}\right)H_{j,11}\left(d\bar{w},d\bar{x}\right),$$

here $\xi_1\left(\cdot,\cdot,\cdot;t,x\right)$ and $\xi_0\left(\cdot,\cdot,\cdot;t,x\right)$ are as defined in (2.5) and (2.6), respectively. Put

$$\eta_{j,i}\left(t,x\right)=\xi_{j}\left(Q_{i},X_{i},D_{i};t,x\right)\gamma_{j,0}\left(Q_{i}\right)\delta_{i}+\gamma_{j,1}\left(Q_{i}\right)\left(1-\delta_{i}\right)-\gamma_{j,2}\left(Q_{i}\right). \tag{2.14}$$

Some remarks are necessary. First, the above representation relies only on the "known" functions ξ_j, $j=0,1$. Then, as discussed in Stute (1995, 1996a), the first term of $\eta_{j,i}\left(t,x\right)$ has expectation $\mathbb{E}\left[\xi_j\left(Q,X,D;t,x\right)\right]$. The second and third terms represent the estimation effect coming from not knowing $G_j\left(\cdot\right)$ in (2.10), and they have identical expectations. Finally, notice that in the absence of censoring, $\gamma_{j,0}\left(\cdot\right)=1$ $a.s.$, and $\gamma_{j,1}\left(\cdot\right)=\gamma_{j,2}\left(\cdot\right)=0$ $a.s.$.

Given that $\hat{I}(\cdot,\cdot)$ is the difference of empirical KM integrals, define

$$\eta_{i}\left(t,x\right)=\eta_{1,i}\left(t,x\right)-\eta_{0,i}\left(t,x\right), \tag{2.15}$$

the difference of (2.14) between the treated and control group.

To discuss the estimation effect coming from not knowing $p\left(\cdot\right)$ in the KM-integrals, let

$$\alpha_{1}\left(X;t,x\right)=-p\left(X\right)1\left\{X\leq x\right\}F_{1}\left(t|X\right), \tag{2.16}$$

$$\alpha_{0}\left(X;t,x\right)=\left(1-p\left(X\right)\right)1\left\{X\leq x\right\}F_{0}\left(t|X\right). \tag{2.17}$$

Notice that $\alpha_1\left(\cdot;t,x\right)$ and $\alpha_0\left(\cdot;t,x\right)$ are nothing more than the conditional expectation of the derivative of ξ_1 and ξ_0, as defined in (2.5) and (2.6), with respect to $p(\cdot)$, respectively. Similarly to (2.15), define

$$\alpha\left(X;t,x\right)=\alpha_{1}\left(X;t,x\right)-\alpha_{0}\left(X;t,x\right). \tag{2.18}$$

Before presenting our asymptotic results, we need to assume some additional regularity conditions.

Assumption 2.6 *(i) The support* χ_X *of the* k*-dimensional covariate* X *is a Cartesian product of compact intervals,* $\chi_X = \prod_{j=1}^{k} [x_{lj}, x_{uj}]$ *;*

(ii) The density of X *is bounded, and bounded away from 0, on* χ_X

(iii) For $j \in \{0,1\}$*, and any given* $t \in \chi_Y$*,* $F_j(t|x)$ *is continuously differentiable in* $x \in \chi_X$*.*

Assumption 2.7 *For all* $x \in \chi_X$*, the propensity score* $p(x)$ *is continuously differentiable of order* $s \geq 13k$*, where* k *is the dimension of* X*.*

Assumption 2.8 *The series logit estimator of* $p(x)$ *uses a power series with* $L = a \cdot N^v$ *for some* $a > 0$ *and* $1/(s/k - 2) < v < 1/11$*.*

Similar assumptions have been done by Hahn (1998), Hirano et al. (2003), Crump et al. (2008), Donald and Hsu (2013), among others. Assumption 2.6 restrict the distribution of X and Y and requires that all covariates are continuous. By imposing these restrictions, we are able to use Newey (1997) results for series estimators. Nonetheless, at the expense of additional notation, we can deal with the case where X has both continuous and discrete components by means of sample splitting based on the discrete covariates. In order to avoid cumbersome notation, we abstract from this point in the rest of the paper. Assumption 2.7 requires sufficient smoothness of the propensity score, whereas Assumption 2.8 restrict the rate at which additional terms are added to the series approximation of $p(x)$, depending on the dimension of X and the number of derivatives of $p(x)$. The restriction on the derivatives in Assumption 2.7 guarantees the existence of a v that satisfy the conditions in Assumption 2.8.

Under the aforementioned conditions, we can state our first asymptotic result, which provides the representation of $\sqrt{n}\left(\hat{I} - I\right)(t, x)$ over $\mathcal{W}$.

Lemma 2.1 *Under Assumptions 2.1-2.8, we have*

$$\sup_{(t,x)\in\mathcal{W}} \left| \sqrt{n}\left(\hat{I}(t,x) - I(t,x)\right) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\left[\eta_i(t,x) - I(t,x)\right] + \alpha(X_i;t,x)(D_i - p(X_i))\right\} \right| = o(1).$$

Lemma 2.1 shows that the estimator $\hat{I}(t,x)$ can be represented as asymptotically linear:

$$\hat{I}(t,x) = I(t,x) + \frac{1}{n}\sum_{i=1}^{n}\left\{\psi_i(t,x) + \tilde{\alpha}_i(t,x)\right\} + o_{\mathbb{P}}\left(n^{-1/2}\right)$$

where

$$\psi_i(t,x) = \eta_i(t,x) - I(t,x),$$

$\eta_i(t,x)$ being defined as in (2.15) and

$$\tilde{\alpha}_i(t,x) = \alpha(X_i;t,x)(D_i - p(X_i)). \tag{2.19}$$

The known-IPW estimator, (2.10) with $\hat{p}(x)$ replaced by $p(x)$, is asymptotically linear with score function $\psi(\cdot,\cdot)$. The function $\tilde{\alpha}(t,x)$ represents the effect on the score function of estimating $p(\cdot)$.

2.3.2 Asymptotic null distribution

Using the uniform representation of Lemma 2.1, we next derive the weak convergence of the processes $\sqrt{n}\hat{I}(t,x)$ under the null hypothesis (2.1).

Theorem 2.1 *Under the null hypothesis (2.1) and Assumptions 2.1-2.8, we have*

$$\sqrt{n}\hat{I}(t,x) \Rightarrow C_\infty,$$

where C_∞ is Gaussian process with zero mean and covariance function

$$V((t_1,x_1),(t_2,x_2)) = \mathbb{E}\left[\left\{\psi(t_1,x_1) + \tilde{\alpha}(t_1,x_1)\right\}\left\{\psi(t_2,x_2) + \tilde{\alpha}(t_2,x_2)\right\}\right]. \tag{2.20}$$

Now, we can apply the continuous mapping theorem in order to characterize the limiting null distributions of our test statistics using the sup and L_2 distances.

Corollary 2.1 *Under the null hypothesis (2.1) and the assumptions of Theorem 2.1,*

$$KS_n \xrightarrow{d} \sup_{(t,x)\in\mathcal{W}} |C_\infty(t,x)|,$$
$$CvM_n \xrightarrow{d} \int_{\mathcal{W}} |C_\infty(t,x)|^2 H(dt,dx).$$

Let T_n be a generic notation for KS_n and CvM_n. From Corollary 2.1, it follows immediately that

$$\lim_{n\to\infty} \mathbb{P}\left\{T_n > c_\alpha^T\right\} = \alpha$$

where

$$c_\alpha^T = \inf\left\{c \in [0,\infty) : \lim_{n\to\infty} \mathbb{P}\{T_n > c\} = \alpha\right\},$$

2.3.3 Asymptotic distribution under fixed and local alternatives

Now, we analyze the asymptotic properties of our tests under the fixed alternative H_1. Under H_1, there is at least one $(t,x) \in \mathcal{W}$ such that $\Upsilon(t,x) \neq 0$, implying that $I(t,x) \neq 0$ for some $(t,x) \in \mathcal{W}$. Therefore, our test statistics KS_n and CvM_n diverge to infinity. Given that the critical values are bounded, it follows that our tests are consistent. We formalize this result in the next theorem.

Theorem 2.2 *Under Assumptions 2.1-2.8 and the alternative hypothesis H_1, we have*

$$\lim_{n\to\infty} \mathbb{P}\left\{KS_n > c_\alpha^{KS}\right\} = 1,$$
$$\lim_{n\to\infty} \mathbb{P}\left\{CvM_n > c_\alpha^{CvM}\right\} = 1.$$

Given that our test statistics diverge to infinity under fixed alternatives, it is desirable studying the asymptotic power of these tests under local alternatives. To this end, we study the asymptotic behavior of $\hat{I}(t,x)$ under alternative hypotheses converging to the null at the parametric rate $n^{-1/2}$.

Consider the following class of local alternatives:

$$H_{1,n} : \Upsilon (t, x) = \frac{1}{\sqrt{n}} h (t, x) \quad \forall (t, x) \in \mathcal{W}, \tag{2.21}$$

In the sequel, we need that (2.21) satisfies the following regularity condition.

Assumption 2.9 *(a) $h(\cdot, \cdot)$ is an F-integrable function;*

(b) the set $h_n \equiv [(t, x) \in \mathcal{W} : h(x, t) \neq 0]$ has positive Lebesgue measure.

Theorem 2.3 *Under the local alternatives (2.21) and Assumptions 2.1- 2.9,*

$$\sqrt{n} \hat{I} (t, x) \Rightarrow C_\infty + R$$

where C_∞ is the process defined in Theorem 2.1 and $R(\cdot)$ is the deterministic function

$$R (t, x) = \mathbb{E} [h (t, X) (p(X) (1 - p (X))) 1 \{X \leq x\}] .$$

The following corollary is a consequence of the continuous mapping theorem and Theorem 2.3.

Corollary 2.2 *Under the local alternatives (2.21),and Assumptions 2.1- 2.9,*

$$\begin{aligned} KS_n &\xrightarrow{d} \sup_{(t,x) \in \mathcal{W}} |C_\infty (t, x) + R (t, x)| , \\ CvM_n &\xrightarrow{d} \int_{\mathcal{W}} |C_\infty (t, x) + R (t, x)|^2 H (dt, dx) , \end{aligned}$$

From the above corollary, we see that our test statistics, under local alternatives of the form of (2.21), converge to a different distribution due to the presence of a deterministic shift function R. This additional term guarantees the good local power property of our test.

2.3.4 Estimation of critical values

From the above theorems, we see that the asymptotic distribution of $\sqrt{n}\hat{I}(\cdot,\cdot)$ depends on the underlying data generating process and standardization is complicated in this case. Therefore, we propose a bootstrap method to estimate the critical values of our test. Our bootstrap procedure is related to the wild bootstrap, but instead of just resampling imposing the restriction under H_0, we use the asymptotic linear representation of $\sqrt{n}\hat{I}(\cdot,\cdot)$. More precisely, we consider the multiplier-type bootstrap as Stute et al. (2000), Delgado and González-Manteiga (2001), Barrett and Donald (2003) and Donald and Hsu (2013) suggest in different contexts. The proposed procedure has good theoretical and empirical properties, is straightforward to verify its asymptotic validity, and is computationally easy to implement.

In order to implement the bootstrap, we need nonparametric estimators for the terms in the asymptotic linear representation of Lemma 2.1, namely the propensity score $p(\cdot)$, $\eta(t,x)$ as defined in (2.15), and $\alpha(\cdot;t,x)$ as in (2.18).

As already discussed, we estimate $p(\cdot)$ using the SLE of Hirano et al. (2003). In order to estimate $\eta(t,x)$, we notice that after plugging in $\hat{p}(\cdot)$, each γ only depends on $H-functions$ and is therefore estimable just replacing the $H-terms$ by their empirical counterparts Then, we estimate $\eta(t,x)$ by its empirical analogue,

$$\hat{\eta}(t,x) = \hat{\eta}_1(t,x) - \hat{\eta}_0(t,x)$$

such that, for $j=0,1$,

$$\begin{aligned}
\hat{\eta}_j(t,x) &= \hat{\xi}_j(Q,X,D;t,x)\,\hat{\gamma}_{j,0}(Q)\,\delta_{j,i} + \hat{\gamma}_{j,1}(Q)(1-\delta) - \hat{\gamma}_{j,2}(Q),\\
\hat{\gamma}_{j,0}(\bar{t}) &= \exp\left\{\int_0^{\bar{t}-} \frac{\hat{H}_{j,0}(d\bar{w})}{1-\hat{H}_j(\bar{w})}\right\},\\
\hat{\gamma}_{j,1}(\bar{t}) &= \frac{1}{1-\hat{H}_j(t)}\int 1\{\bar{t}<\bar{w}\}\,\hat{\xi}_j(\bar{w},\bar{x},\bar{z};t,x)\,\hat{\gamma}_{j,0}(\bar{w})\,\hat{H}_{j,11}(d\bar{w},d\bar{x}),\\
\hat{\gamma}_{j,2}(\bar{t}) &= \int\int \frac{1\{\bar{v}<\bar{t},\bar{v}<\bar{w}\}\,\hat{\xi}_j(\bar{w},\bar{x},\bar{z};t,x)}{\left[1-\hat{H}_j(\bar{v})\right]^2}\hat{\gamma}_{j,0}(\bar{w})\,\hat{H}_{j,0}(d\bar{v})\,\hat{H}_{j,11}(d\bar{w},d\bar{x}),
\end{aligned}$$

where$\hat{\xi}_1(\cdot,\cdot,\cdot;t,x)$ and $\hat{\xi}_0(\cdot,\cdot,\cdot;t,x)$ are defined in (2.8) and (2.9), respectively, and

$$\hat{H}_j(\bar{w}) = \frac{1}{n}\sum_{i=1}^{n} 1\{Q_i \leq \bar{w}\} 1\{D_i = j\},$$
$$\hat{H}_{j,0}(\bar{w}) = \frac{1}{n}\sum_{i=1}^{n} (1-\delta_i) 1\{Q_i \leq \bar{w}\} 1\{D_i = j\}$$
$$\hat{H}_{j,11}(\bar{w},\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} \delta_i 1\{Q_i \leq \bar{w}\} 1\{X_i \leq \bar{x}\} 1\{D_i = j\}$$

are the empirical counterparts of $H_j(\bar{w})$, $H_{j,0}(\bar{w})$ and $H_{j,11}(\bar{w})$, respectively.

Finally, we must consider nonparametric estimate for $\alpha(X;t,x) = \alpha_1(X;t,x) - \alpha_0(X;t,x)$, $\alpha_1(X;t,x)$ and $\alpha_1(X;t,x)$ being defined in (2.16) and (2.17), respectively. To this end, notice that

$$\begin{aligned}\alpha(X;t,x) &= -\mathbb{E}[D1\{Y \leq t\} + (1-D)1\{Y \leq t\}|X] 1\{X \leq x\} \\ &= -\mathbb{E}[1\{Y \leq t\}|X] 1\{X \leq x\}.\end{aligned}$$

If we fully observe (Y,X,D), we could estimate this conditional expectation using nonparametric series regression of $1\{Y \leq \cdot\}$ on X, as similarly adopted by Hirano et al. (2003) and Donald and Hsu (2013). Given that the outcome of interest Y is subjected to censoring, such procedure is not at our disposal. To the best of our knowledge, no nonparametric estimator for $\alpha(\cdot;t,x)$ is yet available.

Notwithstanding, by using the Kaplan-Meier weights as discussed in Sections 2.2 and 2.3.1, we can overcome such problem and estimate $\alpha(\cdot;t,x)$ by the Kaplan-Meier series estimator

$$\begin{aligned}&\hat{\alpha}^{KM}(X;t,x) \\ &= -\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i}{1-\hat{G}_1^{KM}(Q_i-)} + \frac{1-D_i}{1-\hat{G}_0^{KM}(Q_i-)}\right)\delta 1\{Q_i \leq t\} R^L(X_i)\right)' \\ &\qquad \times \left(\frac{1}{n}\sum_{i=1}^{n} R^L(X_i) R^L(X_i)'\right)^{-1} R^L(X) 1\{X \leq x\},\end{aligned}$$

where $R^L(\cdot)$ is the same power series used in SLE estimator, with potentially different number of series. The uniform consistency of the aforementioned nonparametric estimator for $\alpha(X;t,x)$ is proved in Lemma 2.5 in Appendix A.

Once we have nonparametric estimators $p(\cdot)$, $\eta(t,x)$, and $\alpha(\cdot;t,x)$, the bootstrapped version of $\hat{I}(t,x)$ is given by

$$\hat{I}^{*}(t,x)=\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\eta}_i(t,x)+\hat{\alpha}^{KM}(X_i;t,x)(D_i-\hat{p}(X_i))\right]V_i$$

where the random variables $\{V_i\}_{i=1}^{n}$ are *iid* as a random variable V with bounded support, zero mean and variance one, being independent generated from the sample $\{(Q_i,\delta_i,D_i,X_i)\}_{i=1}^{N}$.

Replacing $\hat{I}(t,x)$ with $\hat{I}^{*}(t,x)$, we get the bootstrap versions of KS_n and CvM_n, KS_n^{*} and CvM_n^{*}, respectively. The asymptotic critical values are estimated by

$$\begin{aligned}c_{n,\alpha}^{KS,\ *}&\equiv\inf\left\{c_\alpha\in[0,\infty):\lim_{n\to\infty}\mathbb{P}_n^{*}\{KS_n^{*}>c_\alpha\}=\alpha\right\},\\ c_{n,\alpha}^{CvM,\ *}&\equiv\inf\left\{c_\alpha\in[0,\infty):\lim_{n\to\infty}\mathbb{P}_n^{*}\{CvM_n^{*}>c_\alpha\}=\alpha\right\}\end{aligned}$$

where $\mathbb{P}_n^{*}$ means bootstrap probability, i.e. conditional on the sample $\{(Q_i,\delta_i,D_i,X_i)\}_{i=1}^{n}$. In practice, $c_{n,\alpha}^{KS,\ *}$ and $c_{n,\alpha}^{CvM,\ *}$ are approximated as accurately as desired by $\left(KS_n^{*}\right)_{B(1-\alpha)}$ and $\left(CvM_n^{*}\right)_{B(1-\alpha)}$, the $B(1-\alpha)-th$ order statistic from B replicates $\left\{KS_n^{*}\right\}_{l=1}^{B}$ of KS_n^{*} or $\left\{CvM_n^{*}\right\}_{l=1}^{B}$ of CvM_n^{*}, respectively.

The next theorem proves the validity of the proposed multiplicative bootstrap. Notice that we need an additional smoothness assumption on $F_j(\cdot|X)$, $j\in\{0,1\}$.

Theorem 2.4 *Let Assumptions 2.1-2.9 hold. Additionally, for $j\in\{0,1\}$, assume that $F_j(\cdot|X)$ is continuously differentiable of order $m\geq k$, where k is the dimension of X. Assume $\{V_i\}_{i=1}^{n}$ are iid, independent of the sample $\{(Q_i,\delta_i,D_i,X_i)\}_{i=1}^{N}$, bounded with zero mean and variance one. Then, under the null hypothesis (2.1), any fixed alternative hypothesis or under the local alternatives (2.21)*

$$\sqrt{n}\hat{I}^{*}\underset{*}{\Rightarrow}C_\infty$$

where C_∞ is the same Gaussian process of Theorem 2.1 and $\underset{*}{\Rightarrow}$ denoting weak convergence *a.s. under the the bootstrap law (see Giné and Zinn (1990)).*

Straightforward application of the continuous mapping theorem lead us to conclude that our bootstrap-based tests has correct asymptotic size, are consistent against fixed alternatives and are able to detect contiguous alternatives of the form of (2.21).

2.4 Monte Carlo simulations

In this section, we conduct a small scale Monte Carlo exercise in order to study the finite sample properties of our test statistics for the null hypothesis (2.1). The $\{V_i\}_{i=}^{n}$ used in the bootstrap implementations are independently generated as V with $\mathbb{P}(V = 1 - \kappa) = \kappa/\sqrt{5}$ and $\mathbb{P}(V = \kappa) = 1 - \kappa/\sqrt{5}$, where $\kappa = \left(\sqrt{5}+1\right)/2$, as proposed by Mammen (1993). The bootstrap critical values are approximated by Monte Carlo using 1000 replications and the simulations are based on 10000 Monte Carlo experiments. We report rejection probabilities at the 5% significance level. Results for 10% and 1% significance levels are similar and available upon request.

We consider the following four designs:

$$(i).\;\; Y_0 \sim Exponential\left(\frac{1}{1.1+X}\right),\; Y_1 \sim Exponential\left(\frac{1}{1.1+X}\right),$$
$$C_1 = C_2 \sim a_{11} + b_{11} \times Exponential\,(1)\,;$$

$$(ii).\;\; Y_0 \sim Exponential\left(\frac{1}{1.1+X}\right),\; Y_1 \sim Exponential\left(\frac{1}{1.1+X}\right),$$
$$C_1 \sim a_{12} + b_{12} \times Exponential 1,;\;\; C_2 \sim a_{22} + b_{22} \times Exponential\,(1)\,;$$

$$(iii).\;\; Y_0 \sim Exponential\left(\frac{1}{1.1+X}\right),\; Y_1 \sim Exponential\left(\frac{1}{0.1+X}\right),$$
$$C_1 \sim a_{13} + b_{13} \times Exponential\,(1)\,,\;\; C_2 \sim a_{23} + b_{23} \times Exponential\,(1)\,;$$

$$(iv).\;\; Y_0 \sim Exponential\left(\frac{1}{1.1+2X}\right),\; Y_1 \sim Exponential\left(\frac{1}{0.1+x}\right),$$
$$C_1 \sim a_{14} + b_{14} \times Exponential\,(1)\,,\;\; C_2 \sim a_{24} + b_{24} \times Exponential\,(1)\,;$$

where X is distributed as $U\,[0,1]$, independently of Y_0, Y_1, C_1 and C_0, and the parameters a and b are chosen such that the percentage of censoring is equal to 0, 10 or 30 percent in the whole sample. Design (i) and (ii) fall under the null hypothesis, and designs $(iii)-(iv)$ fall under the alternative. Design (i) differs from design (ii) by the censoring distribution: in (i), the censoring variable is the same for treated and control group, whereas in design (ii) C_1 and C_2 follow different distributions. In design (ii) we set that the censoring level under treated and control groups are different: it is 0, 5, and 20 under control and 0, 15 and 40 under treatment. For the other designs, the censoring proportion is equal for the treatment and control groups. In design (iii), the CDTE does not depend on covariates, whereas in design (iv) it does. In all designs, $\mathbb{P}\,(D=1|X) = X$.

We report the proportion of rejections for sample sizes $n = 100$, 300 and 1000. We estimate $p\,(\cdot)$ using the SLE: with $n = 100$ we use $1, X, X^2$, with $n = 300$ we use $1, X, X^2, X^3$ and with $n = 1000$ we use $1, X, X^2, X^3, X^4, X^5$ as power functions in the estimation procedure.

We compare our proposed tests KS_n and CvM_n as in (2.11)-(2.12), with two others alternatives: the 'naive' procedure where censoring is ignored (KS_n^{naive} and CvM_n^{naive}), and the analogous procedure of KS_n and CvM_n but imposing that the censoring variable is the same under treatment and control groups (KS_n^{same} and CvM_n^{same}), both implemented with the assistance of a bootstrap analogous to the one discussed in Section 2.3.4. The proportion of rejections are presented in Table 2.1.

Table 2.1: Empirical Rejection probabilities.

		n=100			n=300			n=1000		
		% of Censoring			% of Censoring			% of Censoring		
Design	Tests	0%	10%	30%	0%	10%	30%	0%	10%	30%
(i)	KS_n	-	5.02	4.78	-	5.13	4.87	-	5.22	5.00
	CvM_n	-	5.47	5.22	-	5.39	5.01	-	5.22	5.31
	KS_n^{naive}	5.02	5.34	3.34	4.96	4.98	3.09	5.55	6.19	5.53
	CvM_n^{naive}	4.96	5.34	4.17	5.17	4.88	3.74	5.47	6.15	5.29
	KS_n^{same}	-	4.85	4.32	-	5.13	5.28	-	5.14	5.38
	CvM_n^{same}	-	5.32	4.89	-	5.44	5.09	-	5.26	5.22
(ii)	KS_n	-	5.01	5.14	-	5.08	5.17	-	5.33	5.02
	CvM_n	-	5.39	5.44	-	5.28	4.95	-	5.30	5.12
	KS_n^{naive}	4.98	5.65	8.79	5.53	6.19	24.85	5.23	12.40	87.00
	CvM_n^{naive}	4.88	5.74	6.33	5.29	6.15	12.06	5.19	8.72	43.33
	KS_n^{same}	-	6.39	19.78	-	11.21	60.18	-	43.79	98.84
	CvM_n^{same}	-	5.81	12.26	-	6.82	33.32	-	12.03	89.46
(iii)	KS_n	-	83.97	60.99	-	99.83	92.35	-	100	99.27
	CvM_n	-	93.00	87.23	-	100	99.98	-	100	100
	KS_n^{naive}	88.14	84.91	68.28	99.98	99.97	99.32	100	100	100
	CvM_n^{naive}	93.98	90.47	78.34	100	100	99.80	100	100	100
	KS_n^{same}	-	78.31	38.22	-	99.52	86.27	-	100	98.69
	CvM_n^{same}	-	90.17	77.64	-	100	99.90	-	100	100
(iv)	KS_n	-	93.49	73.51	-	99.92	94.75	-	100	99.64
	CvM_n	-	98.34	95.13	-	100	100	-	100	100
	KS_n^{naive}	96.92	94.88	82.97	100	100	99.94	100	100	100
	CvM_n^{naive}	98.74	97.22	89.36	100	100	99.99	100	100	100
	KS_n^{same}	-	77.64	49.75	-	99.81	90.71	-	100	100
	CvM_n^{same}	-	97.25	89.60	-	100	100	-	100	100

Note: One thousand bootstrap replications. Ten thousand Monte Carlo simulations. 5% level.

We observe that our tests KS_n and CvM_n exhibits good size accuracy for both designs (i) and (ii) even when $n = 100$. In design (i), tests based on the 'naive' and the 'common censoring' control size, though once we increase the sample size and the censoring proportion, the size of KS_n^{naive} and CvM_n^{naive} fall bellow the nominal level. Although one may find the result that the 'naive' procedure is able to control the Type-I error surprising, the reason behind this is simple: since $Y_1 = Y_0$, and at the same time $C_1 = C_0$, the censored outcomes $Q_1 = \min(Y_1, C_1)$ and $Q_0 = \min(Y_0, C_0)$ are also equal. Nonetheless, when the C_1 is different than C_2, as in design (ii), this is not true anymore. As one can see from Table 2.1, the tests procedures that either ignore the censoring or incorrectly impose the assumption of common $G's$ are not able to control size in this situation. This size distortions become more evident as we increase the sample size and the censoring level, reaching values higher than 80%.

With respect to power, our tests KS_n and CvM_n reach moderate levels for $n = 100$, but they uniformly increase and reach satisfactory levels when sample size is 300. The power is decreasing with the degree of censoring. For the considered designs, CvM_n tends to have higher power than KS_n. In addition, we can see that our proposed tests has similar and some times even higher power to those based on the 'naive' and the 'common censoring' procedures. Overall, these simulations show that the proposed bootstrap tests KS_n and CvM_n exhibit very good size accuracy and power for relatively small sample sizes. On other hand, the tests KS_n^{naive}, CvM_n^{naive}, KS_n^{same} and CvM_n^{same} may not be reliable due to their inability of controlling size when the censoring distributions differ in the two treatment regimes.

2.5 Some applications of the basic setup

2.5.1 Average treatment effects

So far, we have only discussed tests for the existence of distributional treatment effects. Although the proposed tests for zero distributional treatment effects are able to detect a very broad set of alternative hypotheses, we are still not able to pin down the direction of the departure from the null hypothesis of interest. For instance, if we reject the null of zero distributional treatment effect for all subpopulations defined by covariates, we unfortunately do not know if the policy has affect the conditional mean or, instead, any other particular feature of the outcome distribution. Given that the policy evaluation literature has given a great deal of importance to the average treatment effect, in this section we show how one adapt our tests to focus on this particular measure.

Let $\Upsilon^{CATE}(x) = \mathbb{E}[Y_1 - Y_0 | X = x]$. Remember that, as discussed in Section 2.2, we may be unable to test hypotheses concerning $\Upsilon^{CATE}(x)$ itself because of lack of information in the right tail of the outcome distribution due to the censoring mechanism. Therefore, we focus our attention to the trimmed versions of $\Upsilon^{CATE}(x)$, $\Upsilon_{\tau}^{CATE}(x) = \mathbb{E}[Y_1 1\{Y_1 < \tau\} - Y_0 1\{Y_0 < \tau\} | X = x]$, where $\tau \leq \tau_C$. Similar procedures have been previously considered by Sellero et al. (2005) and Pardo-Fernandez and Van Keilegom (2006).

We are concerned with the following hypothesis:

$$H_0^{CATE} : \Upsilon_{\tau}^{CATE}(x) = 0 \ \forall x \in \mathcal{W}_X \tag{2.22}$$

where $\mathcal{W}_X \subseteq \chi_X$, χ_X denoting the support of X. Under H_0^{CATE}, the trimmed average treatment effect (ATE) is equal to zero for all subpopulations defined by covariates. The alternative hypothesis is the negation of the null.

Following the same steps as in Section 2.2, our Kolmogorov-Smirnov (KS) type test statistic for hypothesis (2.22) is

$$KS_n^{CATE} = \sup_{x \in \mathcal{W}_X} \left| \sqrt{n} \hat{I}_{\tau}^{CATE}(x) \right|, \tag{2.23}$$

where $\hat{I}_{\tau}^{CATE}(x)$ is defined as

$$\hat{I}_{\tau}^{CATE}(x) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{D_i (1 - \hat{p}(X_i))}{1 - \hat{G}_1^{KM}(Q_i -)} - \frac{(1 - D_i) \hat{p}(X_i)}{1 - \hat{G}_0^{KM}(Q_i -)} \right) \delta_i Q_i 1\{Q_i < \tau\} 1\{X_i \leq x\}.$$

The discussion for the Cramér-von Mises test is the same and is therefore omitted. Notice that when $\tau = \tau_C$, $\delta 1\{Q < \tau\} = \delta$, and therefore no user-chosen trimming is necessary. This is of particular importance because, in this case, we are using all the information about the average treatment effect available in the data.

In order to proceed with the asymptotic analysis, we need the following integrability assumption, which guarantees that the variance of our proposed estimators is finite and that the censoring effects do not dominate in the right tails. See Stute (1996*a*) for a detailed discussion.

Assumption 2.10 *For $j \in \{0,1\}$, assume the following integrability condition*

$$\mathbb{E}\left[\left(Q_j \gamma_{j,0}\left(Q\right)\right)^2\right] < \infty,$$

$$\mathbb{E}\left[\left|Y_j\right| C_j^{1/2}\left(Y\right)\right] < \infty, \tag{2.24}$$

where

$$C_j\left(w\right) = \int_{-\infty}^{w-} \frac{G_j(dy)}{\left[1 - H_j\left(y\right)\right]\left[1 - G_j\left(y\right)\right]}$$

For a given $\tau \leq \tau_C$, consider the class of local alternatives

$$H_{1,n}^{CATE} : \Upsilon_{\tau}^{CATE}\left(x\right) = \frac{1}{\sqrt{n}} h^{CATE}\left(x\right) \ \forall x \in \mathcal{W}_X, \tag{2.25}$$

that satisfy the following regularity condition.

Assumption 2.11 *Assume that*

(a) $h^{CATE}\left(\cdot\right)$ is an F-integrable function;

(b) the set $h_n^{CATE} \equiv \left[x \in \mathcal{W}_X : h^{CATE}\left(x\right) \neq 0\right]$ has positive Lebesgue measure.

Using an analogous procedure described in Section 2.3.4, let $c_{\alpha,n}^{CATE,*}$ denote the bootstrap critical value of the KS_n^{CATE}. In the next theorem, we prove that, for a given τ, our tests for CATE are asymptotically unbiased, consistent and are able to detect local alternatives of the form of (2.25).

Theorem 2.5 *Suppose Assumptions 2.1-2.8, 2.10 and 2.11 hold. Additionally, assume that for $j \in \{0,1\}$, $\mathbb{E}\left(Y_j|X\right)$ is continuously differentiable of order $m \geq k$, where k is the dimension of X. Then, for a fixed $\tau \leq \tau_C$,*

1. *Under* H_0^{CATE}, $\lim_{n\to\infty} \mathbb{P}_n \left\{ KS_n^{CATE} > c_{\alpha,n}^{CATE,*} \right\} = \alpha.$

2. *Under* H_1^{CATE}, $\lim_{n\to\infty} \mathbb{P}_n \left\{ KS_n^{CATE} > c_{\alpha,n}^{CATE,*} \right\} = 1.$

3. *Under* $H_{1,n}^{CATE}$, $\lim_{n\to\infty} \mathbb{P}_n \left\{ KS_n^{CATE} > c_{\alpha,n}^{CATE,*} \right\} > \alpha.$

From the above discussion, we conclude that with a simple modification of our tests for distributional treatment effects, we can concentrate on tests for average treatment effects. In general, these tests can complement each other.

2.5.2 Testing within the Local Treatment Effect setup

The goal of this section is to show that, in case the unconfoundedness assumption does not hold, that is, if Assumption 2.2 fails, our tests are still applicable to the local average treatment effect (LTE) setup of Imbens and Angrist (1994) and Angrist et al. (1996).

The LTE setup presumes the availability of a binary instrumental variable Z for the treatment assignment. Denote D_0 and D_1 the value that D would have taken if Z is equal to zero or one, respectively. The realized treatment is $D = ZD_1 + (1 - Z) D_0$.

In order to identify the LTE for the subpopulation of compliers, that is, individuals who comply with their actual assignment of treatment and would have complied with the alternative assignment, we need the following assumption.

Assumption 2.12 *(i)* $(Y_0, Y_1, D_1, D_0, C_1, C_0) \perp\!\!\!\perp Z|X$.

(ii) $\forall x \in \mathcal{W}$, *and some* $\varepsilon > 0$,

$$\varepsilon < \mathbb{P}(Z|X = x) \equiv q(x) < 1 - \varepsilon,$$

and

$$\mathbb{P}\left(D_1 = 1|X = x\right) > \mathbb{P}\left(D_0 = 1|X = x\right) \forall x \in \mathcal{W}_X,$$

(iii)$\mathbb{P}\left(D_1 > D_0|X = x\right) = 1 \ \forall x \in \mathcal{W}$.

The null hypothesis of interest in this setup is

$$H_0^L : \Upsilon^L(t, x) = 0 \ \forall (t, x) \in \mathcal{W},$$

where

$$\Upsilon^{L}(t,x)=\mathbb{E}\left[1\{Y_1\leq t\}-1\{Y_0\leq t\}|X=x,Pop=Comp\right].$$

This hypothesis is the analogous of (2.1) within the LTE setup.

In order to proceed, we must express $\Upsilon^{L}(t,x)$ in terms of (Q,δ,D,X). It turns out that, under Assumptions 2.3-2.5 and 2.12, for $(t,x)\in(-\infty,\tau_C)\times(-\infty,\infty)^k$,

$$\Upsilon^{L}(t,x)=\mathbb{E}\left[\frac{Z\delta 1\{Q\leq t\}}{1-G_1(Q-)\,q(X)}-\frac{(1-Z)\,\delta 1\{Q\leq t\}}{(1-q(X))\,1-G_0(Q-)}|X=x\right]\Bigg/\mathbb{E}\left[\frac{ZD}{q(X)}-\frac{(1-Z)\,D}{1-q(X)}|X=x\right]$$

From Assumption 2.12, the denominator of $\Upsilon^{L}(\cdot,\cdot)$ is always strictly positive. Therefore, from the discussion in Section 2.2, the hypothesis of zero conditional distribution treatment effect among compliers can be equivalently written as

$$H_0^{L}:I^{L}(t,x)=0\;\forall(t,x)\in\mathcal{W},$$

where

$$I^{L}(t,x)=\mathbb{E}\left[\left(\frac{Z(1-q(X))}{1-G_1(Q-)}-\frac{(1-Z)\,q(X)}{1-G_0(Q-)}\right)\delta 1\{Q\leq t\}\,1\{X\leq x\}\right].$$

Noticing that once we replace Z to D, and $q(x)$ to $p(x)$, $I^{L}(t,x)$ is equal $I(t,x)$, that is, the LTE framework reduces to the unconfounded framework. Therefore, we conclude that our tests for zero treatment effects with censored data are valid not only when the treatment assignment is unconfounded, but also to a particular case when the selection to treatment is based on unobservables, namely the local treatment effect setup of Imbens and Angrist (1994) and Angrist et al. (1996).

2.5.3 Dynamic treatment assignments

Until now, all proposed tests rely on individuals entering the treatment at the beginning of the spell. Nonetheless, this setup might be restrictive for some important applications where the treatment might start at any time. A leading example is the active labor market policy (ALMP) programs for the unemployed. The common feature of ALMP is that participation is not instantaneous upon inflow into unemployment, but individuals are observed to enter ALMP programs at any possible time since the start of the unemployment spell. This dynamic selection mechanism introduces some potential problems to select a proper control group. The main issue within this dynamic setup is that individuals currently non-treated might become treated later. Given that the probability of enrollment increases with the elapsed duration, the treatment status depends on the outcome, and therefore, unconfoundedness-based tests like ours may not be suitable. Nonetheless, in this subsection we show that, by focusing on the effect of treatment now versus continuing to wait for treatment, as initially proposed by Sianesi (2004), our test statistics are still suitable.

In order to formalize this idea, we need to introduce some additional notation. The eligible population at time u are those still in the state of interest after u periods. For the eligibles at u, denote $D^{(u)} = 1$ for joining a program at u and $D^{(u)} = 0$ for not joining at least up to u. Denote $Y_1^{(u)}$ and $Y_0^{(u)}$ as the potential outcomes if treated at u and not yet treated up to u, respectively. Note that the potential outcomes $Y_1^{(u)}$ and $Y_0^{(u)}$ are only defined for those who are still in the state of the interest at time u, that is, only for those $Y_1^{(u)} > u$, $Y_0^{(u)} > u$. Assume that $\mathbb{P}(Y_1^{(u)} > u|X)$ and $\mathbb{P}(Y_0^{(u)} > u|X)$ is always between ε and $1-\varepsilon$, for some $\varepsilon > 0$.

The conditional distributional treatment effect is given by

$$\Upsilon^{(u)}(t,x) = \mathbb{E}\left[1\{Y_1^{(u)} \leq t\} - 1\{Y_0^{(u)} \leq t\}|X = x, Y_1^{(u)} > u, Y_0^{(u)} > u\right].$$

Under Assumptions 2.2-2.5, but with $D^{(u)}, Y_1^{(u)}$ and $Y_0^{(u)}$ playing the role of D, Y_1 and Y_0, we have that

$$\Upsilon^{(u)}(t,x) = \mathbb{E}\left[\left(\frac{D^{(u)}\delta 1\{Q \leq t\}}{(1-G_1(Q-))\,p(X)} - \frac{\left(1-D^{(u)}\right)\delta 1\{Q \leq t\}}{(1-p(X))\,(1-G_0(Q-))}\right)|X = x, Q > u\right], \tag{2.26}$$

for $(t, x) \in (-\infty, \tau_C) \times (-\infty, \infty)^k$.

Notice that (2.26) is nothing more than (2.2) restricted to the subpopulation of those who are still at the state of interest at time u. Therefore, once this restriction is applied to the observed data, all the test statistics previously described can be straightforwardly used. Hence, we conclude that our proposal is also suitable for the case of dynamic treatment assignment.

2.6 Evaluation of labor market programs

In this section, we demonstrate that our proposed tests can be useful in practice. We consider one application with experimental data, the Illinois Reemployment Bonus Experiment, and one with observational data, a female job training in Korea.

2.6.1 Illinois Reemployment Bonus Experiment

In this section we analyze data from the Illinois Reemployment Bonus Experiments, which is freely available at the W.E. Upjohn Institute for Employment Research. From mid-1984 to mid-1985, the Illinois Department of Employment Security conducted a social experiment to test the effectiveness of bonus offers in reducing the duration of insured unemployment At the beginning of each claim, the experiment randomly divided newly unemployed people into three groups:

1. Job Search Incentive Group (JSI). The members of this group were told that they would qualify for a cash bonus of \$500, which was about four times the average weekly unemployment insurance benefits, if they found a full-time job within eleven weeks of benefits, and if they held that job for at least four months. 4816 claimants were assigned to this group.

2. Hiring Incentive Group (HI). The members of this group were told that their employer would qualify for a cash bonus of \$500 if the claimant found a full-time job within eleven weeks of benefits, and if they held that job for at least four months. 3963 claimants were assigned to this group.

3. Control Group. All claimants not assigned to the other groups. These members did not know that the experiment was taking place. 3952 individuals were assigned to this group.

Several studies including Woodbury and Spiegelman (1987), Meyer (1996) and Bijwaard and Ridder (2005*a*) have analyzed the impact of the reemployment bonus on the unemployment duration measured by the number of weeks receiving unemployment insurance. It is important to emphasize that spells which reached the maximum amount of benefits or the state maximum number of weeks, 26, are censored, leading to censoring proportions of 38, 41 and 42 percent for the JSI, HI and the control group, respectively. Apart from the duration data, some information about claimants' background characteristics is also available: age, gender (Male =1), ethnicity (White =1), pre-unemployment earning and the weekly unemployment insurance benefits amount. For a complete description of the experiment and the available dataset, see Woodbury and Spiegelman (1987).

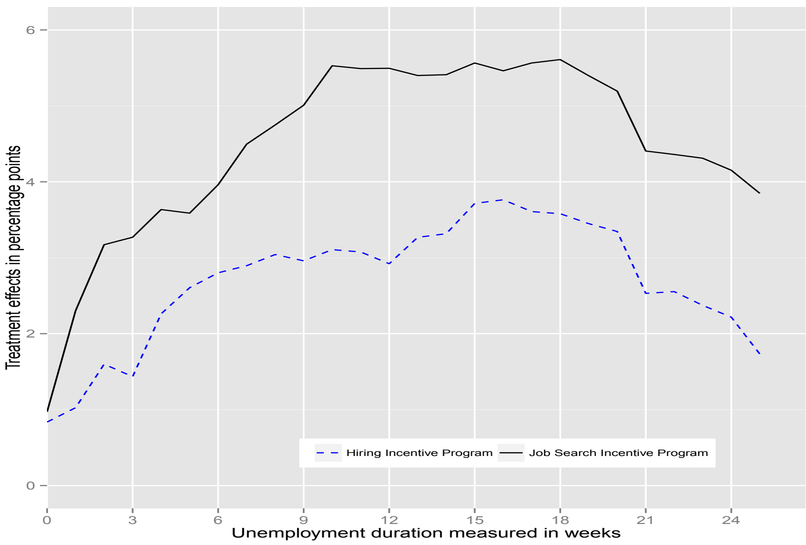
We start our analysis by plotting in Figure 2.1 the estimated overall treatment effect for the Job Search Incentive and the Hiring Incentive groups. From the discussion in Section 2.2.3, one can estimate $\Upsilon(t)$ by

$$\hat{\Upsilon}(t) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i\delta_i 1\{Q_i < \tau\}}{1-\hat{G}_1^{KM}(Q_i-)\,\hat{p}(X_i)} - \frac{(1-D_i)\,\delta_i 1\{Q_i < \tau\}}{\left(1-\hat{G}_0^{KM}(Q_i-)\right)(1-\hat{p}(X_i))}\right),$$

where, given the experimental design, $\hat{p}(\cdot) = n^{-1}\sum_{i=1}^{n} D_i$, which is numerically the same as the series logit estimator using a power series of order zero. Notice that both treatments seems to short the unemployment duration when compared to the control group, with the effects of the JSI group a bit larger than those of HI group.

We are focused on evaluating the effectiveness of the unemployment bonus in affecting the unemployment duration for all subgroups characterized by observable characteristics, and not just the overall effect as displayed in Figure 2.1. To this end, we perform our test for zero CDTE. We compare our results with the one using the semi-parametric Cox (1972) proportional hazard model.

Figure 2.1: Distributional treatment effects of Illinois reemployment bonus program on unemployment duration



The results of the tests are reported in Table 2.2. Using our nonparametric proposals, we reject the null hypothesis of zero CDTE at the 5% level for both treatment groups. Therefore, our tests suggest that the bonus experiment in Illinois were able to affect the unemployment duration. On the other hand, if one uses the proportional hazard model, one cannot reject the null of zero effect for all subpopulations in the hiring incentive group at the 5% level. In fact, by means of Grambsch and Therneau (1994)'s test, the proportionality assumption is rejected in the data at the 1% level. This illustrates how using parametric models to assess the existence of treatment effects might lead to "erroneous" conclusions.

One might be also interested is assessing the direction of the treatment effect, i.e., if the unemployment bonus program has led to a shorter or longer unemployment duration. Given the design of the Illinois experiment, it is plausible to assume that offering a reemployment bonus for the unemployed cannot lead to longer unemployment spells than in the control group, i.e., we might exclude the possibility that the treatment is "harmful", that is, we can impose that

$$F_1(t|X=x) \geq F_0(t|X=x) \;\; \forall (t,x) \in \mathcal{W}. \tag{2.27}$$

With (2.27), we can focus on single direction of departure of the null hypothesis of no distributional treatment effects for all subpopulations characterized by covariates. That is, with the additional information in (2.27), we can test

$$H_0^{one} : F_1\left(t|X=x\right) = F_0\left(t|X=x\right) \ \forall \left(t,x\right) \in \mathcal{W},$$

against

$$\begin{aligned} H_1^{one} : & F_1\left(t|X=x\right) \geq F_0\left(t|X=x\right) \ \forall \left(t,x\right) \in \mathcal{W} \\ & with\ strict\ inequality\ for\ some\ \left(t,x\right) \in \mathcal{W} \end{aligned}$$

using the test statistic

$$KS_n^{one} = \sup_{(t,x)\in\mathcal{W}} \sqrt{n}\hat{I}_n(t,x). \tag{2.28}$$

Critical values are computed as described in Section 2.3.4.

As shown in Table 2.2, we reject the null of zero conditional treatment effects in favor of the one sided alternative that the treatment is non-negative (not-harmful) for all subpopulations, in both treatment groups. Even tough we excluded the possibility of a "negative treatment effect" for the Illinois experiment, as a "robustness check", we also consider the other one-sided alternative, i.e., the one in which the treatment is non-positive (not-helpful) for all subpopulations. In fact, we fail to reject our null hypothesis of zero conditional treatment effect for both treatment groups. Therefore, this evidence suggest that the bonus experiment has reduced the unemployment duration.

An important aspect of the Illinois Reemployment Bonus Experiment is that participation was not mandatory. Once claimants were assigned to the treatment groups, they were asked if they would like to participate in the demonstration or not. For those selected to the Job Search Incentive group, 84% agreed to participate, whereas just 65% of the Hiring Incentive group agreed to participate. This non-compliance issue may raise some selection bias issue. Therefore, the performed tests might be interpreted as tests for zero distributional "intention to treat" effects. Nonetheless, one may be willing to some

Table 2.2: Bootstrap p-values for conditional distributional treatment effects tests for the Illinois bonus experiment.

Intention to Treat		
Treatment Effect tests	Job Search Incentive	Hiring Incentive
Two sided	0.000	0.030
One sided - Not harmful	0.000	0.015
One sided - Not helpful	0.684	0.987
Proportional Hazard Model	0.000*	0.059*
Local Treatment Effects - Compliers		
Treatment Effect tests	Job Search Incentive	Hiring Incentive
Two sided	0.000	0.025
One sided - Not harmful	0.000	0.012
One sided - Not helpful	0.885	0.935

Note: 10,000 bootstrap replications. * denotes p-value based on Gaussian approximation.

extent disentangle the effects of participation and the effects of actual treatment. Using the random assignment as an instrumental variable for the actual participation in the demonstration, we adopt the LTE framework described in Section 2.5.2, using a power series of order two[2]. The results for these tests are displayed in the second part of Table 2.2, and the conclusions of our tests are not changed. Therefore, we argue that indeed there is statistical evidence that the unemployment bonus experiment has helped their participants shorten their unemployment spell in Illinois.

2.6.2 Female job training in Korea.

Our method can also be used with observational data. Therefore, we analyze female job training data from the Department of Labor in South Korea in which the control group consist of unemployed claimants who chose to receive unemployment insurance instead of job training. This dataset is also used by Lee (2009). The data represents about 20% of the Korea population who became unemployed during January 1999 to the end of 1999 and either received job training or used unemployment insurance up to the end of 1999. All individuals were followed until the end of March 2000. Therefore, from the

2. The results are robust to different choices of the number of power series considered in the estimation of the propensity score.

design of the data, the maximum unemployment duration is 450 days, that is, $\tau_C = 450$. Nonetheless, in the dataset, all observations with unemployment duration beyond 420 days are censored. There are 9312 members individuals in the control group and 1554 in the treatment group. For a complete description of the dataset and the characteristics of the job training program, see Lee and Lee (2005).

In addition to the unemployment duration information, we use as covariates informations on the individual characteristics such as the number of days that the woman worked at her last workplace, education (completed high school=1), age in years, and four ex-job categories (1-executive, professional or semiprofessional; 2-clerical; 3-service or sales; 4-mechanic, assembler, operator and menial labour). In the dataset, the proportion of censoring is around 70% for both treated and control groups. Notice that the duration for the treated group includes the duration of job training, which average duration was about 3 months.

As is usual in the policy evaluation literature, we first estimate the unconditional average trimmed treatment effect, $\mathbb{E}\left[Y_1 1\left\{Y_1 < \tau_C\right\} - Y_0 1\left\{Y_0 < \tau_C\right\}\right]$, by

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{D_i\delta_i Q_i}{1-\hat{G}_1^{KM}\left(Q_i-\right)\hat{p}\left(X_i\right)} - \frac{\left(1-D_i\right)\delta_i Q_i}{\left(1-\hat{G}_0^{KM}\left(Q_i-\right)\right)\left(1-\hat{p}\left(X_i\right)\right)}\right). \tag{2.29}$$

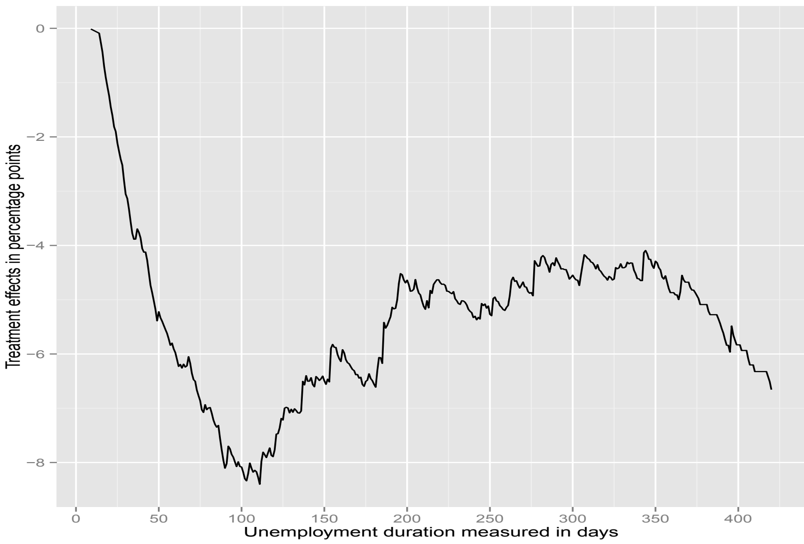
The ATE point estimate is approximately -5 days, i.e. the job training had reduced the overall unemployment duration by 5 days. Nonetheless, following an analogous procedure as describe in Section 2.3, we are not able to reject the null hypothesis that the ATE is equal to zero at the conventional levels. Therefore, looking at the *unconditional* ATE, one may argue that the unemployment duration for those who receive unemployment insurance and those who received job training are the same.

Next, we consider the unconditional distributional treatment effect. Figure 2.2 plots $\hat{\Upsilon}\left(t\right)$, but with the propensity score $p\left(\cdot\right)$ estimated with the series logit estimator using a power series of order two[3]. From the plot, one may argue that job training leads to an increase in the unemployment duration of female Koreans. Indeed, we reject the null

3. Our results are not sensitive to different choices of the number of power series included.

hypothesis that the unconditional DTE is zero. From the results of the unconditional ATE and DTE tests, it seems that the job training has had an effect at the unemployment distribution at some point other than the average.

Figure 2.2: Distributional treatment effects of job training on female unemployment duration in South Korea



In order to analyze if this conclusion holds true after conditioning on a vector of individual characteristics, we apply our tests for both zero conditional average treatment effect and for zero conditional distributional treatment effect, using age, last firm employment days, education level and job categories as covariates. For comparison, we also apply Lee (2009)'s test for zero CDTE based on a 'two-sample" covariate-matching procedure. To avoid dimensionality problems, we only consider matching on the propensity score (estimated with a probit), with bandwidth equal to $0.62n^{-1/5}$, and the bi-weight kernel $K(z) = (15/16)(1 - z^2)^2 1\{|z| < 1\}$.

The results are presented in Table 2.3. Using both Lee (2009)'s and our proposal, we find evidence of treatment effect has an effect on the distributional of unemployment duration at the 5% level. Furthermore, we reject the null hypothesis of zero conditional average treatment effect at the 5% level, which is in contrast to the unconditional case. Hence, we conclude that, after conditioning on a vector of observables, we are able to point out the direction of the departure of the null hypothesis (2.1). This illustrates the complementary of our tests for zero CDTE and CATE, and the additional information these tests can provide when compared to their unconditional counterparts.

Table 2.3: Tests for the Korean job training data . Bootstrap p-values

Treatment Effect tests	Bootstrap p-value
Cond. Dist. Treat. Effect	0.000
Cond. Aver. Treat. Effect	0.004
Cond. Dist. Treat. Effect - Lee(2009)	0.000 *

Note: 10,000 bootstrap replications. * denotes p-value based on Gaussian approximation.

2.7 Conclusion and suggestions for further research

In this paper we proposed different nonparametric tests for treatment effects when the outcome of interest is censored. Once we transform our conditional moment restrictions into an infinite number of moments, we characterize our tests statistics as Kaplan-Meier integrals that can be easily estimated from the observed data. Our tests have asymptotically correct size, are able to detect local alternatives converging to the null hypothesis of interest at the parametric rate $n^{-1/2}$, and are consistent against fixed alternatives. Our simulation study provide evidence that our tests have good finite sample properties. We provide two empirical applications to demonstrate that our tests can be useful in practice.

Our results can be extended to other situations of practical interest. For instance, an interesting extension of our results consists of testing conditional stochastic dominance when the outcome of interest is a duration. In the context of fully observed data, conditional stochastic dominance has recently attracted a lot of interest. See, for example, Lee and Whang (2009), Delgado and Escanciano (2013), and Lee et al. (2013). Adopting either Delgado and Escanciano (2013)'s or Andrews and Shi (2013, 2014)'s approach, one can extend our proposal to the stochastic dominance analysis to censored outcomes.

Another important extension would be to allow the covariates distribution to be different in the treatment and control groups by introducing covariate-matching techniques. With these techniques, the use of smooth estimators cannot be avoided. In particular, proposals by Cabus (1998), Neumeyer and Dette (2003), and Srihera and Stute (2010) designed for testing the equality of nonparametric regression curves in a two-sample context with fully observed data, can be adapted to handle randomly censored outcomes if one use Kaplan-Meier integrals as in this article. For a related approach with censored outcomes, see Pardo-Fernandez and Van Keilegom (2006) and Lee (2009). A detailed analysis of these extensions is beyond the scope of this article and is deferred to future work.

2.8 Appendix

In this appendix, we prove our main results. Before proving the main results of the article, we first introduce some notation. For a generic set $\mathcal{G}$, let $l^{\infty}(\mathcal{G})$ be the Banach space of all uniformly bounded real functions on $\mathcal{G}$ equipped with the uniform metric $\|f\|_{\mathcal{G}} \equiv \sup_{z\in\mathcal{G}} |f(z)|$. We consider convergence in distribution of empirical processes in the metric space $\left(l^{\infty}(\mathcal{G}), \|f\|_{\mathcal{G}}\right)$ in the sense of J. Hoffman-Jørgensen (see, e.g. , van der Vaart and Wellner (1996)). For any generic Euclidean random vector ξ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $\chi_{\boldsymbol{\xi}}$ denotes its state space and $P_{\boldsymbol{\xi}}$ denotes its induced probability measure with corresponding distribution function $F_{\xi}(\cdot) = P_{\xi}(-\infty, \cdot]$. Given *iid* observations $\{\xi_i\}_{i=1}^n$ of ξ, $\mathbb{P}_{\xi_n} f \equiv n^{-1}\sum_{i=1}^n f(\xi_i)$. Let $F_{\xi_n}(\cdot) \equiv \mathbb{P}_{\xi_n}(-\infty, \cdot]$ be the corresponding empirical CDF. Likewise, the expectation is denoted by $P_{\xi} f \equiv \int f dP$. The empirical process evaluated at f is $\mathbb{G}_{\xi_n} f$ with $\mathbb{G}_{\xi_n} \equiv \sqrt{n}\left(\mathbb{P}_{\xi_n} - P_{\xi_n}\right)$. Let $\|\cdot\|_{2,P}$ be the $L_2(P)$ norm, that is, $\|f\|_{2,P}^2 = \int f^2 dP$. When P is clear from the context, we simply write $\|\cdot\|_2 \equiv \|\cdot\|_{2,P}$. Let $|\cdot|$ denote the Euclidean norm, that is, $|A|^2 = A'A$. For a measurable class of functions $\mathcal{G}$ from χ_Z to $\mathbb{R}$, let $\|\cdot\|$ be a pseudo-norm on $\mathcal{G}$, that is, a norm except for the property $\|f\| = 0$ does not imply $f = 0$. Let $N(\varepsilon, \mathcal{G}, \|\cdot\|)$ be the covering number with respect to $\|\cdot\|$ needed to cover $\mathcal{G}$. Given two functions $l, u \in \mathcal{G}$, the bracket $[l, u]$ is the set of functions $f \in \mathcal{G}$ such that $l \leq f \leq u$. An ε-bracket with respect to $\|\cdot\|$ is a bracket $[l, u]$ with $\|l - u\| \leq \varepsilon$. The covering number with bracketing $N_{[\cdot]}(\varepsilon, \mathcal{G}, \|\cdot\|)$ is the minimal number of ε-brackets with respect to $\|\cdot\|$ needed to cover $\mathcal{G}$. Define

$\mathcal{S} \equiv (-\infty, \tau_C) \times \chi_X$. Throughout the appendix, denote $\mathcal{C}$ as a generic constant that may change from expression to expression.

First, we present the proof of the identification result in Proposition 2.1.

Proof of Proposition 2.1: By Assumptions 2.2-2.4 and the law of iterated expectations, we have, for $t < \tau_C$

$$
\begin{aligned}
&\mathbb{E}\left[\frac{D\delta 1\{Q \leq t\}}{(1 - G_1(Q-))\, p(X)}\middle| X\right] \\
&= \mathbb{E}\left[\frac{1\{Y_1 \leq t\}}{(1 - G_1(Y_1-))\, p(X)} \mathbb{E}\left[D\delta_1 | X, Y_1\right] \middle| X\right] \\
&= \mathbb{E}\left[\frac{1\{Y_1 \leq t\}}{(1 - G_1(Y_1-))\, p(X)} \mathbb{E}\left[D | X\right] \mathbb{E}\left[\delta_1 | X, Y_1\right] \middle| X\right] \\
&= \mathbb{E}\left[\frac{1\{Y_1 \leq t\}}{(1 - G_1(Y_1-))\, p(X)} p(X) (1 - G_1(Y_1)) \middle| X\right] \\
&= \mathbb{E}\left[1\{Y_1 \leq t\} | X\right],
\end{aligned}
$$

where the first equality follows from the law of iterated expectations, the second from Assumption 2.2, and the third from Assumption 2.4. Assumption 2.3 guarantees that the expectation is well defined.

Adopting the analogous steps for the control group,

$$
\begin{aligned}
&\mathbb{E}\left[\frac{(1 - D)\, \delta 1\{Q \leq t\}}{(1 - G_0(Q-))\,(1 - p(X))}\middle| X\right] \\
&= \mathbb{E}\left[1\{Y_0 \leq t\} | X\right].
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[\left(\frac{D\delta 1\{Q \leq t\}}{(1 - G_1(Q_1-))\, p(X)} - \frac{(1 - D)\, \delta 1\{Q \leq t\}}{(1 - G_0(Q_1-))\,(1 - p(X))}\right) \middle| X\right] \\
&= \Upsilon(t, x)
\end{aligned}
$$

concluding the proof. ■

Next, we state an auxiliary result from the empirical process literature. Define the generic class of measurable functions $\mathcal{G} \equiv \{Z \to m(Z, \theta, h) : \theta \in \Theta, h \in \mathcal{H}\}$, where Θ and

$\mathcal{H}$ are endowed with the pseudo-norms $|\cdot|_\Theta$ and $|\cdot|_\mathcal{H}$. The following result is part of Theorem 3 in Chen et al. (2003).

Lemma 2.2 *Assume that for all $(\theta_0, h_0) \in \Theta \times \mathcal{H}$, $m(Z,\theta,h)$ is locally uniformly $L_2(P)$ continuous, in the sense that*

$$E\left[\sup_{\theta:|\theta_0-\theta|_\Theta<\delta,\ h:|h_0-h|_\mathcal{H}<\delta} |m(Z,\theta,h) - m(Z,\theta_0,h_0)|^2 \le \mathcal{C}\delta^s\right]$$

for all sufficiently small $\delta > 0$ and some constant $s \in (0,2]$. Then,

$$\begin{aligned} N_{[\cdot]}\left(\varepsilon, \mathcal{G}, \|\cdot\|_2\right) \le N &\left(\left(\frac{\varepsilon}{2\mathcal{C}}\right)^{2/s}, \Theta, \|\cdot\|_\Theta\right) \\ &\times N\left(\left(\frac{\varepsilon}{2\mathcal{C}}\right)^{2/s}, \mathcal{H}, \|\cdot\|_\mathcal{H}\right). \end{aligned}$$

Before we introduce the proofs of our main theorems, we prove two useful lemmas that are crucial to the derivation of our result. Recall that, for a given $(t,x) \in \mathcal{W}$,

$$\begin{aligned} \xi_1(\bar{y},\bar{x},\bar{z};t,x) &= \bar{z}(1-p(\bar{x}))\,1\{\bar{y}\le t\}\,1\{\bar{x}\le x\}, \\ \xi_0(\bar{y},\bar{x},\bar{z};t,x) &= (1-\bar{z})\,p(\bar{x})\,1\{\bar{y}\le t\}\,1\{\bar{x}\le x\}. \end{aligned}$$

where $p(\cdot)$ is the true propensity score. Define the infeasible estimator

$$\begin{aligned} \bar{I}(t,x) = &\int \xi_1(\bar{y},\bar{x},\bar{z};t,x)\hat{F}^{KM}(d\bar{y},d\bar{x},\bar{z}=1) \\ &- \int \xi_0(\bar{y},\bar{x},\bar{z};t,x)\hat{F}^{KM}(d\bar{y},d\bar{x},\bar{z}=0), \end{aligned} \tag{2.30}$$

the analogous of (2.10) but with the true propensity score.

Lemma 2.3 *Under Assumptions 2.1-2.5,*

$$\sup_{(t,x)\in\mathcal{S}}\left|\bar{I}(t,x) - \frac{1}{n}\sum_{i=1}^n \eta_i(t,x)\right| = o\left(n^{-1/2}\right)$$

Proof To prove this lemma, it suffices to apply Theorem 1 of Sellero et al. (2005). Toward this goal, define the following class of real-valued measurable functions on $\chi_Y \times \chi_X \times \{0,1\}$

$$\mathcal{G}_1 \equiv \{(\bar{\omega}, \bar{x}, \bar{z}) \to \xi_1(\bar{\omega}, \bar{x}, \bar{z}; t, x) : (t, x) \in \mathcal{S}\}. \tag{2.31}$$

Notice that $\mathcal{G}_1$ is a VC-subgraph class of functions with VC index smaller or equal than $k+2$ and admits the envelope $\Phi(\bar{\omega}, \bar{x}, \bar{z}) = 1$ that satisfies the required moment conditions of Theorem 1 of Sellero et al. (2005). The same holds for the class of functions

$$\mathcal{G}_2 \equiv \{(\bar{\omega}, \bar{x}, \bar{z}) \to \xi_0(\bar{y}, \bar{x}, \bar{z}; t, x) : (t, x) \in \mathcal{S}\}. \tag{2.32}$$

Hence, we have

$$\bar{I}(t, x) = \frac{1}{n}\sum_{i=1}^{n} \eta_i(t, x) + R_n(t, x) \tag{2.33}$$

where

$$\sup_{(t,x)\in\mathcal{S}} |R_n(t, x)| = O\left(\frac{\ln^3 n}{n}\right) \text{ a.s.,}$$

concluding our proof. ■

In the next lemma we focus on the treated group. The result for the control group is analogous.

Lemma 2.4 *Under Assumptions 2.1-2.8, we have, uniformly in* $(t, x) \in \mathcal{S}$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{1-\hat{G}_1^{KM}(Q_i-)} - \frac{1}{1-G_1(Q_i-)}\right)\delta_i D_i 1\{Q_i \le t\} 1\{X_i \le x\}(\hat{p}(X_i) - p(X_i))$$
$$= o_{\mathbb{P}}\left(n^{-1/2}\right),$$

where $G_1(t-) \equiv \mathbb{P}(C_1 < t)$.

Proof Denote

$$Z_1(t) = \frac{\hat{G}_1^{KM}(t-) - G_1(t-)}{1-\hat{G}_1^{KM}(t-)}.$$

Now, one can write

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\hat{G}_{1}^{KM}\left(Q_{i}-\right)}\delta_{i}D_{i}1\left\{Q_{i}\leq t\right\}1\left\{X_{i}\leq x\right\}\left(\hat{p}\left(X_{i}\right)-p\left(X_{i}\right)\right)\\
&=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-G_{1}\left(Q_{i}-\right)}\delta_{i}D_{i}1\left\{Q_{i}\leq t\right\}1\left\{X_{i}\leq x\right\}\left(\hat{p}\left(X_{i}\right)-p\left(X_{i}\right)\right)\\
&+\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-G_{1}\left(Q_{i}-\right)}Z_{1}\left(Q_{i}\right)\delta_{i}D_{i}1\left\{Q_{i}\leq t\right\}1\left\{X_{i}\leq x\right\}\left(\hat{p}\left(X_{i}\right)-p\left(X_{i}\right)\right)\\
&\equiv A_{1}\left(t,x\right))+A_{2}\left(t,x\right).
\end{aligned}
$$

It suffices to show that, uniformly in $(t,x)\in\mathcal{W}$, $\sqrt{n}A_{2}\left(\cdot,\cdot\right)=o_{\mathbb{P}}\left(1\right)$. First, rewrite $\sqrt{n}A_{2}\left(\cdot,\cdot\right)$ as

$$
\begin{aligned}
\sqrt{n}A_{2}\left(t,x\right)=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_{i}\left(\hat{G}_{1}^{KM}\left(Q_{i}-\right)-G_{1}\left(Q_{i}-\right)\right)\frac{1-G_{1}\left(Q_{i}-\right)}{1-\hat{G}_{1}^{KM}\left(Q_{i}-\right)}\frac{1}{\left(1-G_{1}\left(Q_{i}-\right)\right)^{2}}\\
&\times D_{i}\delta_{i}1\left\{Q_{i}\leq t\right\}1\left\{X_{i}\leq x\right\}\left(\hat{p}\left(X_{i}\right)-p\left(X_{i}\right)\right).
\end{aligned}
$$

We have

$$
\begin{aligned}
\sqrt{n}\sup_{(t,x)\in\mathcal{W}}\left|A_{2}\left(t,x\right)\right|&\leq\mathcal{C}\sqrt{n}\sup_{t}\left|\hat{G}_{1}^{KM}\left(t\right)-G_{1}\left(t\right)\right|\times\sup_{t}\left|\frac{1-G_{1}\left(t\right)}{1-\hat{G}_{1}^{KM}\left(t\right)}\right|\times\sup_{x}\left|\left(\hat{p}\left(x\right)-p\left(x\right)\right)\right|\\
&=O_{\mathbb{P}}(1)\times O_{\mathbb{P}}(1)\times\left[O_{\mathbb{P}}\left(\sqrt{\frac{L^{3}}{n}}\right)+O\left(L^{-\frac{s}{2k}+1}\right)\right]
\end{aligned}
$$

in which the last step follows from

$$
\sqrt{n}\sup_{t}\left|\hat{G}_{1}^{KM}\left(t\right)-G_{1}\left(t\right)\right|=O_{\mathbb{P}}\left(1\right), \tag{2.34}
$$

$$
\sup_{t}\left|\frac{1-G_{1}\left(t\right)}{1-\hat{G}_{1}^{KM}\left(t\right)}\right|=O_{\mathbb{P}}\left(1\right), \tag{2.35}
$$

$$
\sup_{x}\left|\left(\hat{p}_{n}\left(x\right)-p\left(x\right)\right)\right|=O_{\mathbb{P}}\left(\sqrt{\frac{L}{n}^{3}}\right)+O\left(L^{-\frac{s}{2k}+1}\right) \tag{2.36}
$$

see Gill (1983) for (2.34), Zhou (1992) for (2.35), and Hirano et al. (2003) for (2.36).

Taking $L = a \cdot N^v$ as in Assumption 2.8,

$$O_{\mathbb{P}}\left(\sqrt{\frac{L^3}{n}}\right) + O_{\mathbb{P}}\left(L^{-\frac{s}{2k}+1}\right) = O_{\mathbb{P}}\left(n^{\frac{3v-1}{2}}\right) + O\left(n^{-\left(\frac{s}{2k}+1\right)v}\right)$$
$$= o_{\mathbb{P}}(1)$$

if $v < 1/3$ and $s/k > 2$. From Assumptions 2.7 and 2.8, these two conditions are fulfilled, concluding our proof. ■

Now we are ready to proceed with the proofs of our main results.

Proof of Lemma 2.1: Notice that

$$\hat{I}(t,x) = \bar{I}(t,x)$$
$$+ \frac{n_1}{n}\sum_{i=1}^{n_1} W_{1,i:n_1}\left(\hat{\xi}_1\left(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1[i:n_1]}; t, x\right) - \xi_1\left(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1[i:n_1]}; t, x\right)\right) \tag{2.37}$$
$$- \frac{n_0}{n}\sum_{l=1}^{n_0} W_{0,l:n_0}\left(\hat{\xi}_0\left(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t, x\right) - \xi_0\left(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t, x\right)\right)$$

First, by Lemma 2.3, we have that, uniformly in $(t,x) \in \mathcal{S}$,

$$\bar{I}(t,x) = \frac{1}{n}\sum_{i=1}^{n} \eta_i(t,x) + o_{\mathbb{P}}\left(n^{1/2}\right). \tag{2.38}$$

We now focus on the second term of (2.37). Our goal is to show that

$$\sup_{(t,x)\in\mathcal{S}}\left|\frac{n_1}{n}\sum_{i=1}^{n_1} W_{1,i:n_1}\left(\hat{\xi}_1\left(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x\right) - \xi_1\left(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x\right)\right)\right. \tag{2.39}$$
$$\left. - \frac{1}{n}\sum_{i=1}^{n} F_1(t|X_i)\, p(X_i)\, 1\{X_i \le x\}\,(D_i - p(X_i))\right| = o_{\mathbb{P}}\left(n^{-1/2}\right)$$

As discussed in Section 2.3, we have that

$$W_{1,i:n_1} = \frac{\delta_{[i:n_1]}}{n_1} \frac{1}{1 - \hat{G}_1^{KM}(Q_{i:n_1}-)},$$

where $\hat{G}_1^{KM}$ is the Kaplan-Meier estimator of G_1. By Lemma 2.4, we have that, uniformly in $(t, x) \in \mathcal{S}$

$$\begin{aligned}
&\frac{n_1}{n} \sum_{i=1}^{n_1} W_{1,i:n_1} \left(\hat{\xi}_1 \left(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x \right) - \xi_1 \left(Q_{i:n_1}, X_{[i:n_1]}, D_{[i:n_1]}; t, x \right) \right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{D_i \delta_i}{1 - G_1(Q_i-)} \left(\hat{\xi}_1 (Q_i, X_i, D_{i:}; t, x) - \xi_1 (Q_i, X_i, D_i; t, x) \right) + o_{\mathbb{P}} \left(n^{-1/2} \right), \quad (2.40)
\end{aligned}$$

that is, that there is no estimation effect due to the replacing G_1 by estimation of $\hat{G}_1^{KM}$ in the second term of (2.37).

By adding and subtracting a number of terms, we have that (2.40) is equal to

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\frac{D\delta_i}{1 - G_1(Q_i-)} \left(\hat{\xi}_1 (Q_i, X_i, D_{i:}; t, x) - \xi_1 (Q_i, X_i, D_i; t, x) \right) \\
&\quad - \int_{\mathcal{X}} F_1(t, x|\bar{x}) p(\bar{x}) (\hat{p}_n(\bar{x}) - p(\bar{x})) \mathbb{P}(d\bar{x})] \quad (2.41) \\
&\quad + \left[\sqrt{n} \int_{\mathcal{X}} F_1(t, x|\bar{x}) p(\bar{x}) (\hat{p}_n(\bar{x}) - p(\bar{x})) \mathbb{P}(d\bar{x}) \right. \quad (2.42) \\
&\qquad \left. - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{u}_n(X_i) \frac{(D_i - p_L(X_i))}{\sqrt{p_L(X_i)(1 - p_L(X_i))}} \right] \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{u}_n(X_i) - u_n(X_i)) \frac{(D_i - p_L(X_i))}{\sqrt{p_L(X_i)(1 - p_L(X_i))}} \quad (2.43) \\
&\quad + \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} u_n(X_i) \frac{(D_i - p_L(X_i))}{\sqrt{p_L(X)(1 - p_L(X))}} - u(X_i) \frac{(D_i - p(X_i))}{\sqrt{p(X_i)(1 - p(X_i))}} \right) \\
&\qquad (2.44) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} F_1(t|X_i) 1\{X_1 \leq x\} p(X_i)(D_i - p(X_i)) \quad (2.45)
\end{aligned}$$

where

$$
\begin{aligned}
\tilde{u}_n(z) &= \int_{\mathcal{X}} F_1\left(t, x|\bar{x}\right) p\left(\bar{x}\right) \mathcal{L}'\left(R^L\left(\bar{x}\right)' \tilde{\pi}_L\right) R^L\left(\bar{x}\right)' \mathbb{P}\left(d\bar{x}\right) \tilde{\Sigma}_L^{-1} \sqrt{p_L\left(X_i\right)\left(1 - p_L\left(X_i\right)\right)} R^L\left(z\right), \\
u_n(z) &= \int_{\mathcal{X}} F_1\left(t, x|\bar{x}\right) p\left(\bar{x}\right) \mathcal{L}'\left(R^L\left(\bar{x}\right)' \pi_L\right) R^L\left(\bar{x}\right)' \mathbb{P}\left(d\bar{x}\right) \Sigma_L^{-1} \sqrt{p_L\left(X_i\right)\left(1 - p_L\left(X_i\right)\right)} R^L\left(z\right), \\
u(z) &= F_1\left(t|z\right) 1\{z \leq x) p\left(z\right) \sqrt{p\left(z\right)\left(1 - p\left(z\right)\right)},
\end{aligned}
$$

with

$$
\begin{aligned}
\Sigma_L^{-1} &= \mathbb{E}\left[R^L\left(X\right) R^L\left(X\right)' \mathcal{L}'\left(R^L\left(\bar{x}\right)' \pi_L\right)\right] \\
\tilde{\Sigma}_L^{-1} &= \frac{1}{n} \sum_{i=1}^{n} R^L\left(X_i\right) R^L\left(X_i\right)' \mathcal{L}\left(R^L\left(\bar{x}\right)' \tilde{\pi}_L\right),
\end{aligned}
$$

and $\tilde{\pi}_L$ between $\hat{\pi}_L$ and π_L.

By the same arguments as in the Addendum of **?**[4], we have that, uniformly in $(t, x) \in \mathcal{S}$,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{D_i \delta_i}{1 - G_1\left(Q_i-\right)} \left(\hat{\xi}_1\left(Q_i, X_i, D_{i:}; t, x\right) - \xi_1\left(Q_i, X_i, D_i; t, x\right)\right) \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} F_1\left(t|X_i\right) p\left(X_i\right) 1\left\{X_1 \leq x\right\}\left(D_i - p\left(X_i\right)\right) \\
&= \left[O_{\mathbb{P}}\left(L^{-\frac{s}{2k}+1}\right) + O_{\mathbb{P}}\left(L^2 n^{-\frac{1}{2}}\right)\right] + O_{\mathbb{P}}\left(\sqrt{n} L^{-\frac{s}{2k}+1}\right) \\
&\quad + O_{\mathbb{P}}\left(n^{-\frac{1}{2}} L^{\frac{11}{2}}\right) + O_{\mathbb{P}}\left(\max\left(L^{1-\frac{s}{2k}}, L^{-\frac{1}{2k}}\right)\right) \\
&= o_{\mathbb{P}}\left(1\right)
\end{aligned}
$$

4. The step-by-step procedure is available upon request.

under Assumptions 2.7 and 2.8. Therefore, by the above results we obtain that

$$\begin{aligned}\sup_{(t,x)\in\mathcal{S}}\Bigg|\frac{n_1}{n}\sum_{i=1}^{n_1}W_{1,i:n_1}\left(\hat{\xi}_1\left(Q_{i:n_1},X_{[i:n_1]},D_{[i:n_1]};t,x\right)-\xi_1\left(Q_{i:n_1},X_{[i:n_1]},D_{[i:n_1]};t,x\right)\right)\\ -\frac{1}{n}\sum_{i=1}^{n}F_1\left(t|X_i\right)1\left\{X_i\leq x\right\}p\left(X_i\right)\left(D_i-p\left(X_i\right)\right)\Bigg| \\ =o_{\mathbb{P}}\left(n^{-1/2}\right)\end{aligned} \tag{2.46}$$

as desired.

By applying the same arguments, we have that

$$\begin{aligned}\sup_{(t,x)\in\mathcal{S}}\Bigg|\frac{n_0}{n}\sum_{l=1}^{n_0}W_{0,l:n_0}\left(\hat{\xi}_0\left(Q_{l:n_0},X_{[l:n_0]},D_{[l:n_0]};t,x\right)-\xi_0\left(Q_{l:n_0},X_{[l:n_0]},D_{[l:n_0]};t,x\right)\right)\\ -\frac{1}{n}\sum_{i=1}^{n}F_0\left(t|X_i\right)\left(1-p\left(X_i\right)\right)1\left\{X_1\leq x\right\}\left(D_i-p\left(X_i\right)\right)\Bigg|=o_{\mathbb{P}}\left(n^{-1/2}\right).\end{aligned} \tag{2.47}$$

Combining (2.38), (2.46) and (2.47), we have that, uniformly in $(t,x)\in\mathcal{S}$,

$$\begin{aligned}&\hat{I}\left(t,x\right)-I\left(t,x\right)\\ &=\frac{1}{n}\sum_{i=1}^{n}\left[\left(\eta_i\left(t,x\right)-I\left(t,x\right)\right)-\alpha\left(X_i;t,x\right)\left(D_i-p\left(X_i\right)\right)\right]+o_P\left(n^{-1/2}\right)\end{aligned}$$

where

$$\alpha\left(X_i;t,x\right)=\left[F_1\left(t|X_i\right)p\left(X_i\right)+F_0\left(t|X_i\right)\left(1-p\left(X_i\right)\right)\right]1\left\{X_i\leq x\right\}$$

concluding our proof.■

Proof of Theorem 2.1: From the asymptotic representation of Lemma 2.1, it suffices to prove the convergence of the dominant term. To this end, define the class of real-valued measurable functions on $\chi_Y\times\chi_X\times\{0,1\}\times\{0,1\}$

$$\begin{aligned}\mathcal{F}=\{&\left(\bar{\omega},\bar{x},\bar{z},\bar{\delta}\right)\rightarrow\varphi_{(t,x)}\equiv\eta\left(\bar{\omega},\bar{x},\bar{z},\bar{\delta};t,x\right)\\ &-\left(F_1\left(t|\bar{x}\right)p\left(\bar{x}\right)+F_0\left(t|\bar{x}\right)\left(1-p\left(\bar{x}\right)\right)\right)1\left\{\bar{x}\leq x\right\}\left(\bar{z}-p\left(\bar{x}\right)\right):(t,x)\in\mathcal{S}\}\end{aligned}$$

where $\eta\left(\bar{\omega}, \bar{x}, \bar{z}, \bar{\delta}; t, x\right)$ is defined as in (2.15)

Our goal is to show that class of functions $\mathcal{F}$ is Donsker. By Theorem 2.10.6 in van der Vaart and Wellner (1996) it suffices to show that the classes of functions $\mathcal{G}_1$ and $\mathcal{G}_2$, as defined in (2.31) and (2.32), and for $j = \{0,1\}$, $\{\gamma_{j,0}(\cdot)\}$,$\{\delta\}$, $\{\gamma_{j,1}(\cdot)\}$, $\{\gamma_{j,2}(\cdot)\}$, $\{F_j(t,x|\cdot)\}$, $\{(D - p(\cdot))\}$ are Donsker.

For $j = \{0,1\}$, define the class of real-valued measurable functions on χ_X

$$\mathcal{G}_{3,j} \equiv \{\bar{x} \to \phi_2(\bar{x}) \equiv F_j(t|\bar{x}) : t \in \mathcal{S}\}. \tag{2.48}$$

Now, notice that both $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_{3,j}$ are VC-Class with square integrable envelope functions. Therefore, by Theorem 2.6.8 in van der Vaart and Wellner (1996), these classes of functions are Donsker. The functions $\gamma_{0,j}$, $(D - p(\cdot))$ and δ does not depend on t nor on x and so they are clearly Donsker.

We next consider $\gamma_{j,1}$. For $j = \{0,1\}$,define the classes of real-valued measurable functions

$$\mathcal{F}_{1,j} = \{\omega \in [-\infty, t_H) \to \gamma_{j,1}(\omega) \equiv \frac{1}{1 - H_j(\omega)} \int 1\{\omega < \bar{\omega}\} \xi_j(\bar{\omega}, \bar{x}, \bar{z}; t, x)$$
$$\gamma_{j,0}(\bar{w}) H_{j,11}(d\bar{\omega}, d\bar{x}) : (t,x) \in \mathcal{S}\}$$

$$\mathcal{F}_{2,j} = \{\omega \in [-\infty, t_H) \to \gamma_{j,2}(\omega) \equiv \int\int \frac{1\{\bar{v} < \omega, \bar{v} < \bar{\omega}\} \xi_j(\bar{\omega}, \bar{x}, \bar{z}; t, x)}{[1 - H_j(\bar{v})]^2}$$
$$\gamma_{j,0}(\bar{\omega}) H_{j,0}(d\bar{v}) H_{j,11}(d\bar{\omega}, d\bar{x}) : (t,x) \in \mathcal{S}\}$$

In order to prove that these classes of functions are Donsker, by Theorem 2.5.6 of van der Vaart and Wellner (1996), it suffices to show that, for $i = 1, 2$,

$$\int_0^\infty \sqrt{\ln N_{[\cdot]}(\varepsilon, \mathcal{F}_{i,j}, L_2(P))} d\varepsilon < \infty \tag{2.49}$$

where P is the probability measure corresponding to the joint distribution of (Q, δ, D, X), and $L_2(P)$ is the $L_2 - norm$. Notice that both $\mathcal{F}_{1,j}$ and $\mathcal{F}_{2,j}$ are classes of monotone bounded functions. Therefore, by Theorem 2.7.5 in van der Vaart and Wellner (1996), we have that, for a fixed $\varepsilon > 0$ and $i = 1, 2$, $\ln N_{[\cdot]}\left(\varepsilon, \mathcal{F}_{i,j}, L_2\left(P\right)\right) \leq K\varepsilon^{-1}$, where K is an arbitrary constant. Hence, for $i = 1, 2$, the integral in (2.49) is finite, and the classes of functions $\mathcal{F}_{1,j}$ and $\mathcal{F}_{2,j}$, $j = \{0, 1\}$, are Donsker.

We have just shown that $\mathcal{F}$ is Donsker, that is, we have proved that

$$\sqrt{n}\left(\hat{I} - I\right) \Rightarrow C_{\infty}$$

where C_{∞} is a tight Gaussian process in $l^{\infty}\left(\mathcal{S}\right)$ with zero mean and covariance function given by (2.20). Since under H_0, $I\left(t, x\right) = 0 \; \forall \left(t, x\right) \in \mathcal{W} \subseteq \mathcal{S}$, the proof is completed.■

Proof of Theorem 2.2: Notice that we can always write

$$\begin{aligned} \sqrt{n}\hat{I}\left(t, x\right) &= \sqrt{n}\left(\hat{I} - I\right)\left(t, x\right) + \sqrt{n}I\left(t, x\right) \\ &= D_{1,n}\left(t, x\right) + D_{2,n}\left(t, x\right). \end{aligned}$$

From the proof of Theorem 2.1, we have that

$$\sqrt{n}\left(\hat{I} - I\right) \Rightarrow C_{\infty},$$

and therefore $D_{1,n}\left(t, x\right) = O_{\mathbb{P}}\left(1\right)$. On the other hand, under the alternative $I\left(t, x\right) \neq 0$ for some $\left(t, x\right)$. Therefore $D_{2,n}\left(t, x\right) = O_{\mathbb{P}}\left(n^{1/2}\right)$. Hence, under H_1,

$$\sqrt{n} \sup_{(t,x)\in\mathcal{W}} \left|\hat{I}\left(t, x\right)\right| \rightarrow^{p} \infty,$$

Since under H_0, $I\left(t, x\right) = 0$ for all $\left(t, x\right)$, $KS_n = O_{\mathbb{P}}\left(1\right)$, and therefore $c_{\alpha}^{KS} = O\left(1\right)$ almost surely, we conclude that

$$\lim_{n\to\infty} P\left\{KS_n > c_{\alpha}^{KS}\right\} = 1.$$

Analogously, we have that

$$\lim_{n\to\infty} P\left\{CvM_n > c_\alpha^{CvM}\right\} = 1.$$

■

Proof of Theorem 2.3: As in the proof of Theorem 2.2, we can always write

$$\begin{aligned}\sqrt{n}\hat{I}(t,x) &= \sqrt{n}\left(\hat{I}-I\right)(t,x) + \sqrt{n}I(t,x)\\ &= D_{1,n}(t,x) + D_{2,n}(t,x)\end{aligned}$$

From the proof of Theorem 2.1, we have that

$$\sqrt{n}\left(\hat{I}-I\right) \Rightarrow C_\infty,$$

and therefore $D_{1,n}(t,x) = O_{\mathbb{P}}(1)$. On the other hand, under the local alternatives of the type $H_{1,n}$, $\sqrt{n}I(t,x) = \mathbb{E}\left[h(t,x)(p(X)(1-p(X)))1\{X \le x\}\right] = O_{\mathbb{P}}(1)$. Hence, under $H_{1,n}$,

$$\sqrt{n}\hat{I}(t,x) \Rightarrow C_\infty + R(t,x)$$

in $l^\infty(\mathcal{W})$.■

Before we proceed with the proof of Theorem 2.4, we prove the uniformly consistency of our estimator for

$$\begin{aligned}\alpha(X;t,x) &= \left(\mathbb{E}\left[\frac{D\delta 1\{Q\le t\}}{1-G_1(Q-)}\middle|X\right] - \mathbb{E}\left[\frac{(1-D)\delta 1\{Q\le t\}}{1-G_0(Q-)}\middle|X\right]\right)1\{X\le\cdot\}\\ &= \left[F_1(t|X)p(X) - F_0(t|X)(1-p(X))\right]1\{X\le\cdot\}\end{aligned}$$

To this end, if suffices to show that

$$\sup_{(t,\bar{x})\in\mathcal{S}} \left|\hat{\alpha}_1^{KM}(\bar{x};t) - F_1(t|\bar{x})p(\bar{x})\right| = o_{\mathbb{P}}(1),$$

where

$$\hat{\alpha}_1^{KM}(\bar{x};t) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{D_i\delta_i 1\{Q_i \le t\} R^L(X_i)}{1-\hat{G}_1^{KM}(Q_i-)}\right)'\left(\frac{1}{n}\sum_{i=1}^{n} R^L(X_i) R^L(X_i)'\right)^{-1} R^L(\bar{x})$$

The analogous result applies to the other expectation

Lemma 2.5 *Suppose Assumptions 2.1-2.8 hold. Additionally, assume that $F_1(\cdot,\cdot|X)$ is continuously differentiable of order $m \ge k$, where k is the dimension of X. Then,*

$$\sup_{(t,\bar{x})\in\mathcal{S}}\left|\hat{\alpha}_1^{KM}(\bar{x};t) - F_1(t|\bar{x})\,p(\bar{x})\right| = o_{\mathbb{P}}(1)$$

Proof For a matrix A, let $\|A\|$ denote the matrix norm of A such that $\|A\| = \sqrt{tr(A'A)}$. Define

$$\begin{aligned}
\Phi_L(t) &= \frac{1}{n}\sum_{i=1}^{n}\delta_i D_i 1\{Q_i \le t\}\,\gamma_{1,0}(Q_i)\,R^L(X_i),\\
\Phi_L^{KM}(t) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i D_i 1\{Q_i \le t\} R^L(X_i)}{1-\hat{G}_1^{KM}(Q_i-)},\\
\zeta_L &= \frac{1}{n}\sum_{i=1}^{n} R^L(X_i)\,R^L(X_i)'.
\end{aligned}$$

Notice that

$$\hat{\alpha}_1^{KM}(t|\bar{x}) = \Phi_L^{KM}(t)'\,\zeta_L^{-1} R^L(\bar{x})$$

From Theorem 1 of Stute (1993), we have that $\Phi_L^{KM}(t) = \Phi_L(t)$ *a.s.*. Given that the conditional variance of $\delta_i 1\{Q \le \cdot\}\,\gamma_0(Q_i)$ conditional on X is bounded, the uniform bound in Newey (1997) for power series estimators applies:

$$\sup_{(t,\bar{x})\in\mathcal{S}}\left|\Phi_L(t)'\,\zeta_L^{-1} R^L(\bar{x}) - F_1(t|\bar{x})\,p(\bar{x})\right| \le \mathcal{C}\left(L^{\frac{3}{2}} n^{-\frac{1}{2}} + L^{1-\frac{m}{k}}\right)$$

where m is the number of continuous derivatives of $F(\cdot|\bar{x})$.

Taking $L = a \cdot N^v$ as in Assumption 2.8, and from the results above, we have that

$$\sup_{(t,\bar{x})\in\mathcal{S}} \left|\hat{\alpha}_1^{KM}(\bar{x};t) - F_1(t|\bar{x})\,p(\bar{x})\right| = o_{\mathbb{P}}(1)$$

if $v < 1/3$, and $m \geq k$. Given that these conditions are fulfilled, we conclude our proof. ■

Next, we proceed with the proof of Theorem 2.4.

Proof of Theorem 2.4: For $j \in \{0,1\}$, denote

$$\hat{\eta}_{j,i}(t,x) = \hat{\xi}_j(Q_i, X_i, D_i; t, x)\,\hat{\gamma}_{j,0}(Q_{j,i})\,\delta_{j,i} + \hat{\gamma}_{j,1}(Q_{j,i})(1-\delta_{j,i}) - \hat{\gamma}_{j,2}(Q_{j,i})$$

and

$$\hat{\eta}_i(t,x) = \hat{\eta}_{1,i}(t,x) - \hat{\eta}_{0,i}(t,x)$$

and $\hat{\gamma}_{j,0}$, $\hat{\gamma}_{j,1}$ and $\hat{\gamma}_{j,2}$ are the empirical analogous of $\gamma_{j,0}$, $\gamma_{j,1}$ and $\gamma_{j,2}$, $j = \{0,1\}$, respectively, as defined in (2.15).

The proof follows two steps. In the first step in this proof is to show that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\hat{\eta}_i(t,x) - \hat{\alpha}^{KM}(X_i;t,x)(D_i - \hat{p}_n(X_i))\Big)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\eta_i(t,x) - \alpha(X_i;t,x)(D_i - p(X_i))) + o_{\mathbb{P}}(1) \qquad (2.50)
\end{aligned}$$

uniformly in $(t,x) \in \mathcal{S}$, that is, there is no estimation effect coming from replacing the true $\eta(t,x)$, $\alpha(X;t,x)$ and $p(X)$ with their nonparametric estimators.

In the second step, we prove that, under H_0, H_1 or $H_{1,n}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\eta_i(t,x,p) - \alpha(X_i;t,x)(D_i - p(X_i)))\,V_i \qquad (2.51)$$

converges weakly to the same limit process as in Theorem 2.1.

We proceed with the proof of the first step. For $j = \{0,1\}$, consider the class of measurable functions

$$\tilde{\mathcal{F}}_j = \{ \left(\bar{\omega}, \bar{x}, \bar{z}, \bar{\delta}\right) \to \xi_j\left(\bar{\omega}, \bar{x}, \bar{z}; t, x, p\right) \gamma_{j,0}\left(\bar{\omega}\right) \bar{\delta}$$
$$+ \gamma_{j,1}\left(\bar{\omega}\right)\left(1-\bar{\delta}\right) - \gamma_{j,2}\left(\bar{\omega}\right) : (t,x) \in \mathcal{S},\ p \in \mathcal{H}\},$$

where $\mathcal{H}$ is the collection of all distribution functions that satisfy Assumption 2.7.We prove that the $\tilde{\mathcal{F}}_j$ is Donsker. First, similar to Theorem 2.1, define the class of real-valued measurable functions on $\chi_Y \times \chi_X \times \{0,1\}$

$$\mathcal{F}_{0,3} \equiv \{ \left(\bar{\omega}, \bar{x}, \bar{z}\right) \to \xi_0\left(\bar{\omega}, \bar{x}, \bar{z}; t, x, p\right) \equiv p\left(\bar{x}\right) 1\left\{\bar{\omega} \le t\right\}$$
$$\times 1\left\{\bar{x} \le x\right\} : (t,x) \in \mathcal{S},\ p \in \mathcal{H}\},$$

$$\mathcal{F}_{1,3} \equiv \{ \left(\bar{\omega}, \bar{x}, \bar{z}\right) \to \xi_1\left(\bar{\omega}, \bar{x}, \bar{z}; t, x, p\right) \equiv \left(1 - p\left(\bar{x}\right)\right) 1\left\{\bar{\omega} \le t\right\}$$
$$\times 1\left\{\bar{x} \le x\right\} : (t,x) \in \mathcal{S},\ p \in \mathcal{H}\}.$$

Note that, for each $((t,x),p) \in \mathcal{S} \times \mathcal{H}$, we have that, for $j = \{0,1\}$,

$$E\left[\sup\left|\xi_j\left(\bar{\omega}, \bar{x}, \bar{z}; t, x, p_1\right) - \xi_j\left(\bar{\omega}, \bar{x}, \bar{z}; t, x, p\right)\right|^2\right] \le \mathcal{C}\delta^2,$$

where the supremum is over the set $(t_1, x_1) \in \mathcal{S}$ and $p_1 \in \mathcal{H}$ such that $|(t_1,x_1) - (t,x)| \le \delta$ and $\sup_{x \in \chi_X} |p_1(x) - p(x)| \le \delta$, respectively. By Lemma 2.2 and Theorem 19.5 in van der Vaart (1998), the classes of functions $\mathcal{F}_{0,3}$ and $\mathcal{F}_{1,3}$ are Donsker. Then, by Theorem 2.1 of Bae and Kim (2003), we have that $\tilde{\mathcal{F}}_1$ and $\tilde{\mathcal{F}}_2$ are Donsker. Therefore, by a stochastic equicontinuity argument and the Glivenko-Cantelli Theorem

$$\sup_{(t,x)\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\hat{\eta}_i(t,x,\hat{p}_n) - \eta_i(t,x,p)\right) \right| = o_{\mathbb{P}}(1). \tag{2.52}$$

Now, consider the class of functions

$$
\begin{aligned}
\mathcal{F}_{0,4} \equiv & \{(\bar{\omega},\bar{x},\bar{z}) \to \tilde{\alpha}_0(\bar{\omega},\bar{x},\bar{z};t,x,p,F_0) \equiv p(\bar{x})\, 1\{\bar{x} \le x\}\, F_0(t|\bar{x})(\bar{z}-p(\bar{x})) \\
& : (t,x) \in \mathcal{S},\ p \in \mathcal{H}_1,\ F_0 \in \mathcal{H}_2\},
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{F}_{1,4} \equiv & \{(\bar{\omega},\bar{x},\bar{z}) \to \tilde{\alpha}_1(\bar{\omega},\bar{x},\bar{z};t,x,p,F_1) \equiv (1-p(\bar{x}))\, 1\{\bar{x} \le x\}\, F_1(t|\bar{x})(\bar{z}-p(\bar{x})) \\
& : (t,x) \in \mathcal{S},\ \ p \in \mathcal{H}_1,\ F_q \in \mathcal{H}_2\}.
\end{aligned}
$$

Again, for each $((t,x),p,F_j) \in \mathcal{S} \times \mathcal{H}_1 \times \mathcal{H}_2$, we have that, for $j = \{0,1\}$,

$$
E\left[\sup |\tilde{\alpha}_j(\bar{\omega},\bar{x},\bar{z};t_1,x_1,p_1;F_{j,1}) - \tilde{\alpha}_j(\bar{\omega},\bar{x},\bar{z};t,x,p;F_j)|^2\right] \le \mathcal{C}\delta^2,
$$

where the supremum is over the set $(t_1,x_1) \in \mathcal{S}$, $p_1 \in \mathcal{H}_1$ and $F_j \in \mathcal{H}_2$ such that $|(t_1,x_1)-(t,x)| \le \delta$, $\sup_{x\in\chi_X} |p_1(x)-p(x)| \le \delta$ and $\sup_{(t,x)\in\mathcal{S}} |F_{j,1}(t|x) - F_j(t|x)| \le \delta$ respectively. By Lemma 2.2 and Theorem 19.5 in van der Vaart (1998), the classes of functions $\mathcal{F}_{0,4}$ and $\mathcal{F}_{1,4}$ are Donsker. Therefore, by a stochastic equicontinuity argument, the Glivenko-Cantelli Theorem and the triangle inequality, we have

$$
\sup_{(t,x)\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\hat{\alpha}^{KM}(X_i;t,x)(D_i - \hat{p}(X_i)) - \alpha(X_i;t,x)(D_i - p(X_i)) \right) \right| = o_{\mathbb{P}}(1). \tag{2.53}
$$

Combining (2.52) and (2.53), we have established (2.50), finishing the proof of the first step.

Next, let's consider (2.51). Define the classes of real measurable functions

$$
\begin{aligned}
\mathcal{G}_{0,1,*} \equiv & \left\{ \left(\bar{w},\bar{x},\bar{z},\bar{\delta},\bar{v}\right) \in \chi_Y \times \chi_{\boldsymbol{X}} \times \{0,1\} \times \{0,1\} \times \chi_v \to g_0\left(\bar{w},\bar{x},\bar{z},\bar{\delta},\bar{v};t,x\right) \equiv \right. \\
& (1\{\bar{w} \le t\}\, 1\{\bar{x} \le x\}\, p(\bar{x})\gamma_{0,0}(\bar{w})\,\bar{\delta} + \gamma_{0,1}(\bar{w})(1-\bar{\delta}) - \gamma_{0,2}(\bar{w}) \\
& + (1-p(\bar{x}))\, 1\{\bar{x} \le x\}\, F_0(y|X_i)(\bar{z}-p(\bar{x})))\bar{v}\ \ : (t,x) \in \mathcal{S}\},
\end{aligned}
$$

and

$$\mathcal{G}_{1,1,*} \equiv \{\left(\bar{w},\bar{x},\bar{z},\bar{\delta},\bar{v}\right) \in \chi_{Y} \times \chi_{\mathbf{X}} \times \{0,1\} \times \{0,1\} \times \chi_{v} \to g_1\left(\bar{w},\bar{x},\bar{z},\bar{\delta},\bar{v};t,x\right) \equiv$$
$$(1\{\bar{w} \leq t\} 1\{\bar{x} \leq x\}(\bar{z}-p(\bar{x}))\gamma_{1,0}(\bar{w})\bar{\delta} + \gamma_{1,1}(\bar{w})(1-\bar{\delta}) - \gamma_{1,2}(\bar{w})$$
$$- p(\bar{x}) 1\{\bar{x} \leq x\} F_1(y|\bar{x})(\bar{z}-p(\bar{x})))\bar{v} \ : (t,x) \in \mathcal{S}\}.$$

For $j = \{0,1\}$, the classes $\mathcal{G}_{j,1,*}$ are $P_{\left(\bar{w},\bar{x},\bar{z},\bar{\delta},\bar{v}\right)}$-Donsker, since $\mathcal{G}_{j,1}$ are $P_{\bar{w},\bar{x},\bar{z},\bar{\delta}}$-Donsker, see Theorem 2.9.6 in van der Vaart and Wellner (1996). Then, since $\mathbb{P}_n^* g_j = 0$ for all $g_j \in \mathcal{G}_{j,1,*}$,

$$I^*(t,x) = \frac{1}{n}\sum_{i=1}^{n}\left(\eta_i(t,x) - \alpha(X_i;t,x)(D_i - p(X_i))\right)V_i + o_{\mathbb{P}_n^*}\left(n^{-1/2}\right), \tag{2.54}$$

uniformly in $(t,x) \in \mathcal{S}$.

The expansion (2.54), and the multiplier functional central limit theorem, see Theorem 2.9.6 in van der Vaart and Wellner (1996), imply that $\sqrt{n}I^*(t,x)$ converges weakly (almost surely) to the same weak limit as $\sqrt{n}\hat{I}(t,x)$ in $l^{\infty}(\mathcal{S})$ under H_0, H_1 or H_{1n}.

This completes the proof of Theorem 2.4.■

Proof of Theorem 2.5 First, we must derive the asymptotic linear representation of the process $\left(\hat{I}_{\tau}^{CATE} - I_{\tau}^{CATE}\right)(x)$. To consider the most general case, we set $\tau = \tau_C$. Then, we can rewrite $\hat{I}_{\tau}^{CATE}(\cdot)$ as

$$\begin{aligned}
\hat{I}_{\tau}^{CATE} &= \bar{I}_{\tau}^{CATE}(x) \\
&+ \frac{n_1}{n}\sum_{i=1}^{n_1} W_{1,i:n_1}\left(\hat{\xi}_1^{CATE}\left(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1[i:n_1]}; t, x\right) - \xi_1^{CATE}\left(Q_{1,i:n_1}, X_{1,[i:n_1]}, D_{1[i:n_1]}; x\right)\right) \\
&- \frac{n_0}{n}\sum_{l=1}^{n_0} W_{0,l:n_0}\left(\hat{\xi}_0^{CATE}\left(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; t, x\right) - \xi_0^{CATE}\left(Q_{0,l:n_0}, X_{0,[l:n_0]}, D_{0,[l:n_0]}; x\right)\right),
\end{aligned} \tag{2.55}$$

where $\bar{I}_{\tau}^{CATE}(x)$ is defined similarly to (2.30) but replacing ξ_1 and ξ_1 with

$$\xi_1^{CATE}(\bar{y},\bar{x},\bar{z};x) = \bar{z}(1-p(\bar{x}))\bar{y}1\{\bar{x}\leq x\},$$
$$\xi_0^{CATE}(\bar{y},\bar{x},\bar{z};x) = (1-\bar{z})p(\bar{x})\bar{y}1\{\bar{x}\leq x\}.$$

Additionally, $\hat{\xi}_1^{CATE}$ and $\hat{\xi}_0^{CATE}$ are defined similarly to ξ_1^{CATE} and ξ_0^{CATE}, but replacing the true propensity score $p(\cdot)$ by the SLE $\hat{p}(\cdot)$.

We will derive the uniform representation of each term separately, as in Theorem 2.1. To this end, define the classes of real-value measurable functions on $\chi_Y \times \chi_X \times \{0,1\}$

$$\mathcal{H}_{0,1} \equiv \{(\bar{\omega},\bar{x},\bar{z}) \to \xi_0^{CATE}(\bar{y},\bar{x},\bar{z};x) \equiv p(\bar{x})\bar{\omega}1\{\bar{x}\leq x\} : x \in \mathbb{R}^k\},$$

$$\mathcal{H}_{1,1} \equiv \{(\bar{\omega},\bar{x},\bar{z}) \to \xi_1^{CATE}(\bar{y},\bar{x},\bar{z};x) \equiv (1-p(\bar{x}))\bar{\omega}\times 1\{\bar{x}\leq x\} : x \in \mathbb{R}^k\}$$

Notice that $\mathcal{H}_{j,1}$ are a VC-subgraph classes of functions with VC index smaller or equal than $k+2$ and admits the envelope $\Phi(\bar{\omega},\bar{x},\bar{z}) = |\bar{\omega}|$ that satisfies, under Assumption 2.10, the required moment conditions of Theorem 1 of Sellero et al. (2005). Thus,

$$\bar{I}_{\tau}^{CATE}(x) = \frac{1}{n}\sum_{i=1}^{n}\eta_i^{CATE}(x) + R_n^{CATE}(x) \tag{2.56}$$

where

$$\eta_i^{CATE}(x) = \eta_{1,i}^{CATE}(x) - \eta_{0,i}^{CATE}(x),$$

and for $j=\{0,1\}$,

$$\eta_{j,i}^{CATE}(x) = \xi_j^{CATE}(Q_i,X_i,D_i;x)\gamma_{j,0}(Q_{j,i})\delta_{j,i} + \gamma_{j,1}^{CATE}(Q_{j,i})(1-\delta_{j,i}) - \gamma_{j,2}^{CATE}(Q_{j,i}),$$

$$\gamma_{j,0}(\bar{t}) = \exp\left\{\int_0^{\bar{t}-} \frac{H_{j,0}(d\bar{w})}{1-H_j(\bar{w})}\right\},$$
$$\gamma_{j,1}(\bar{t}) = \frac{1}{1-H_j(t)}\int 1\{\bar{t}<\bar{w}\}\,\xi_j^{CATE}(\bar{w},\bar{x},\bar{z};x)\,\gamma_{j,0}(\bar{w})\,H_{j,11}(d\bar{w},d\bar{x}),$$
$$\gamma_{j,2}(\bar{t}) = \int\int \frac{1\{\bar{v}<\bar{t},\bar{v}<\bar{w}\}\,\xi_j^{CATE}(\bar{w},\bar{x},\bar{z};x)}{[1-H_j(\bar{v})]^2}\gamma_{j,0}(\bar{w})\,H_{j,0}(d\bar{v})\,H_{j,11}(d\bar{w},d\bar{x}).$$

and

$$\sup_{x\in\mathbb{R}^k}\left|R_n^{CATE}(x)\right| = O\left(\frac{\ln^3 n}{n}\right) \text{ a.s.}$$

Now, we look for the second term of (2.55). Using similar arguments as in the proof of Lemma 2.1, we can establish that

$$\begin{aligned}
&\frac{n_1}{n}\sum_{i=1}^{n_1} W_{1,i:n_1}\left(\hat{\xi}_1^{CATE}\left(Q_{1,i:n_1},X_{1,[i:n_1]},D_{1[i:n_1]};t,x\right)-\xi_1^{CATE}\left(Q_{1,i:n_1},X_{1,[i:n_1]},D_{1[i:n_1]};x\right)\right)\\
&-\frac{n_0}{n}\sum_{l=1}^{n_0} W_{0,l:n_0}\left(\hat{\xi}_0^{CATE}\left(Q_{0,l:n_0},X_{0,[l:n_0]},D_{0,[l:n_0]};t,x\right)-\xi_0^{CATE}\left(Q_{0,l:n_0},X_{0,[l:n_0]},D_{0,[l:n_0]};x\right)\right)\\
&=\frac{1}{n}\sum_{i=1}^{n}\alpha^{CATE}(X_i;x)\,(D_i-p(X_i))+o_{\mathbb{P}}\left(n^{-1/2}\right) \qquad (2.57)
\end{aligned}$$

uniformly in $x\in\mathcal{W}_X$, where

$$\alpha^{CATE}(\bar{x};x) = \alpha_1^{CATE}(\bar{x};x)-\alpha_0^{CATE}(\bar{x};x)$$

and

$$\alpha_1^{CATE}(\bar{x};x) = -p(\bar{x})\,1\{\bar{x}\le x\}\,\mathbb{E}(Y_1|\bar{x}),$$
$$\alpha_0^{CATE}(\bar{x};x) = (1-p(\bar{x}))\,1\{\bar{x}\le x\}\,\mathbb{E}(Y_0|\bar{x}).$$

Combining (2.56) and (2.57), we conclude that

$$\hat{I}_{\tau}^{CATE} - I_{\tau}^{CATE} = \frac{1}{n}\sum_{i=1}^{n}\Bigg[\left(\eta_i^{CATE}(x) - I_{\tau}^{CATE}(x)\right) - \alpha^{CATE}(X_i;x)\left(D_i - p\left(X_i\right)\right)\Bigg] + o_P\left(n^{-1/2}\right) \tag{2.58}$$

uniformly in $x \in \mathcal{W}_X$, concluding the proof of the asymptotic linear representation.

Once we have proved the validity of the uniform linear representation (2.58), the proof of the weak converge of the process $\sqrt{n}\left(\bar{I}_n^{\tau} - I^{\tau}\right)(x)$ under H_0^{CATE}, H_1^{CATE} and $H_{1,n}^{CATE}$ follows the same steps of Theorems 2.1, 2.2 and 2.3, and the validity of the bootstrap follows the reasoning of Theorem 2.4 in a routine fashion. Details are omitted.■

Chapter 3

A simple GMM for randomly censored data (with Miguel A. Delgado)

3.1 Introduction

Endogeneity and right-censoring are common problems in many areas of applied economics. For example, right censored outcomes naturally appear when one is interest in analyze if labor market programs affect the length of unemployment, if correctional programs affect recidivism of criminal activities, or whether the survival time is affected by a new clinical therapy. Endogeneity is also a widespread phenomenon both in experimental studies due to noncompliance, and in observational studies due to simultaneity, measurement error, sample selection or omitted relevant variables. In this paper, we propose a simple yet powerful generalized method of moments (GMM) estimator that deals with both problems, and name this estimator the Kaplan-Meier GMM (KM-GMM) estimator.

The semiparametric literature which studies censored data can be classified in two branches. The first one, assumes that the censoring values for the dependent variable are known for all observations, even those not censored. A number of consistent estimator has been proposed exploiting different identification assumptions: in the absence of endogeneity, see e.g. Powell (1984, 1986), Horowitz (1986), Newey and Powell (1990), Buchinsky and Hahn (1998), Chernozhukov and Hong (2002), among others, whereas Lewbel (2000), Hong and Tamer (2003), Blundell and Powell (2007) and Chernozhukov et al. (2015) proposed methods suitable for endogenous (continuous) regressors, and Frandsen (2014) for endogenous binary treatments (without covariates). We refer to such models as *fixed* censoring models.

The second branch of literature has been concerned with *randomly* censored data. In this class of models, the outcome of interest, typically a duration, may be right-censored at random points which are observed only when the observation is censored. This class of models has seen many applications in economics, see e.g. van den Berg (2001) for a survey. Studies that proposes semiparametric methods for randomly right-censored data under exogeneity conditions includes Cox (1972), Miller (1976), Buckley and James (1979), Koul et al. (1981), Ritov (1991), Ying et al. (1995), Stute (1996*a*, 1999), Yang (1999), Honore et al. (2002), Portnoy (2003), Cosslett (2004), Wang and Wang (2009), among others. When one regressor is endogenous, Bijwaard and Ridder (2005*b*) propose a two-stage rank estimator to correct for noncompliance in randomized experiments, whereas Khan and Tamer (2009) proposed a quantile instrumental variable estimator based on conditional moment inequalities.

In this paper we describe a general methodology to estimate finite-dimensional parameters β_0 with *randomly* censored data. Different than the aforementioned proposals, our procedure does not focus on a particular model. Instead, the parameters of interest are characterized by a vector of orthogonality conditions, which can be nonlinear or non-smooth in β_0. Our procedure can be seen as an extension of Hansen (1982) and Pakes and Pollard (1989) GMM estimators to *randomly* censored data. By doing so, our proposal naturally extends many models widely used for complete data to the random censoring setup. Examples include the instrumental variables (IV), linear and nonlinear (two stage) least squares, and the linear and nonlinear IV quantile regression models.

In order to tackle the censoring problem, we characterize our estimators as Kaplan-Meier integrals, as in Stute and Wang (1993*b*) and Stute (1993, 1995, 1996*a*). One of the attractiveness of this approach is that our procedure is fully data-driven, and does not rely on choice of tuning parameters such as bandwidths. Moreover, in the absence of censoring, our estimator reduces to the standard GMM. We show under sufficient condition that the KM-GMM estimator is consistent and we derive its large sample distribution.

The remainder of the paper is organized as follows. In Section 2, we present the KM-GMM model. In Section 3 and 4, we derive the general asymptotic distribution for our estimators, and illustrate how one can verify the conditions for the general limit theorems, focusing our attention to the linear GMM model. In Section 5, we study the finite-sample properties of our proposal by means of a Monte Carlo exercise. Finally, we offer concluding remarks and suggest extensions for future research. Mathematical proofs are gathered in an appendix at the end of the article.

3.2 The Kaplan-Meier GMM

In this paper, we are concerned with inference on the parameter vector $\beta_0 \in \Theta \subset \mathbb{R}^d$ defined by the set of orthogonality conditions of the form

$$\mathbb{E}\left[g\left(Y, X, Z; \beta_0\right)\right] = 0, \tag{3.1}$$

where Y is the outcome of interest, and X is a d-dimensional vector of potentially endogenous covariates, and Z a k-dimensional vector of instruments. $g(\cdot)$ is a k-dimensional vector of generalized residuals with functional forms known up to β_0. We allow $g(\cdot)$ to be nonlinear or non-smooth in β_0.

Given that the outcome variable is typically a duration, Y is usually is subjected to random right-censoring. Therefore, rather than Y, one observes $Q = \min(Y, C)$, together with an indicator $\delta = 1\{Y \leq C\}$, where C is a censoring random variable. Censoring might appear for different reasons such as the end of a follow-up or drop out. Henceforth, we assume that $\{(Q_i, \delta_i, X_i, Z_i)\}_{i=1}^n$ is a random sample, and all random variables are defined on a common probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

To fix ideas, it is worth to consider the case in which (Y, X, Z) is fully observed. In such cases, one can consistently estimate β_0 using the GMM, that is

$$\hat{\beta}_{GMM} = \arg\min_{\beta \in \Theta} \left\|\mathbb{E}_n\left[g\left(\beta\right)\right]\right\|_{A_n} \tag{3.2}$$

where

$$\left\|\mathbb{E}_n\left[g\left(\beta\right)\right]\right\|_{A_n} = \left[\frac{1}{n}\sum_{i=1}^{n} g\left(Y_i, X_i, Z_i; \beta\right)\right]' A_n \left[\frac{1}{n}\sum_{i=1}^{n} g\left(Y_i, X_i, Z_i; \beta\right)\right]$$

A_n is an $k \times k$, possibly random positive definite weight matrix, whose rank is greater than or equal to d, and $A_n = A + o_p(1)$, see e.g. Hansen (1982), Pakes and Pollard (1989) and Newey and McFadden (1994, Section 7).

Notice that the key to estimate β_0 is to define a quadratic empirical distance that uses only information in the data. When $\{Y_i, X_i, Z_i\}_{i=1}^n$ is fully observed, such distance is given by $\|\mathbb{E}_n [g(\beta)]\|_{A_n}$, which is defined as the sample analog of

$$\left[\int g(y,x,z;\beta) F(dy,dx,dz)\right]' A \left[\int g(y,x,z;\beta) F(dy,dx,dz)\right]. \qquad (3.3)$$

That is, in order to get $\|\mathbb{E}_n [g(\beta)]\|_{A_n}$, one replace the true unknown distribution $F(y,x,z)$ in (3.3), by its empirical analog

$$\hat{F}_n(y,x,z) = n^{-1}\sum_{i=1}^{n} 1\{Y_i \leq y, X_i \leq x, Z_i \leq z\}.$$

Unfortunately, when the outcome of interest Y is subjected to right-censoring, the aforementioned GMM procedure cannot be applied because the empirical distribution $\hat{F}_n(\cdot,\cdot,\cdot)$ is not at our disposal. Nonetheless, one can exploit other possibilities. Since the Kaplan and Meier (1958) estimator is the analogous to the empirical CDF when the outcome is subjected to right censoring, a convenient way to proceed involves using some multivariate Kaplan-Meier (KM) estimator of $F(\cdot,\cdot,\cdot)$, which would use only the information available at the sample $\{(Q_i, \delta_i, X_i, Z_i)\}_{i=1}^n$.

To define the KM estimator of $F(y,x,z)$, let, $Q_{1:n} \leq \cdots \leq Q_{n:n}$ be the ordered Q values, where ties within Y or within C are ordered arbitrarily and ties among Y and C are treated as if the former precedes the later, and let $X_{[i:n]}$ and $Z_{[i:n]}$ be the concomitant of the ith order statistics, i.e. the X and Z paired with $Q_{i:n}$. Stute (1993, 1996a) defines the multivariate Kaplan-Meier estimator of $F(y,x,z)$ as

$$\hat{F}_n^{KM}(y,x,z) = \sum_{i=1}^{n} W_{i:n} 1\{Q_{i:n} \leq t\} 1\left\{X_{[i:n]} \leq x\right\} 1\left\{Z_{[i:n]} \leq z\right\},$$

where

$$W_{k:n} = \frac{\delta_{[k:n]}}{n-k+1}\prod_{l=1}^{k-1}\left(\frac{n-l}{n-l+1}\right)^{\delta_{[l:n]}}$$

denotes its "jump" at observation k.

With the KM estimators $\hat{F}_n^{KM}(\cdot,\cdot,\cdot)$ at hands, the KM analogue of (3.3) is given by

$$\begin{aligned}
&\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]\right\|_{A_n^{KM}} \\
&= \left(\int g\left(y,x,z;\beta\right)\hat{F}_n^{KM}\left(dy,dx,dz\right)\right)' A_n^{KM}\left(\int g\left(y,x,z;\beta\right)\hat{F}_n^{KM}\left(dy,dx,dz\right)\right) \\
&= \left[\sum_{i=1}^{n} W_{i:n} g\left(Q_{i:n},X_{[i:n]},Z_{[i:n]};\beta\right)\right]' A_n^{KM}\left[\sum_{i=1}^{n} W_{i:n} g\left(Q_{i:n},X_{[i:n]},Z_{[i:n]};\beta\right)\right],
\end{aligned}$$

where A_n^{KM} is defined analogously to A_n, but potentially replacing the empirical integral by the suitable Kaplan-Meier empirical measure.

Therefore, given a random sample $\{(Q_i,\delta_i,X_i,Z_i)\}_{i=1}^n$, we form the estimator of β_0 as

$$\hat{\beta}_{GMM}^{KM} = \arg\min_{\beta\in\Theta}\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]\right\|_{A_n^{KM}}, \tag{3.4}$$

and name this estimator the Kaplan-Meier GMM (KM-GMM) estimator. It is important to notice that, in the absence of censoring, for $i=1,\ldots,n$, $Q_i=Y_i$, $\delta_i=1$, $W_{i:n}=n^{-1}$ *a.s.*, and therefore, (3.4) naturally reduces to (3.1). Hence, one can see that the KM-GMM is a natural extension of the GMM to handle right-censored data.

3.3 The large sample theory

Our large sample theory is a direct extension of the theory of Pakes and Pollard (1989) to allow for random censoring. To do so, we will also use the results on Kaplan-Meier integrals from Stute and Wang (1993*b*) and Stute (1993, 1995, 1996*a*). In the absence of censoring, our Theorem 1 becomes Pakes and Pollard (1989)'s Corollary 3.2 and our Theorem 2 becomes their Theorem 3.3.

3.3.1 Consistency

Before we proceed with our analysis, let us discuss some assumptions regarding the censoring mechanism.

Assumption 3.1 *(i) Y and C are independent.*

(ii) $P(Y \leq C|X, Z, Y) = P(Y \leq C|Y)$.

Assumption 3.1 is needed for identifiability. As pointed out in Stute (1996*b*), almost surely and distributional convergence of Kaplan-Meier integrals carries through if only the $\{(Q_i, \delta_i)\}_{i=1}^n$ are *iid*. Independence of Y and C is needed to the identify the limit. Assumption 3.1 (ii) states that, given the "time of death" Y, the covariates do not provide any further information whether censoring will take place, that is, δ and X are conditionally independent given Y. A particular case in which it holds is when C is independent of (Y, X), as assumed in Honore et al. (2002), Lee and Lee (2005) and Frandsen (2014), for example. Nonetheless, Assumption 3.1 is more general and does not put obvious restrictions on the joint distribution of (X, Z, C). We notice that similar assumptions have been used in different contexts, see e.g. Chen (2001), Tang et al. (2003), D'Haultfoeuille (2010) and Breunig et al. (2014). An alternative to Assumption 3.1 is $Y \perp\!\!\!\perp C|X$. In this case the use of smoothing techniques and trimming procedures are required, see Akritas (1994), González-Manteiga and Cadarso-Suárez (1994), and Iglesias Pérez and González-Manteiga (1999) for examples in different contexts. With Assumption 3.1, the use of smothers and trimming is avoided.

Let $F(\cdot) = \mathbb{P}(Y_j \leq \cdot)$, $G(\cdot) = \mathbb{P}(C \leq \cdot)$ and $H(\cdot) = \mathbb{P}(Q \leq \cdot)$. By Assumption 3.1, $1 - H = (1 - F)(1 - G)$. Let $\tau_H = \inf\{t : H(t) = 1\}$ be the least upper bound for the support of H. Similarly for F and G. Then, $\tau_H = \min(\tau_F, \tau_G)$. Clearly, there will be no data beyond τ_H. So, if (3.3) is a parameter of interest, we can only consistently estimate it up to $y \leq \tau_H$. More precisely, from Stute (1993), we have that, with probability one

$$\begin{aligned} \mathbb{E}_\infty^{KM}[g(\beta)] \equiv \lim_{n\to\infty} \mathbb{E}_n^{KM}[g(\beta)] = &\left[\int_{\{Y<\tau_H\}} g(Y, X, Z; \beta)\, d\mathbb{P}\right] \\ &+ 1\{\tau_H \in B\}\left[\int_{\{Y=\tau_H\}} g(\tau_H, X, Z; \beta)\, d\mathbb{P}\right], \end{aligned}$$

where B is a set of H atoms, possibly empty. Notice that when $\tau_F < \tau_G$, $\mathbb{E}_\infty^{KM}[g(\beta_0)] = \mathbb{E}[g(\beta)]$. If $\tau_F > \tau_G$, $\mathbb{E}_\infty^{KM}[g(\beta)] = \mathbb{E}[g(\beta)]$ if $g = 0$ on $[\tau_G, \tau_H]$, and when $\tau_F = \tau_G$, $\mathbb{E}_\infty^{KM}[g(\beta_0)] = \mathbb{E}[g(\beta)]$ provided the cases where $F\{\tau_H\} > 0$ but $1 - G(\tau_H-) = 0$ are excluded. To achieve the maximum generality possible and avoid discussion related to this support restriction, we will assume that $\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\beta_0\right)\right]\right\|_A = 0$.

Now, we are in a position to state our first theorem. For consistency, the weight matrix A_n^{KM} plays no important role, and therefore, we focus our attention to the case where $A_n^{KM} = A$. A crucial example that this is satisfied is when one sets $A_n^{KM} = I_k$.

Theorem 3.1 *Under Assumption 3.1, suppose that* $\beta_0 \in \Theta$ *satisfies* $\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\beta_0\right)\right]\right\| = 0$ *and that:*

(i) $\left\|\mathbb{E}_n^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|_A \leq \inf_{\beta\in\Theta}\left\|\mathbb{E}_n^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|_A + o_p(1)$.

(ii) For all $\varepsilon > 0$, *there exist* $\eta(\varepsilon) > 0$ *such that* $\inf_{\|\beta-\beta_0\|>\varepsilon}\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\beta\right)\right]\right\|_A \geq \eta(\varepsilon) > 0$.

(iii)

$$\sup_{\beta\in\Theta}\frac{\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right] - \mathbb{E}_{\infty}^{KM}\left[g\left(\beta\right)\right]\right\|_A}{1 + \left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]\right\|_A + \left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\beta\right)\right]\right\|_A} = o_p(1).$$

Then, $\hat{\beta}_{GMM}^{KM} - \beta_0 = o_p(1)$.

Remark 3.1 A sufficient condition for *(iii)* is

$$\sup_{\beta\in\Theta}\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right] - \mathbb{E}_{\infty}^{KM}\left[g\left(\beta\right)\right]\right\|_A = o_p(1).$$

3.3.2 Asymptotic Normality

In this sub-section, we assume that $\hat{\beta}_{GMM}^{KM}$ is a consistent estimator for β_0.Then, establishing asymptotically normality only requires local assumptions on the behavior of $\mathbb{E}_n^{KM}\left[g\left(\cdot\right)\right]$ and $\mathbb{E}_{\infty}^{KM}\left[g\left(\cdot\right)\right]$ in small neighborhoods of β_0. Therefore, the parameter space Θ can be replace by small or even shrinking set. Define $\Theta_\delta = \{\beta \in \Theta : \|\beta - \beta_0\| \leq \delta\}$ for some small $\delta > 0$.

Next, let us introduce some additional notation. Define the sub-distributions $\tilde{H}_{11}\left(y\right) = P\left(Q \leq y, X \leq x, Z \leq z, \delta = 1\right)$ and $\tilde{H}_0\left(y\right) = P\left(Q \leq y, \delta = 0\right)$. Let

$$\gamma_0\left(y\right) = \exp\left(\int_{-\infty}^{y-}\frac{\tilde{H}_0\left(dw\right)}{\left[1 - H\left(w\right)\right]}\right),$$

$$\gamma_1^g\left(y\right) = \frac{1}{1 - H\left(y\right)}\int 1\left\{y < w\right\}g\left(w, x, z; \beta_0\right)\gamma_0\left(w\right)\tilde{H}_{11}\left(dw, dx, dz\right),$$

and

$$\gamma_2^g(y) = \int\int \frac{1\{v < y, v < w\} g(w, x, z; \beta_0) \gamma_0(w)}{(1 - H(v))^2} \tilde{H}_0(dv) \tilde{H}_{11}(dw, dx, dz).$$

Note that, for continuous H, $\gamma_0 = (1 - G)^{-1}$, see Stute and Wang (1993b).

Stute (1996a) show that, under weak moment assumptions,

$$\begin{aligned}
&\int g(\beta_0) d\left(F_n^{KM} - F\right) \\
&= \frac{1}{n} \sum_{i=1}^{N} \left[g(Q_i, X_i, Z_i; \beta_0) \gamma_0(Q_i) \delta_i + \gamma_1^g(Q_i)(1 - \delta_i) - \gamma_2^g(Q_i)\right] + o_p\left(n^{-1/2}\right) \\
&= \frac{1}{n} \sum_{i=1}^{N} \left[\psi_i^g(\beta_0)\right] + o_p\left(n^{-1/2}\right),
\end{aligned}$$

where the $\psi's$ are *iid* with expectation zero. To proceed with our analysis, set

$$V = \mathbb{E}\left[\psi^g(\beta_0) \psi^g(\beta_0)'\right], \tag{3.5}$$

and let

$$T(w) = \int_0^{w-} \frac{G(dy)}{[1 - H(y)][1 - G(y)]}$$

Next, we introduce the following integrability assumptions.

Assumption 3.2 $\int [g(Q, X, Z; \beta_0) \gamma_0(Q) \delta]^2 d\mathbb{P} < \infty$

Assumption 3.3 $\int |g(Q, X, Z; \beta_0)| T^{1/2}(Q) d\mathbb{P} < \infty$

Assumption 3.2 guarantees that $V < \infty$ and Assumption 3.3 is mainly to control the bias of Kaplan-Meier integrals. For extensive discussion of these two assumptions, see Stute (1994).

Next theorem gives conditions under which $\hat{\beta}_{GMM}^{KM}$, which is now assumed to converge in probability to β_0, satisfies a central limit theorem.

Theorem 3.2 *Let $\hat{\beta}_{GMM}^{KM}$ be a consistent estimator of β_0, the unique point of Θ for which $\left\|\mathbb{E}_\infty^{KM}[g(\beta_0)]\right\|_A = 0$. If Assumptions 3.1-3.3 are satisfied and that:*

(i) $\left\|\mathbb{E}_n^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|_A \leq \inf_{\beta\in\Theta_\delta}\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]\right\|_A + o_p\left(n^{-1/2}\right).$

(ii) $\mathbb{E}_\infty^{KM}\left[g\left(\beta\right)\right]$ *is differentiable at* β_0 *with derivative matrix* Γ *of full rank.*

(iii) For every sequence $\{\delta_n\}$ *of positive numbers that converges to zero,*

$$\sup_{\|\beta-\beta_0\|\leq\delta_n}\frac{\sqrt{n}\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]-\mathbb{E}_\infty^{KM}\left[g\left(\beta\right)\right]-\mathbb{E}_n^{KM}\left[g\left(\beta_0\right)\right]\right\|_A}{1+\sqrt{n}\left(\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]\right\|_A+\left\|\mathbb{E}_\infty^{KM}\left[g\left(\beta\right)\right]\right\|_A\right)} = o_p\left(1\right).$$

(iv) $\beta_0 \in int\left(\Theta\right).$

(v) Let $A_n^{KM} = A_n^{KM}\left(\hat{\beta}_{GMM}^{KM}\right)$, *where* $\left\{A_n^{KM}\left(\beta\right):\beta\in\Theta\right\}$ *is a family of sequences of random matrices such that*

$$\sup_{\|\beta-\beta_0\|\leq\delta_n}\left\|A_n^{KM}\left(\beta\right)-A\right\|_{I_k} = o_p\left(1\right)$$

for all positive values $\delta_n = o\left(1\right)$*.*

Then,

$$\sqrt{n}\left(\hat{\beta}_{GMM}^{KM}-\beta_0\right)\xrightarrow{d}N\left(0,\Omega\right),$$

where

$$\Omega = \left(\Gamma' A\Gamma\right)^{-1}\Gamma' AVA\Gamma\left(\Gamma' A\Gamma\right)^{-1},$$

and V *is defined in (3.5).*

Remark 3.2 A sufficient condition for *(iii)* is

$$\sup_{\beta\in\Theta}\sqrt{n}\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]-\mathbb{E}_\infty^{KM}\left[g\left(\beta\right)\right]-\mathbb{E}_n^{KM}\left[g\left(\beta_0\right)\right]\right\|_A = o_p\left(1\right).$$

3.4 Kaplan-Meier Linear GMM

This section provides a detailed examination of the asymptotic behavior of the Kaplan-Meier linear GMM, and we illustrate how the general conditions in Theorems 3.1 and 3.2 can be verified. Moreover, we discuss how one can conduct valid inference in this class of models.

Economists often use linear regression models to quantify a relationship between economic variables. In many cases, however, some covariates included in the model may be endogenous. More formally, one is interest in the relationship

$$Y = X^{'}\beta_0 + \varepsilon \tag{3.6}$$

where Y is a scalar outcome of interest, X is a d-dimensional vector of potentially endogenous covariates, and Z a k-dimensional vector of instrumental variables that a priori can be assumed to be uncorrelated with ε. Hence, the parameters $\beta_0 \in \Theta \subset \mathbb{R}^d$ can be regarded as the solution to the moment equations of the form

$$\mathbb{E}\left[Z\left(Y - X'\beta_0\right)\right] = 0.$$

In the absence of censoring, one could estimate β_0 by means of the GMM (3.2). When Y is subjected to random right-censoring, one can estimate β_0 using the Kaplan-Meier linear GMM, that is,

$$\hat{\beta}^{KM} = \arg\min_{\beta\in\Theta}\left[\sum_{i=1}^{n} W_{i:n} Z_{[i:n]}\left(Q_{i:n} - X'_{[i:n]}\beta\right)\right]' A_n^{KM}\left[\sum_{i=1}^{n} W_{i:n} Z_{[i:n]}\left(Q_{i:n} - X'_{[i:n]}\beta\right)\right], \tag{3.7}$$

where A_n^{KM} is an $k \times k$ positive definite weight matrix, such that $A_n^{KM} = A + o_p(1)$. In this setup, we have $g(\beta) = Z(Y - X'\beta)$. Moreover, as standard in the complete data case, see e.g. Newey and McFadden (1994), we assume that Θ is compact and that $\mathbb{E}\left[\sup_{\beta\in\Theta}\|Z(Y - X'\beta)\|_A\right] < \infty$.

Now, we verify the general conditions of Theorem 3.1. First, notice the function $g(\beta)$ is linear in β. To verify Assumptions 1.i and 1.ii, the first order conditions of (3.7) leads to the moment equation

$$\left(\left[\sum_{i=1}^{N} W_{i:n} Z_{[i:n]} X'_{[i:n]}\right]' A_n^{KM}\left[\sum_{i=1}^{N} W_{i:n} Z_{[i:n]}\left(Q_{i:n} - X'_{[i:n]}\hat{\beta}^{KM}_{2SLS}\right)\right]\right) = 0.$$

As it is well known, a necessary and sufficient condition for identification of β_0 is that

$\mathbb{E}\left[ZX'\right]$ exist and has full column rank. Notice that this implies that we must have $k \geq d$, that is, at least as many instruments as parameters. With this conditions, we have that

$$\hat{\beta}^{KM} = \left(\left[\sum_{i=1}^{N} W_{i:n} Z_{[i:n]} X_{[i:n]}^{'} \right]^{'} A_n^{KM} \left[\sum_{i=1}^{N} W_{i:n} Z_{[i:n]} X_{[i:n]}^{\prime} \right] \right)^{-1} \left[\sum_{i=1}^{N} W_{i:n} Z_{[i:n]} X_{[i:n]}^{'} \right]^{'} A_n^{KM} \left[\sum_{i=1}^{N} W_{i:n} Z_{[i:n]} Q_{i:n} \right],$$

and Assumptions $1.i$ and $1.ii$ are satisfied.

Next we verify condition $1.iii$. From the Stute (1993)'s law of large numbers we have that, for each fixed β,

$$\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right] = \mathbb{E}_\infty^{KM}\left[g\left(\beta\right)\right] + o_p\left(1\right).$$

Then, it suffices to show that $\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]$ is stochastically equicontinuous. This follows from $\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]$ being Lipschitz, that is,

$$\left\| \mathbb{E}_n^{KM}\left[g\left(\tilde{\beta}\right)\right] - \mathbb{E}_n^{KM}\left[g\left(\beta\right)\right] \right\| \leq \left\| \sum_{i=1}^{n} W_{i:n} Z_{[i:n]} \left(Y_{i:n} - X_{[i:n]}^{\prime}\right) \right\| \left\| \tilde{\beta} - \beta \right\|,$$

where $\sum_{i=1}^{n} W_{i:n} Z_{[i:n]} \left(Y_{i:n} - X_{[i:n]}^{\prime}\right) = \mathbb{E}_\infty^{KM}\left[Z\left(Y - X'\right)\right] + o_p\left(1\right) = O_p\left(1\right)$ from Stute (1993)'s law of large numbers. Hence, condition $1.iii$ is satisfied.

Next, we proceed with the asymptotic normality. Assumptions $2.i$ follows from $\left\|\mathbb{E}_n^{KM}\left[g\left(\beta\right)\right]\right\|_A$ being uniquely minimized at $\hat{\beta}^{KM}$, and Assumption $2.ii$ follows from $\mathbb{E}\left[ZX'\right]$ having full column rank. We directly assume that $\beta_0 \in int\left(\Theta\right)$. Examples of A_n that trivially satisfy Assumption $2.v$ are $A_n = I_k$, or $A_n = \left(n^{-1}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1}$. This last choice of A_n leads to the Kaplan-Meier Two Stage Least Squares. Since other options of A_n are possible, we directly assume that Assumption $2.v$ is satisfied.

In order to verify Assumption $2.iii$, we will combine the results of Stute (1996a) with empirical process theory. First, define

$$\gamma_1^1(y;\beta) = \frac{1}{1-H(y)} \int 1\{y < w\} z(w - x'\beta)\gamma_0(w) \tilde{H}_{11}(dw, dx, dz),$$
$$\gamma_2^1(y;\beta) = \int\int \frac{1\{v < y, v < w\} z(w - x'\beta)\gamma_0(w)}{(1-H(v))^2} \tilde{H}_0(dv) \tilde{H}_{11}(dw, dx, dz).$$

From Stute (1996a), we have, for each $\beta \in \Theta_\delta$,

$$\begin{aligned}
&\sum_{i=1}^{n} W_{i:n} Z_{[i:n]} \left(Y_{i:n} - X'_{[i:n]}\beta\right) \\
&= \frac{1}{n} \sum_{i=1}^{N} \left[Z_i(Q_i - X_i'\beta)\gamma_0(Q_i)\delta_i + \gamma_1^1(Q_i;\beta)(1-\delta_i) - \gamma_2^2(Q_i;\beta)\right] + o_p\left(n^{-1/2}\right) \\
&= \frac{1}{n} \sum_{i=1}^{N} [\psi_{1,i}(\beta)] + o_p\left(n^{-1/2}\right).
\end{aligned}$$

It suffices to show that the class of functions

$$\begin{aligned}
\mathcal{F}_1 &= \{z(q - x'\beta) : \beta \in \Theta\}, \\
\mathcal{F}_2 &= \left\{\gamma_1^1(q;\beta) : \beta \in \Theta\right\}, \\
\mathcal{F}_3 &= \left\{\gamma_1^2(q;\beta) : \beta \in \Theta\right\}
\end{aligned}$$

are Donsker. It follows that, since $\Theta \subset \mathbb{R}^d$, by Lemma 2.6.15 of van der Vaart and Wellner (1996) these three classes of functions are Donsker. Hence, Assumption 2.iii is satisfied and we conclude that

$$\sqrt{n}\left(\hat{\beta}^{KM} - \beta_0\right) \xrightarrow{d} N(0, \Omega_1),$$

where

$$\begin{aligned}
\Omega_1 &= (\Gamma_1' A_1 \Gamma_1)^{-1} \Gamma_1' A V_1 A \Gamma_1 (\Gamma_1' A \Gamma_1)^{-1}, \\
\Gamma_1 &= \mathbb{E}\left[ZX'\right], \\
V_1 &= \mathbb{E}\left[\psi_1(\beta_0)\psi_1(\beta_0)'\right].
\end{aligned}$$

Notice that if one set $A_1 = V_1^{-1}$, the variance matrix simplifies to $\Omega_1 = \left(\Gamma_1' V_1^{-1} \Gamma_1\right)^{-1}$. Indeed, this is the well-known "optimal" weight matrix.

In order to conduct valid inference, one need to consistently estimate Ω_1. Notice that a simple consistent estimator of Γ_1 is given by

$$\hat{\Gamma}_{1,n} = \frac{1}{n} \sum_{i=1}^{n} Z_i X_i',$$

and a consistent estimator of A is given by A_n^{KM}. Hence, the main challenge to estimate Ω_1 is to consistently estimate V_1. If γ_0, $\gamma_1^1(\cdot; \beta_0)$ and $\gamma_2^2(\cdot; \beta_0)$ were known, V_1 could be estimated by the sample covariance $\hat{V}_{1,n} = n^{-1} \sum \left[\psi_i(\beta_0) \psi_i(\beta_0)'\right]$. In practice, however, they are not. Nonetheless, note that each of the $\gamma's$ is a function of the $H's$, which can be estimated from the data. Hence, replacing β_0 by $\hat{\beta}^{KM}$, $\tilde{H}_0(v)$ by

$$\hat{H}_{0,n}(v) = n^{-1} \sum_{i=1}^{n} 1\{Q_i \leq v, \delta_i = 0\}$$

and $\tilde{H}_{11}(v, x, x)$ by

$$\hat{H}_{11,n}(v) = n^{-1} \sum_{i=1}^{n} 1\{Q_i \leq v, X_i \leq x, Z_i \leq z, \delta_i = 1\},$$

we obtain an estimator of V_1. An alternative method to estimate V_1 would be an adaptation of the Jackknife to the GMM case, as studied for ordinary Kaplan-Meier integrals in Stute (1996*b*) and for Kaplan-Meier integrals with covariates by Azarang et al. (2014).

A third alternative is to use the ordinary nonparametric bootstrap for censored data introduced by Efron (1981), and estimate directly Ω_1. Let $\{(Q_i^*, \delta_i^*, X_i^*, Z_i^*)\}_{i=1}^n$ be drawn randomly with replacement from $\{(Q_i, \delta_i, X_i, Z_i)\}_{i=1}^n$. Let

$$\mathbb{E}_n^{KM,*}[g(\beta)] = \sum_{i=1}^{n} W_{i:n}^* Z_{[i:n]}^* \left(Q_{i:n}^* - X_{[i:n]}^{*\prime} \beta\right)$$

where $Q_{1:n}^* \leq \cdots \leq Q_{n:n}^*$ be the ordered Q^* values, $X_{[i:n]}^*$ and $Z_{[i:n]}^*$ be the concomitant of the ith order statistics, and

$$W^*_{k:n} = \frac{\delta^*_{[k:n]}}{n-k+1} \prod_{l=1}^{k-1} \left(\frac{n-l}{n-l+1} \right)^{\delta^*_{[l:n]}}.$$

Following Hall and Horowitz (1996), it is necessary to recenter the moment condition when one has an overidentified model, that is, when $k > d$. Define the recentered bootstrapped moment function

$$\mathbb{E}_n^{KM,\ c,\ *} [g(\beta)] = \sum_{i=1}^{n} W^*_{i:n} Z^*_{[i:n]} \left(Q^*_{i:n} - X^{*\prime}_{[i:n]}\beta \right) - \sum_{i=1}^{n} W_{i:n} Z_{[i:n]} \left(Q_{i:n} - X'_{[i:n]}\hat{\beta}^{KM} \right).$$

Thus, define the bootstrap estimator $\hat{\beta}^{KM,\ *}$ to be any sequence that satisfies

$$\mathbb{E}_n^{KM,\ c,\ *} \left[g\left(\hat{\beta}^{KM,\ *}\right) \right]' A_n \mathbb{E}_n^{KM,\ c,\ *} \left[g\left(\hat{\beta}^{KM,\ *}\right) \right]$$
$$= \inf_{\beta \in \Theta} \left[\mathbb{E}_n^{KM,\ c,\ *} [g(\beta)]' A_n \left[\mathbb{E}_n^{KM,\ c,\ *} [g(\beta)] \right] \right] + o_{p^*}\left(n^{-1/2}\right). \quad (3.8)$$

For $b = 1, \ldots, B$, we compute $\hat{\beta}^{KM,\ *}$ solving the problem (3.8), and, under some regularity conditions, we can use the empirical distribution of $\sqrt{n}\left(\hat{\beta}^{KM,\ *} - \hat{\beta}^{KM}\right)$ to make asymptotically valid inference on β_0 for large B. For instance, let $\hat{\beta}_j^{KM,\ *}$ and $\beta_{0,j}$ denote the jth components of $\hat{\beta}^{KM,\ *}$ and β_0. Then, the 0.025 and 0.975 quantiles of $\left(\hat{\beta}_{1,j}^{KM,\ *}, \ldots, \hat{\beta}_{B,j}^{KM,\ *}\right)$ form a 95% asymptotic confidence interval for $\beta_{0,j}$.

3.5 Monte Carlo simulations

In this section we develop a Monte Carlo numerical example aimed at analyzing the finite-sample performance of the Kaplan-Meier linear GMM models described in Section 3.4. The Monte Carlo designs considered here are chosen to illustrate the method for simple examples, and are not meant to mimic a design that would be encounter for a particular data set. Nevertheless, with these simple designs we can illustrate how our method perform when some of the assumptions related to the censoring process are violated.

Let (X, X_u, Z) be uniform random variables such that $Corr(X, X_u) = 0.5$, $Corr(X, Z) =$

0.5, and X_u is independent of Z.

We consider the following designs:

$$(i) \quad Y = 1 + \beta X + u,$$
$$u = X_u + \varepsilon,$$
$$C = Uniform[0, a_1];$$

$$(ii) \quad Y = 1 + \beta X + u,$$
$$u = X_u + \varepsilon \left| X \right|,$$
$$C = Uniform[0, a_2];$$

$$(iii) \quad Y = 1 + X + u,$$
$$u = X_u + \varepsilon,$$
$$C = 0.5X + Uniform[0, a_3];$$

where $\beta = 1$, u is the unobserved error, ε is an *iid* standard normal random variable, and a_i, $i = 1, 2, 3$, is set such that one achieves 10 and 30 % of censoring. Note that in all designs $E[Xu] \neq 0$ but $E[Zu] = 0$. Design (i) we have a homoskedastic regression, whereas in design (ii) we have heteroskedastic errors u. Finally, in design (iii) we have the case of dependent censoring, that is, Assumption 3.1 is violated.

Though we can observe Y in the simulated data, we censor the data to $Q = \min(Y, C)$ and also generate $\delta = 1\{Q \leq C\}$. Hence, to use the KM-GMM, we consider the random sample $\{(Q_i, \delta_i, X_i, Z_i)\}_{i=1}^n$. We compare the performance of the KM-GMM estimator with the unfeasible IV estimator in which one observes $\{(Y_i, X_i, Z_i)\}_{i=1}^n$. The comparisons are in terms of mean bias, root mean squared error (rmse), and 95% coverage probability for β. In order to compute the coverage probability, we adopt the bootstrap approach describe in the previous section. Given that we have a just-identified model, the choice of A_n plays no role. Moreover, one does not need to recenter the bootstrap moment equations, speeding up the simulation results. Our results are based 1000 Monte-Carlo experiments and 1000 bootstrap draws. We report our results for sample sizes $n = 100$, 300 and 1000 in Table 3.1

From the results of Table 3.1, one can see that the KM-GMM procedure produces estimators as good as the unfeasible GMM procedure in terms of average bias, rmse and 95% coverage probability. As the censoring increases, both the mean bias and the rmse increases. This is expected given that the "effective" sample size is reduced. As the sample size grows, the bias and the rmse of the KM-GMM estimator reduces in all designs, including (*iii*), where the censoring is informative. Furthermore, the coverage probability is close to the nominal level. Overall, our results shows that the finite sample properties of the KM-GMM in a linear model are satisfactory.

3.6 Conclusion

In this paper we have developed a general GMM estimator which is suitable for randomly right-censored data, and can easily accommodate endogenous regressors. In order to tackle the censoring problem, our estimators are characterized by means of Kaplan-Meier integrals. We illustrate our general results by examining of the asymptotic behavior of the Kaplan-Meier linear GMM with endogenous covariates. In a small Monte Carlo experiment, we showed that our estimator has good finite sample properties.

Many interesting extensions of our procedure are feasible. For instance, one can construct a Durbin–Wu–Hausman endogeneity test, a J-test for overidentifying restrictions, and tests for weak instruments that are suitable for randomly censored data. This can be done by replacing the empirical distribution function with the multivariate Kaplan-Meier distribution function as we have used to define the KM-GMM. Analogously, one can also propose Kaplan-Meier versions of the LIML or the Fulller-k estimators that are more robust against weak instruments.

Table 3.1: Simulation results for Kaplan-Meier linear GMM.

DGP		$n = 100$ $\hat{\beta}$	$\hat{\beta}_{10}^{KM}$	$\hat{\beta}_{30}^{KM}$	$n = 300$ $\hat{\beta}$	$\hat{\beta}_{10}^{KM}$	$\hat{\beta}_{30}^{KM}$	$n = 1000$ $\hat{\beta}$	$\hat{\beta}_{10}^{KM}$	$\hat{\beta}_{30}^{KM}$
	Mean bias	-0.0339	-0.0259	-0.0280	-0.0038	-0.0124	-0.0075	-0.0146	-0.0128	-0.0116
(i)	RMSE	0.7891	0.8464	0.9810	0.4419	0.4655	0.5799	0.2438	0.2585	0.3129
	95% Coverage	0.9390	0.9400	0.9540	0.9350	0.9410	0.9490	0.9380	0.9390	0.9520
	Mean bias	-0.0186	-0.0327	-0.0853	-0.0042	-0.0116	-0.0312	-0.0004	-0.0014	-0.0077
(ii)	RMSE	0.4994	0.5487	0.6845	0.2721	0.2881	0.3597	0.1487	0.1580	0.2039
	95% Coverage	0.9290	0.9370	0.9270	0.9380	0.9510	0.9540	0.9520	0.9490	0.9290
	Mean bias	-0.0294	-0.0152	0.0240	-0.0077	0.0000	-0.0342	-0.0076	0.0003	0.0326
(iii)	RMSE	0.7907	0.8678	1.0627	0.4395	0.4706	0.5787	0.2419	0.2580	0.3166
	95% Coverage	0.9420	0.9400	0.9470	0.9440	0.9380	0.9430	0.9240	0.9450	0.9470

Note: $\hat{\beta}$ denotes the unfeasible GMM estimator (no censoring), and $\hat{\beta}_k^{KM}$ the KM-GMM estimator in the design with k% of censoring.

Another interesting extension is to allow for cases where the criterion function may depend on some nonparametric estimators that can themselves depend on the finite-dimensional parameters of interest, and therefore, extend the results of Chen et al. (2003) to the randomly censored case. Alternatively, one can also extend our procedure to the setup where one is interested in semi or nonparametric conditional moment models as described in Chen and Pouzo (2009, 2012). By doing so, we would be able to extend the nonparametric mean IV and quantile IV regression models for the right-censoring setup, see e.g. Hall and Horowitz (2005), Chernozhukov et al. (2007), Horowitz and Lee (2007) and Darolles et al. (2011). A detailed analysis of these extensions is beyond the scope of this article and is deferred to future work.

3.7 Appendix

In this appendix, we prove our main results. First, we present the proof of the consistency result in Theorem 3.1

Proof of Theorem 3.1. We assume that $A_n^{KM} = A$, and for notational ease we suppress the dependence of the norm on the fixed symmetric positive definite matrix A.

The proof follows the same steps as in Corollary 3.2. of Pakes and Pollard (1989). Fix ε. Then, by Assumption $1.ii$ and using that if an event A implies an event B, $\mathbb{P}(A) \leq \mathbb{P}(B)$, we have

$$\mathbb{P}\left(\left\|\hat{\beta}_{GMM}^{KM} - \beta_0\right\| > \varepsilon\right) \leq \mathbb{P}\left(\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| > \eta\right).$$

Hence it will suffice to show that $\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| = o_p(1)$. To do this, by triangle inequality and Assumption $1.iii$,

$$\begin{aligned}
&\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| = \left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right] + \mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right] - \mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| \\
&\leq \left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| + \left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right] - \mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| \\
&\leq \left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| + o_p(1)\left[1 + \left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| + \left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|\right].
\end{aligned}$$

Rearranging the terms, we have

$$\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|\left[1-o_{p}\left(1\right)\right]\leq o_{p}\left(1\right)+\left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|\left[1+o_{p}\left(1\right)\right]. \tag{3.9}$$

From Assumptions 1.i and 1.iii, and the fact that $\mathbb{E}_{\infty}^{KM}\left[g\left(\beta_{0}\right)\right]=0$,

$$\begin{aligned}\left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| &\leq o_{p}\left(1\right)+\left\|\mathbb{E}_{n}^{KM}\left[g\left(\beta_{0}\right)\right]\right\|\\ &\leq o_{p}\left(1\right),\end{aligned}$$

concluding the proof. ■

Next, we move to the proof of Theorem 3.2.

Proof of Theorem 3.2: We assume that $A_{n}^{KM}=A$, and for notational ease we suppress the dependence of the norm on the fixed symmetric positive definite matrix A. The proof for general random matrices A_{n}^{KM} that satisfy Assumption 2.v follows from Lemma 3.5 of Pakes and Pollard (1989).

The proof follows similar steps as Theorem 3.3. of Pakes and Pollard (1989). First, we prove $\sqrt{n}$-consistency. By $\hat{\beta}_{GMM}^{KM}$ being consistent, we can choose a sequence $\{\delta_{n}\}$ of positive numbers that converges to zero slowly enough to ensure that

$$\mathbb{P}\left(\left\|\hat{\beta}_{GMM}^{KM}-\beta_{0}\right\|>\delta_{n}\right)\to 0,\text{ as } n\to\infty.$$

Then, for such sequence the supremum in Assumption 2.iii runs over an interval that includes $\hat{\beta}_{GMM}^{KM}$ with probability one. Thus, by triangle inequality and Assumption 2.iii,

$$\begin{aligned}&\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|-\left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|-\left\|\mathbb{E}_{n}^{KM}\left[g\left(\beta_{0}\right)\right]\right\|\\ &\leq\left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]-\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]-\mathbb{E}_{n}^{KM}\left[g\left(\beta_{0}\right)\right]\right\|\\ &\leq o_{p}\left(n^{-1/2}\right)+o_{p}\left(\left\|\mathbb{E}_{n}^{KM}\left[g\left(\beta\right)\right]\right\|\right)+o_{p}\left(\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\beta\right)\right]\right\|\right).\end{aligned}$$

Rearranging the terms, we have

$$\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|\left[1-o_p\left(1\right)\right] \leq o_p\left(n^{-1/2}\right) \qquad (3.10)$$
$$+\left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|\left[1+o_p\left(1\right)\right]$$
$$+\left\|\mathbb{E}_{n}^{KM}\left[g\left(\beta_0\right)\right]\right\|.$$

From the fact that $\mathbb{E}_{\infty}^{KM}\left[g\left(\beta_0\right)\right]=0$, Assumptions 3.1-3.3 and Stute (1996a)'s central limit theorem for censored data, we have

$$\sqrt{n}\mathbb{E}_{n}^{KM}\left[g\left(\beta_0\right)\right] \xrightarrow{d} N\left(0,V\right) \text{ as } n\to\infty, \qquad (3.11)$$

where V is defined as in (3.5). Hence, from Assumptions 2.i and (3.11)

$$\left\|\mathbb{E}_{n}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| \leq \left\|\mathbb{E}_{n}^{KM}\left[g\left(\beta_0\right)\right]\right\|+o_p\left(n^{-1/2}\right)=O_p\left(n^{-1/2}\right). \qquad (3.12)$$

Therefore, plugging in (3.12) on (3.10), it follows that

$$\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\|=O_p\left(n^{-1/2}\right).$$

Note that Assumption 2.ii implies that there exist a positive constant $\mathcal{C}<\infty$ such that, in a neighborhood of β_0,

$$\left\|\mathbb{E}_{\infty}^{KM}\left[g\left(\hat{\beta}_{GMM}^{KM}\right)\right]\right\| \geq \mathcal{C}\left\|\hat{\beta}_{GMM}^{KM}-\beta_0\right\|.$$

Therefore, we have

$$\left\|\hat{\beta}_{GMM}^{KM}-\beta_0\right\|=O_p\left(n^{-1/2}\right). \qquad (3.13)$$

Next, we must establish asymptotic normality of $\sqrt{n}\left(\hat{\beta}_{GMM}^{KM}-\beta_0\right)$. To do so, we will argue that $\mathbb{E}_{n}^{KM}\left[g\left(\cdot\right)\right]$ is well approximated by the linear function

$$L_{n}^{KM}\left(\beta\right)=\Gamma\left(\beta-\beta_0\right)+\mathbb{E}_{n}^{KM}\left[g\left(\beta_0\right)\right] \qquad (3.14)$$

within a $O_p\left(n^{-1/2}\right)$ neighborhood of β_0. In particular, we need to verify if the approximation error is of order $o_p\left(n^{-1/2}\right)$ at $\hat{\beta}^{KM}_{GMM}$ and at $\tilde{\beta}_{GMM}$ that globally minimizes $\left\|L^{KM}_n(\cdot)\right\|$. The assertion for $\hat{\beta}^{KM}_{GMM}$ follows from the $\sqrt{n}-consistency$ established in (3.13), Assumptions 2.ii and 2.iii, the triangle inequality and a Taylor expansion argument:

$$\begin{aligned}
&\left\|\mathbb{E}^{KM}_n\left[g\left(\hat{\beta}^{KM}_{GMM}\right)\right] - L^{KM}_n(\beta)\right\| \\
&= ||\mathbb{E}^{KM}_n\left[g\left(\hat{\beta}^{KM}_{GMM}\right)\right] - \mathbb{E}^{KM}_\infty\left[g\left(\hat{\beta}^{KM}_{GMM}\right)\right] - \mathbb{E}^{KM}_n\left[g(\beta_0)\right] \\
&+ \mathbb{E}^{KM}_\infty\left[g\left(\hat{\beta}^{KM}_{GMM}\right)\right] - \Gamma(\beta-\beta_0)\,|| \\
&\le \left\|\mathbb{E}^{KM}_n\left[g\left(\hat{\beta}^{KM}_{GMM}\right)\right] - \mathbb{E}^{KM}_\infty\left[g\left(\hat{\beta}^{KM}_{GMM}\right)\right] - \mathbb{E}^{KM}_n\left[g(\beta_0)\right]\right\| \\
&+ \left\|\mathbb{E}^{KM}_\infty\left[g\left(\hat{\beta}^{KM}_{GMM}\right)\right] - \Gamma\left(\hat{\beta}^{KM}_{GMM}-\beta_0\right)\right\| \\
&\le o_p\left(n^{-1/2}\right) + o_p\left(\left\|\mathbb{E}^{KM}_n\left[g(\beta)\right]\right\|\right) + o_p\left(\left\|\mathbb{E}^{KM}_\infty\left[g(\beta)\right]\right\|\right) \\
&+ o_p\left(\left\|\hat{\beta}^{KM}_{GMM}-\beta_0\right\|\right) \\
&= o_p\left(n^{-1/2}\right).
\end{aligned} \tag{3.15}$$

Next, we verify the approximation error of (3.14) at $\tilde{\beta}_{GMM}$. Notice that

$$\sqrt{n}\left(\tilde{\beta}_{GMM}-\beta_0\right) = -\left(\Gamma' A \Gamma\right)^{-1}\Gamma' A\sqrt{n}\mathbb{E}^{KM}_n\left[g(\beta_0)\right] \tag{3.16}$$

is the minimizer of $\left\|L^{KM}_n(\cdot)\right\|$, and From (3.11), the right-hand side of (3.16) has the asymptotic distribution specified in Theorem 3.2. Using the exact same arguments as Pakes and Pollard (1989), we have that, by Assumptions 2.ii and 2.iii,

$$\left\|\mathbb{E}^{KM}_n\left[g\left(\tilde{\beta}_{GMM}\right)\right] - L^{KM}_n\left(\tilde{\beta}_{GMM}\right)\right\| = o_p\left(n^{-1/2}\right). \tag{3.17}$$

To conclude the proof, it suffices to show that $\sqrt{n}\left(\hat{\beta}^{KM}_{GMM}-\tilde{\beta}_{GMM}\right) = o_p(1)$. From

(3.15) and (3.17), it follows that

$$\begin{aligned} \left\| L_n^{KM}\left(\hat{\beta}_{GMM}^{KM}\right)\right\| - o_p\left(n^{-1/2}\right) &\leq \left\| \mathbb{E}_n^{KM}\left[g\left(\tilde{\beta}_{GMM}\right)\right]\right\| \\ &\leq \left\| \mathbb{E}_n^{KM}\left[g\left(\tilde{\beta}_{GMM}\right)\right]\right\| + o_p\left(n^{-1/2}\right) \\ &\leq \left\| L_n^{KM}\left(\tilde{\beta}_{GMM}\right)\right\| + o_p\left(n^{-1/2}\right). \end{aligned}$$

That is,

$$\left\| L_n^{KM}\left(\hat{\beta}_{GMM}^{KM}\right)\right\| = \left\| L_n^{KM}\left(\tilde{\beta}_{GMM}\right)\right\| + o_p\left(n^{-1/2}\right). \tag{3.18}$$

Squaring both sides of (3.18) we get

$$\left\| L_n^{KM}\left(\hat{\beta}_{GMM}^{KM}\right)\right\|^2 = \left\| L_n^{KM}\left(\tilde{\beta}_{GMM}\right)\right\|^2 + o_p\left(n^{-1}\right). \tag{3.19}$$

Note that, about its global minimum,

$$\left\| L_n^{KM}\left(\beta\right)\right\|^2 = \left\| L_n^{KM}\left(\tilde{\beta}_{GMM}\right)\right\|^2 + \left\| \Gamma\left(\beta - \tilde{\beta}_{GMM}\right)\right\|^2. \tag{3.20}$$

Putting β equal to $\hat{\beta}_{GMM}^{KM}$, and equating (3.19) and (3.20) to conclude that

$$\left\| \Gamma\left(\hat{\beta}_{GMM}^{KM} - \tilde{\beta}_{GMM}\right)\right\| = o_p\left(n^{-1/2}\right).$$

Since Γ has full rank, this is equivalent to $\sqrt{n}\left(\hat{\beta}_{GMM}^{KM} - \tilde{\beta}_{GMM}\right) = o_p\left(1\right)$, and therefore the proof is completed. ■

Bibliography

Abadie, A. (2002), "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the American Statistical Association*, 97(457), 284–292.

Abbring, J. H., and van den Berg, G. J. (2003), "The nonparametric identification of treatment effects in duration models," *Econometrica*, 71(5), 1491–1517.

Abbring, J. H., and van den Berg, G. J. (2005), "Social experiments and instrumental variables with duration outcomes," *IFS Working Papers*, 19.

Agosto, A., Cavaliere, G., Kristensen, D., and Rahbek, A. (2015), "Modeling corporate defaults: Poisson autoregressions with exogenous covariates (PARX)," *CREATES Research Paper*, 11.

Akritas, M. G. (1994), "Nearest neighbor estimation of a bivariate distribution under random censoring," *The Annals of Statistics*, 22(3), 1299–1327.

Anderson, T. (1993), "Goodness of Fit Tests for Spectral Distributions," *The Annals of Statistics*, 21(2), 830–847.

Andrews, D. W. K., and Shi, X. (2013), "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

Andrews, D. W. K., and Shi, X. (2014), "Nonparametric inference based on conditional moment inequalities," *Journal of Econometrics*, 179(1), 31–45.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), "Identification of causal effects using instrumental variables," *Journal of the American Statistical Association*, 91(434), 444–455.

Azarang, L., de Uña Álvarez, J., and Stute, W. (2014), "The Jackknife estimate of covariance of two KaplanMeier integrals with covariables," *Statistics*, 33(3), 1–21.

Azizpour, S., Giesecke, K., and Schwenkler, G. (2014), "Exploring the sources of default clustering," *Preprint*, pp. 1–18.

Bae, J., and Kim, S. (2003), "The uniform central limit theorem for the Kaplan-Meier integral process," *Bulletin of the Australian Mathematical Society*, 67, 467–480.

Barrett, G., and Donald, S. G. (2003), "Consistent tests for stochastic dominance," *Econometrica*, 71(1), 71–104.

Bierens, H. J., and Ploberger, W. (1997), "Asymptotic theory of integrated conditional moment tests," *Econometrica*, 65(5), 1129–1151.

Bijwaard, G. E., and Ridder, G. (2005*a*), "Correcting for selective compliance in a re-employment bonus experiment," *Journal of Econometrics*, 125, 77–111.

Bijwaard, G. E., and Ridder, G. (2005*b*), "Correcting for selective compliance in a re-employment bonus experiment," *Journal of Econometrics*, 125(1-2 SPEC. ISS.), 77–111.

Billingsley, P. (1999), *Convergence of probability measures*, 2nd editio edn, New York: Wiley.

Bitler, M. O., Gelbach, J. B., and Hoynes, H. H. (2006), "What mean impacts miss: Distributional effects of welfare reform experiments," *The American Economic Review*, 96(4), 988–1012.

Bitler, M. O., Gelbach, J. B., and Hoynes, H. H. (2014), Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment,.

Bitler, M. P., Gelbach, J. B., and Hoynes, H. W. (2008), "Distributional impacts of the Self-Sufficiency Project," *Journal of Public Economics*, 92(3-4), 748–765.

Blundell, R., and Powell, J. L. (2007), "Censored regression quantiles with endogenous regressors," *Journal of Econometrics*, 141(1), 65–83.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31(3), 307–327.

Box, G., and Pierce, D. A. (1970), "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal of the American Statistical Association*, 65(332), 1509–1526.

Breunig, C., Mammen, E., and Simoni, A. (2014), "Semi/Nonparametric Estimation in case of Endogeneous Selection," *Mimeo*, .

Brown, R. L., Durbin, J., and Evans, J. M. (1975), "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 37(2), 149–192.

Buchinsky, M., and Hahn, J. (1998), "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66(3), 653–671.

Buckley, J., and James, I. (1979), "Linear regression with censored data," *Biometrika*, 66(3), 429–436.

Cabus, P. (1998), "Un test de type Kolmogorov-Smirnov dans le cade de comparaison de fonctions de régression," *Comptes Rendus de lAcad emie des Sciences*, Series I(327), 939–942.

Chen, K. (2001), "Parametric models for response-biased sampling," *Journal of the Royal Statistical Society: Series B*, 63(4), 775–789.

Chen, X., Linton, O., and Van Keilegom, I. (2003), "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71(5), 1591–1608.

Chen, X., and Pouzo, D. (2009), "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals," *Journal of Econometrics*, 152(1), 46–60.

Chen, X., and Pouzo, D. (2012), "Estimation of Nonparametric Conditional Moment Models With Possibly Nonsmooth Generalized Residuals," *Econometrica*, 80(1), 277–321.

Chernozhukov, V., Fernandez-Val, I., and Kowalski, A. (2015), "Quantile Regression with Censoring and Endogeneity," *Journal of Econometrics*, 186, 50–.

Chernozhukov, V., and Hong, H. (2002), "Three-Step Censored Quantile Regression and Extramarital Affairs," *Journal of the American Statistical Association*, 97(459), 872–882.

Chernozhukov, V., Imbens, G. W., and Newey, W. K. (2007), "Instrumental variable estimation of nonseparable models," *Journal of Econometrics*, 139(1), 4–14.

Christou, V., and Fokianos, K. (2014), "Quasi-likelihood inference for negative binomial time series models," *Journal of Time Series Analysis*, 35(1), 55–78.

Cosslett, S. R. (2004), "Efficient Semiparametric Estimation of Censored and Truncated Regressions via a Smoothed Self-Consistency Equation," *Econometrica*, 72(4), 1277–1293.

Cox, D. R. (1972), "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34(2), 187–220.

Creal, D., Koopman, S. J., and Lucas, A. (2013), "Generalized autoregressive score models with applications," *Journal of Applied Econometrics*, 28(5), 777–795.

Crépon, B., Ferracci, M., Jolivet, G., and van den Berg, G. J. (2009), "Active labor market policy effects in a dynamic setting," *Journal of the European Economic Association*, 7(May 2009), 595–605.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2008), "Nonparametric tests for treatment effect heterogeneity," *The Review of Economics and Statistics*, 90(3), 389–405.

Darolles, S., Fan, Y., Florens, J. P., and Renault, E. (2011), "Nonparametric Instrumental Regression," *Econometrica*, 79(5), 1541–1565.

Das, S. R., Duffie, D., Kapadia, N., and Saita, L. (2007), "Common failings: How corporate defaults are correlated," *The Journal of Finance*, LXII(1), 93–118.

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2003), "Observation-Driven Models for Poisson Counts," *Biometrika*, 90(4), 777–790.

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2005), "Maximum likelihood estimation for an observation driven model for Poisson counts," *Methodology and Computing in Applied Probability*, 7(2), 149–159.

Davis, R. A., Dunsmuir, W. T., and Wang, Y. (2000), "On autocorrelation in a Poisson regression model," *Biometrika*, 87(3), 491–505.

Davis, R. A., and Wu, R. (2009), "A negative binomial model for time series of counts," *Biometrika*, 96(3), 735–749.

Delgado, M. A. (1993), "Testing the equality of nonparametric regression curves," *Statistics & Probability Letters*, 17(3), 199–204.

Delgado, M. a., and Escanciano, J. C. (2013), "Conditional stochastic dominance testing," *Journal of Business & Economic Statistics*, 31(1), 16–28.

Delgado, M. A., and González-Manteiga, W. (2001), "Significance testing in nonparametric regression based on the bootstrap," *The Annals of Statistics*, 29(5), 1469–1507.

Delgado, M. A., Hidalgo, J., and Velasco, C. (2005), "Distribution free goodness-of-fit tests for linear processes," *The Annals of Statistics*, 33(6), 2568–2609.

Delgado, M. A., and Velasco, C. (2010), "Distribution-free tests for time series models specification," *Journal of Econometrics*, 155(2), 128–137.

Delgado, M. A., and Velasco, C. (2011), "An Asymptotically Pivotal Transform of the Residuals Sample Autocorrelations with Application to Model Checking," *Journal of the American Statistical Association*, 116(495), 946–958.

D'Haultfoeuille, X. (2010), "A new instrumental method for dealing with endogenous selection," *Journal of Econometrics*, 154(1), 1–15.

Donald, S. G., and Hsu, Y.-C. (2013), "Estimation and inference for distribution functions and quantile functions in treatment effect models," *Journal of Econometrics*, .

Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009), "Frailty Correlated Default," *The Journal of Finance*, LXIV(5), 2089–2123.

Eberwein, C., Ham, J. C., and Lalonde, R. J. (1997), "The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: Evidence from experimental data," *The Review of Economic Studies*, 64(4), 655–682.

Efron, B. (1981), "Censored data and the bootstrap," *Journal of the American Statistical Association*, 76(374), 312–319.

Engle, R. (2002), "New frontiers for arch models," *Journal of Applied Econometrics*, 17(5), 425–446.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1007.

Escanciano, J. C. (2006*a*), "A consistent diagnostic test for regression models using projections," *Econometric Theory*, 22, 1030–1051.

Escanciano, J. C. (2006*b*), "Goodness-of-Fit Tests for Linear and Nonlinear Time Series Models," *Journal of the American Statistical Association*, 101(474), 531–541.

Escanciano, J. C., and Lobato, I. N. (2009), "An automatic Portmanteau test for serial correlation," *Journal of Econometrics*, 151(2), 140–149.

Escanciano, J. C., Lobato, I. N., and Zhu, L. (2013), "Automatic specification testing for vector autoregressions and multivariate nonlinear time series models," *Journal of Business & Economic Statistics*, 31(4), 426–437.

Fokianos, K., and Neumann, M. H. (2013), "A goodness-of-fit test for Poisson count processes," *The Electronic Journal of Statistics*, 7, 793–819.

Fokianos, K., Rahbek, A., and Tjostheim, D. (2009), "Poisson Autoregression," *Journal of the American Statistical Association*, 104(488), 1430–1439.

Fokianos, K., and Tjostheim, D. (2011), "Log-linear Poisson autoregression," *Journal of Multivariate Analysis*, 102(3), 563–578.

Francq, C., Roy, R., and Zakoïan, J.-M. (2005), "Diagnostic Checking in ARMA Models With Uncorrelated Errors," *The Journal of the American Statistical Association*, 100(470), 532–544.

Frandsen, B. R. (2014), "Treatment Effects With Censoring and Endogeneity," *Mimeo*, .

Frey, R., and McNeil, A. J. (2002), "VaR and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights," *Journal of Banking & Finance*, 26(7), 1317–1334.

Giesecke, K., Longstaff, F. a., Schaefer, S., and Strebulaev, I. (2011), "Corporate bond default risk: A 150-year perspective," *Journal of Financial Economics*, 102(2), 233–250.

Gill, R. (1983), "Large sample behaviour of the product-limit estimator on the whole line," *The Annals of Statistics*, 11(1), 49–58.

Giné, E., and Zinn, J. (1990), "Bootstrapping general empirical measures," *The Annals of Probability*, 18(2), 851–869.

González-Manteiga, W., and Cadarso-Suárez, C. (1994), "Asymptotic properties of a generalized kaplan-meier estimator with some applications," *Journal of Nonparametric Statistics*, 4(1), 65–78.

Gordy, M. B. (2000), "A comparative anatomy of credit risk models," *Journal of Banking & Finance*, 24(1-2), 119–149.

Gordy, M. B. (2003), "A risk-factor model foundation for ratings-based bank capital rules," *Journal of Financial Intermediation*, 12(3), 199–232.

Grambsch, P., and Therneau, T. (1994), "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, 81(3), 515–526.

Hahn, J. (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66(2), 315–331.

Hall, P., and Horowitz, J. L. (1996), "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64(4), 891–916.

Hall, P., and Horowitz, J. L. (2005), "Nonparametric methods for inference in the presence of instrumental variables," *Annals of Statistics*, 33(6), 2904–2929.

Ham, J. C., and Lalonde, R. J. (1996), "The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training," *Econometrica*, 64(1), 175–205.

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

Heckman, J. J., and Vytlacil, E. J. (2007), "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation," *Handbook of Econometrics*, 6B(70), 4779–4874.

Hirano, K., Imbens, G. W., and Ridder, G. (2003), "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71(4), 1161–1189.

Hong, H., and Tamer, E. (2003), "Inference in censored models with endogenous regressors," *Econometrica*, 71(3), 905–932.

Hong, Y. (1996), "Consistent Testing for Serial Correlation of Unknown Form," *Econometrica*, 64(4), 837.

Honore, B., Khan, S., and Powell, J. (2002), "Quantile regression under random censoring," *Journal of Econometrics*, 109, 67–105.

Horowitz, J. L. (1986), "A distribution-free least squares estimator for censored linear regression models," *Journal of Econometrics*, 32(1), 59–84.

Horowitz, J. L., and Lee, S. (2007), "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, 75(4), 1191–1208.

Hsu, Y.-C. (2013), "Consistent tests for conditional treatment effects," *Mimeo*, .

Hubbard, A. E., Laan, M. J. V. D., and Robins, J. M. (2000), "Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies," in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, eds. M. Halloran, and D. Berry, Vol. 116, New York: Springer, pp. 135–177.

Ichimura, H., and Linton, O. (2005), "Asymptotic expansions for some semiparametric program evaluation estimators," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. W. K. Andrews, and J. H. Stock Cambridge University Press, chapter 8, pp. 149–170.

Iglesias Pérez, C., and González-Manteiga, W. (1999), "Strong representation of a generalized product-limit estimator for truncated and censored data with some applications," *Journal of Nonparametric Statistics*, 10(3), 213–244.

Imbens, G. W. (2004), "Nonparametric estimation of average treatment effects under exogeneity: A review," *The Review of Economics and Statistics*, 86(1), 4–29.

Imbens, G. W., and Angrist, J. D. (1994), "Identification and estimation of local average treatment effects," *Econometrica*, 62(2), 467–475.

Imbens, G. W., and Wooldridge, J. M. (2009), "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 47(1), 5–86.

Jung, R. C., and Tremayne, A. R. (2003), "Testing for serial dependence in time series models of counts," *Journal of Time Series Analysis*, 24(1), 65–84.

Kaplan, E. L., and Meier, P. (1958), "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, 53(282), 457–481.

Keenan, S. C., Sobehart, J., and Hamilton, D. T. (1999), "Predicting default rates: a forecasting model for Moodys issuer-based default rates.," *Working Paper, Moodys Investors Services*, .

Khan, S., and Tamer, E. (2009), "Inference on endogenously censored regression models using conditional moment inequalities," *Journal of Econometrics*, 152(2), 104–119.

Khan, S., and Tamer, E. (2010), "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78(6), 2021–2042.

Koopman, S. J., Lucas, A., and Schwaab, B. (2011), "Modeling frailty-correlated defaults using many macroeconomic covariates," *Journal of Econometrics*, 162(2), 312–325.

Koopman, S. J., Lucas, A., and Schwaab, B. (2012), "Dynamic Factor Models With Macro, Frailty, and Industry Effects for U.S. Default Counts: The Credit Crisis of 2008," *Journal of Business & Economic Statistics*, 30(4), 521–532.

Koul, H. L., and Stute, W. (1999), "Nonparametric model checks for time series," *The Annals of Statistics*, 27(1), 204–236.

Koul, H. L., Susarla, V., and Van Ryzin, J. (1981), "Regression Analysis with Randomly Right-Censored Data," *The Annals of Statistics*, 9(6), 1276–1288.

Lee, M.-j. (2009), "Non-parametric tests for distributional treatment effect for randomly censored responses," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1), 243–264.

Lee, M.-j., and Lee, S.-j. (2005), "Analysis of job-training effects on Korean women," *Journal of Applied Econometrics*, 20(4), 549–562.

Lee, S., Song, K., and Whang, Y.-J. (2013), "Testing functional inequalities," *Journal of Econometrics*, 172(1), 14–32.

Lee, S., and Whang, Y. (2009), "Nonparametric tests of conditional treatment effects," *Cowles Foundation Discussion Paper*, 1740.

Lewbel, A. (2000), "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables," *Journal of Econometrics*, 97(1), 145–177.

Li, Q., Racine, J. S., and Wooldridge, J. M. (2009), "Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data," *Journal of Business & Economic Statistics*, 27(2), 206–223.

Liboschik, T., Fokianos, K., and Fried, R. (2015), "tscount : An R Package for Analysis of Count Time Series Following Generalized Linear Models," *Mimeo*, .

Ljung, G. M., and Box, G. E. P. (1978), "On a measure of lack of fit in time series models," *Biometrika*, 65(2), 297.

Lobato, I. N., Nankervis, J. C., and Savin, N. E. (2002), "Testing for Zero Autocorrelation in the Presence of Statistical Dependence," *Econometric Theory*, 18(03), 730–743.

Mammen, E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models," *The Annals of Statistics*, 21(1), 255–285.

Meyer, B. D. (1996), "What have we learned from the Illinois reemployment bonus experiment?," *Journal of Labor Economics*, 14(1), 26–51.

Miller, R. G. (1976), "Least squares regression with censored data," *Biometrika*, 63(3), 449–464.

Neumann, M. H. (2011), "Absolute regularity and ergodicity of Poisson count processes," *Bernoulli*, 17(4), 1268 – 1284.

Neumeyer, N., and Dette, H. (2003), "Nonparametric comparison of regression curves: an empirical process approach," *The Annals of Statistics*, 31(3), 880–920.

Newey, W. K. (1997), "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79(1), 147–168.

Newey, W. K., and McFadden, D. (1994), "Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 4, 2111–2245.

Newey, W. K., and Powell, J. L. (1990), "Efficient estimation of linear and type I censored regression models under conditional quantile restrictions," *Econometric Theory*, 6(3), 295–317.

Pakes, A., and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027–1057.

Pardo-Fernandez, J. C., and Van Keilegom, I. (2006), "Comparison of regression curves with censored responses," *Scandinavian Journal of Statistics*, 33(3), 409–434.

Personal, M., Archive, R., Siew-ling, S., Mansor, A., and Khim-sen, V. (2014), "Poisson qmle of count time series models," *MPRA Paper*, 59804.

Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98(464), 1001–1012.

Powell, J. L. (1984), "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, 25(3), 303–325.

Powell, J. L. (1986), "Censored regression quantiles," *Journal of Econometrics*, 32(1), 143–155.

Ritov (1991), "Estimation in linear regression with censored data," *The Annals of Statistics*, 18(1), 303–328.

Romano, J., and Thombs, L. (1996), "Inference for autocorrelations under weak assumptions," *Journal of the American Statistical Association*, 91(434), 590–600.

Rosenbaum, P. R., and Rubin, D. B. (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

Roussas, G. G., and Ioannides, D. (1987), "Moment Inequalities for Mixing Sequences of Random Variables," *Stochastic Analysis and Applications*, 5(1), 61–120.

Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66(5), 688–70.

Satten, G. A., and Datta, S. (2001), "The KaplanMeier estimator as an inverse-probability-of-censoring weighted average," *The American Statistician*, 55(3), 207–210.

Sellero, C. S., González-Manteiga, W., and van Keilegom, I. (2005), "Uniform representation of product-limit integrals with applications," *Scandinavian Journal of Statistics*, 32, 563–581.

Shorack, G. R., and Wellner, J. A. (1986), *Empirical processes with applications to statistics*, New York: Wiley.

Sianesi, B. (2004), "An evaluation of the Swedish system of active labor market programs in the 1990s," *The Review of Economics and Statistics*, 86(1), 133–155.

Srihera, R., and Stute, W. (2010), "Nonparametric comparison of regression functions," *Journal of Multivariate Analysis*, 101(9), 2039–2059.

Stinchcombe, M. B., and White, H. (1998), "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric theory*, 14, 295–325.

Stute, W. (1993), "Consistent estimation under random censorship when covariables are present," *Journal of Multivariate Analysis*, 45, 89 – 103.

Stute, W. (1994), "The bias of Kaplan-Meier integrals," *Scandinavian Journal of Statistics*, 21(4), 475–484.

Stute, W. (1995), "The central limit theorem under random censorship," *The Annals of Statistics*, 23(2), 422–439.

Stute, W. (1996*a*), "Distributional convergence under random censorship when covariables are present," *Scandinavian Journal of Statistics*, 23(4), 461–471.

Stute, W. (1996*b*), "The jackknife estimate of variance of a Kaplan-Meier integral," *The Annals of Statistics*, 24(6), 2679–2704.

Stute, W. (1997), "Nonparametric model checks for regression," *The Annals of Statistics*, 25(2), 613–641.

Stute, W. (1999), "Nonlinear censored regression," *Statistica Sinica*, 9, 1089–1102.

Stute, W., González-Manteiga, W., and Quindimil, M. P. (1998), "Bootstrap Approximations in Model Checks for Regression," *Journal of the American Statistical Association*, 93(441), 141–149.

Stute, W., Manteiga, W. G., and Sellero, C. S. (2000), "Nonparametric model checks in censored regression," *Communications in Statistics - Theory and Methods*, 29(7), 1611 – 1629.

Stute, W., and Wang, J.-L. (1993*a*), "Multi-sample U-statistics for censored data," *Scandinavian journal of statistics*, 20(4), 369–374.

Stute, W., and Wang, J.-L. (1993*b*), "The strong law under random censorship," *The Annals of Statistics*, 21(3), 1591–1607.

Sun, J., and Mccabe, B. (2012), "Score Statistics for Testing the Serial Dependence in Count Data," *Journal of Time Series Analysis*, .

Tang, G., Little, R., and Raghunathan, T. (2003), "Analysis of multivariate missing data with nonignorable nonresponse," *Biometrika*, 90(4), 747–764.

Tarullo, D. K. (2010), "Equipping Financial Regulators with the Tools Necessary to Monitor Systemic Risk," *Statement before the Banking Subcommittee on Security and International Trade and Finance U.S. Senate, 12 February 2010*, .

Taylor, S. (1986), *Modelling Financial Time Series*, New York: Wiley.

van den Berg, G. J. (2001), "Duration models: specification, identification and multiple durations," in *Handbook of Econometrics*, eds. J. J. Heckman, and E. Leamer, Vol. 5 Elsevier, chapter 55, pp. 3381–3460.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.

Wang, H. J., and Wang, L. (2009), "Locally weighted censored quantile regression," *Journal of the American Statistical Association*, 104(487), 1–32.

Woodbury, S. A., and Spiegelman, R. G. (1987), "Bonuses to workers and employers to reduce unemployment: Randomized trials in Illinois," *The American Economic Review*, 77(4), 513–530.

Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Survival and Hazard Functions," *Journal of the American Statistical Association*, 94(445), 137–145.

Ying, Z., Jung, S., and Wei, L. (1995), "Survival analysis with median regression models," *Journal of the American Statistical Association*, 90(429), 178–184.

Zeger, S. L. (1988), "A Regression Model for Time Series of Counts," *Biometrika*, 75(4), 621.

Zhou, M. (1992), "Asymptotic Normality of the 'Synthetic Data' Regression Estimator for Censored Survival Data," *The Annals of Statistics*, 20(2), 1002–1021.