11-1-2013

New Goodness-of-fit Diagnostics for Conditional Discrete Response Models

Igor Kheifets

Carlos Velasco

Recommended Citation

Kheifets, Igor and Velasco, Carlos, "New Goodness-of-fit Diagnostics for Conditional Discrete Response Models" (2013). *Cowles Foundation Discussion Papers*. 2317.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/2317

NEW GOODNESS-OF-FIT DIAGNOSTICS
FOR CONDITIONAL DISCRETE RESPONSE MODELS

By

Igor Kheifets and Carlos Velasco

November 2013
Revised May 2017

COWLES FOUNDATION DISCUSSION PAPER NO. 1924R

New goodness-of-fit diagnostics for conditional discrete response models

Igor Kheifets* and Carlos Velasco†

May 31, 2017

Abstract

This paper proposes new specification tests for conditional models with discrete responses, which are key to apply efficient maximum likelihood methods, to obtain consistent estimates of partial effects and to get appropriate predictions of the probability of future events. In particular, we test the static and dynamic ordered choice model specifications and can cover infinite support distributions for e.g. count data. The traditional approach for specification testing of discrete response models is based on probability integral transforms of a jittered discrete data which leads to continuous uniform iid series under the true conditional distribution. Then, standard specification testing techniques for continuous variables could be applied to the transformed series, but the extra randomness from jitters affects the power properties of these methods. We investigate in this paper an alternative transformation based only on original discrete data that avoids any randomization. We analyze the asymptotic properties of goodness-of-fit tests based on this new transformation and explore the properties in finite samples of a bootstrap algorithm to approximate the critical values of test statistics which are model and parameter dependent. We show analytically and in simulations that our approach dominates the methods based on randomization in terms of power. We apply the new tests to models of the monetary policy conducted by the Federal Reserve.

Keywords: Specification tests, count data, dynamic discrete choice models, conditional probability integral transform.
JEL classification: C12, C22, C52.

---

*ITAM, Mexico. Email: igor.kheifets@itam.mx
†Department of Economics, Universidad Carlos III de Madrid. Email: carlos.velasco@uc3m.es

1 INTRODUCTION

Many statistical models specify the conditional distribution of a discrete response variable given some explanatory variables, including the description of binary, multinomial, ordered choice and count data. In this paper we analyze goodness-of-fit tests for both static models with covariates as well as dynamic ordered choice and count data models, where the conditioning information set may also include past information on the discrete variable and a set of (contemporaneous) explanatory variables which frequently appear in the social sciences, see Kedem and Fokianos (2002) and Greene and Hensher (2010). For example, dynamic models are popular in macroeconomic applications, see for instance Hamilton and Jordá (2002), Dolado and Maria-Dolores (2002) and Basu and de Jong (2007) for modeling central banks decisions or Kauppi and Saikkonen (2008) and Startz (2008) for predicting US recessions; in finance, see e.g. Rydberg and Shephard (2003) for modeling the size of asset price movements and Fokianos et al. (2009) for the number of transactions per minute of a particular stock.

Suppose we observe the random variables $\{Y_t, X_t'\}_{t=1}^T$ and consider the information sets $\Omega_t = \{X_t, Y_{t-1}, X_{t-1}, Y_{t-2}, X_{t-2}, \ldots\}$ for each period $t = 1, 2, \ldots, T$. We are interested in testing the null hypothesis that the distribution of Y_t conditional on Ω_t is in the parametric family $F_{t,\theta}(\cdot \mid \Omega_t)$, i.e.

$$H_0 : Y_t \mid \Omega_t \sim F_{t,\theta_0}(\cdot \mid \Omega_t) \text{ for some } \theta_0 \in \Theta, \ t = 1, 2, \ldots, T,$$

where $\Theta \subset \mathbb{R}^m$ is the parameter space, while the alternative hypothesis (H_1) for the omnibus test would be the negation of H_0.

We consider a class $\mathcal{M}$ of discrete conditional distributions defined on $\mathcal{K} = \{1, 2, \ldots, K\}$, for integer $K > 1$ or on $\mathcal{K} = \{1, 2, \ldots, \infty\}$ such that for all $F \in \mathcal{M}$ it holds that $F(0) = 0$, $f(k) := F(k) - F(k-1) > 0$ for all $k = 1, 2, \ldots$ and $\sum_{k \in \mathcal{K}} f(k) = 1$. This setup includes numerous models that have been used extensively in applied work both for dynamic and for iid data, here we describe briefly two of them.

Example 1 (Dynamic multinomial ordered choice model)**.** The discrete responses Y_t are

assumed to be generated by the rule

$$Y_t = \begin{cases} 1 & \text{if} & V_t^* \leq \tau_1 \\ 2 & \text{if} & \tau_1 < V_t^* \leq \tau_2 \\ \vdots & & \\ K & \text{if} & V_t^* > \tau_{K-1}, \end{cases}$$

where V_t^* is a continuous latent variable and $\tau_1, \ldots, \tau_{K-1}$ are threshold parameters that define K intervals in $\mathbb{R}$. In a simple model, e.g. Basu and de Jong (2007), the latent variable is determined through the linear equation

$$V_t^* = X_t'\beta + \rho Y_{t-1} + \varepsilon_t,$$

where X_t is a vector of stationary exogenous regressors, β a vector of regression parameters, ε_t is the shock in each period, and Y_{t-1} could be replaced by any function of the past $\{Y_{t-1}, \ldots, Y_{t-n}\}$ for some finite n. The cdf of ε_t, F_ε, is going to determine the class of multinomial model, i.e. ordered multinomial probit (if ε_t is standard normal) or logit (if ε_t is logistic), since F_{t,θ_0} is defined at once from

$$\Pr\left(Y_t = k \mid \Omega_t\right) = \Pr\left(\tau_{k-1} < V_t^* \leq \tau_k \mid \Omega_t\right)$$
$$= F_\varepsilon\left(\tau_k - X_t'\beta - \rho Y_{t-1}\right) - F_\varepsilon\left(\tau_{k-1} - X_t'\beta - \rho Y_{t-1}\right),$$

with $\tau_0 = -\infty$ and $\tau_K = \infty$ and $\theta_0 = (\beta', \rho, \tau_1, \ldots, \tau_{K-1})'$.

Example 2 (Poisson Model)**.** The variate $Y_t = Y_t^* + 1$ is defined on the counts $Y_t^* = 0, 1, 2, \ldots$ which are assumed to follow a conditional Poisson distribution

$$Y_t^* \mid \Omega_t \sim \text{Poisson}(\lambda_t),$$

where the conditional mean can depend on covariates through an exponential link as $\lambda_t = \exp(X_t'\beta)$ or on previous observations through an identity link as $\lambda_t = \alpha_0 + \alpha_1 \lambda_{t-1} + \rho Y_{t-1}^*$, e.g. Fokianos et al. (2009), or through the logarithmic canonical link as $\log(\lambda_t) = X_t'\beta + \rho e_{t-1}$, where $e_t = (Y_t^* - \lambda_t)/\lambda_t$ are scaled and centered errors, e.g. Davis et al. (2003).

Despite that a correct specification is key to apply efficient maximum likelihood methods, to obtain consistent estimates of partial effects and to get appropriate predictions of the probability of future events, empirical researchers typically do not perform goodness of fit testing of such models as they would do in a continuous case. In general, there are only a few specification tests available for discrete data, see Mora and Moro-Egido (2007). Two of them, the test of the Generalized Linear Model (GLM) of Stute and Zhu (2002) and the conditional Kolmogorov test of Andrews (1997), based on the specification of the conditional mean for binary data, can be adapted for this purpose and we discuss this possibility and compare it to our approach in Section 6. A related test to Andrews derived for time series by Corradi and Swanson (2006) could be adapted also for discrete data, but this is testing a different null hypothesis concerning a distribution given a finite conditioning set not characterizing the complete dynamics of the process. There are also tests designed specifically for Poisson models (see e.g. Neumann 2011; Fokianos and Neumann, 2013).

In what follows we propose conditional, dynamic discrete analogs of the Kolmogorov-Smirnov goodness of fit measure that can exploit different restrictions derived from the martingale difference property of a particular transformation of the data under the null hypothesis. This property is derived from the specification of a complete dynamic model given the information set generated by all the past observations of the discrete response and other explanatory variables and is used to build the asymptotic theory for our tests. Under i.i.d. assumptions this martingale difference property leads to an exact independence of the transformation sequence under the null and a much simpler parallel asymptotic theory.

When the fitted distribution is continuous, the *relative distribution* of Y_t compared to F_{t,θ_0} defined as the cdf of the Rosenblatt's (1952) transforms, also called conditional Probability Integral Transforms (PIT),

$$U_t(\theta_0) := F_{t,\theta_0}(Y_t \mid \Omega_t), \qquad t = 1, 2, \ldots, T$$

is standard uniform and $U_t(\theta_0)$ are distributed as independent $[0, 1]$ uniform random

4

variables under H_0. This serves as a basis for several specification tests of H_0, see e.g. Bai (2003) and Kheifets (2015) for dynamic models and Delgado and Stute (2008) for independent and identical distributed (iid) data. However Rosenblatt transformation is not appropriate for discrete support random variables, producing non-iid pseudo residuals even under the null of correct specification. To solve the limitations of PIT-based testing techniques for discrete data, several alternative transforms have been proposed, see Jung, Kukuk and Liesenfeld (2006), Czado, Gneiting and Held (2009) and references therein. An easy and popular way is to randomize, i.e. to interpolate the discrete values of Y_t with independent noise in $[0,1]$, recent references include Kheifets and Velasco (2013) and Lee (2014). Unfortunately the additional simulated noise affects the power of the tests and may lead to different conclusions depending on the simulation outcome.

In this paper instead, we consider a *nonrandomized transform* $Y_t \mapsto I_{t,\theta_0}(u)$ for $u \in [0,1]$,

$$
I_{t,\theta_0}(u) := \begin{cases} 0, & u \leq U_t^-(\theta_0); \\ \dfrac{u - U_t^-(\theta_0)}{U_t(\theta_0) - U_t^-(\theta_0)}, & U_t^-(\theta_0) \leq u \leq U_t(\theta_0); \\ 1, & U_t(\theta_0) \leq u, \end{cases} \tag{1}
$$

where $U_t^-(\theta_0) := F_{t,\theta_0}(Y_t - 1 \mid \Omega_t)$. This transform, conditional on data, is nonrandomized in the sense that it does not depend on extra sources of randomness, as opposed to interpolation transforms discussed in the next section. The unconditional version of this transform appears in Handcock and Morris (1999) and more recently in Czado, Gneiting and Held (2009) where it is used for calibration, but no formal tests are proposed there. This transformation can also be seem as a particular case of the multilinear extension as defined in Genest, Nešlehová and Rémillard (2014). As we show below, for every $u \in [0,1]$, $I_{t,\theta_0}(u) - u$ constitute a martingale difference sequence (MDS) with respect to Ω_t under H_0 and can be used for testing H_0 as $I_{t,\theta_0}(u)$ loses this property when the model is misspecified. For instance, we can compute the pseudo empirical relative distribution of Y_t compared to F_{t,θ_0}

$$
\widetilde{F}_{\theta_0}(u) := \frac{1}{T} \sum_{t=1}^{T} I_{t,\theta_0}(u), \quad u \in [0,1],
$$

5

which can be contrasted with the uniform cdf using the following empirical process

$$S_{1T}(u) := \frac{1}{T^{1/2}} \sum_{t=1}^{T} \{I_{t,\theta_0}(u) - u\} = T^{1/2}\left(\widetilde{F}_{\theta_0}(u) - u\right),$$

which converges weakly to a Gaussian process. In addition, in order to control dynamics in $I_{t,\theta_0}(u)$, we can compare the joint pseudo empirical cdf with the uniform on a square using the biparameter process

$$S_{2T}(u) := \frac{1}{(T-1)^{1/2}} \sum_{t=2}^{T} \{I_{t,\theta_0}(u_1) I_{t-1,\theta_0}(u_2) - u_1 u_2\}, \tag{2}$$

where $u = (u_1, u_2)$. To obtain feasible tests we need to consider norms of S_{jT} for $j = 1, 2$. We use the Cramer-von Mises $\int S_{jT}(u)^2 d\varphi(u)$ for some absolute continuous measure φ in $[0, 1]^j$, or Kolmogorov-Smirnov $\sup_{u \in [0,1]^j} |S_{jT}(u)|$ norms.

When the parameter θ_0 is unknown under the null, we use an estimate $\widehat{\theta}_T$ and account for the parameter estimation effect in the p-value computation with a parametric bootstrap method. It might be possible also to derive, e.g. martingale, distribution-free transforms, but since they typically need to be programmed on a case by case basis for each model, so can be impractical, and are beyond the scope of this paper. As far as we know, our proposal is the first formal specification test of ordered discrete choice models which accounts properly for parameter uncertainty and is based on a nonrandomized transform, which makes it attractive in terms of power against a wide set of alternative hypotheses.

The rest of the paper is organized as follows. In the next section, we describe different alternatives to the PIT. In Sections 3 and 4, we provide the main asymptotic properties of the nonrandomized transforms and of the resulting univariate and bivariate empirical processes using martingale theory. In particular, we establish weak limits under fixed and local alternatives accounting for parameter estimation effect. Section 5 discusses the implementation of new tests with a simple bootstrap algorithm. Section 6 provides a small simulation exercise and an application exploring the properties of specification tests based on both randomized and non randomized transformations. Then we conclude. All proofs are contained in the Appendix.

2   ALTERNATIVES TO PIT FOR DISCRETE DATA

In order to further motivate the nonrandomized transform I_{t,θ_0} defined in (1), we introduce the *randomized* PIT,

$$U_t^r(\theta_0) := U_t^-(\theta_0) + Z_t^U\left(U_t(\theta_0) - U_t^-(\theta_0)\right), \tag{3}$$

where $\{Z_t^U\}_{t=1}^T$ are independent standard uniform random variables, and independent of Y_t. Alternatively, U_t^r can be obtained by applying the standard continuous PIT to the *continuous* random variable $Y_t^\dagger := Y_t - 1 + Z_t$, where $\{Z_t\}_{t=1}^T$ are iid with any continuous cdf F_Z on $[0,1]$. Indeed, we can construct the cdf of $Y_t^\dagger$,

$$F_{t,\theta_0}^\dagger(y \mid \Omega_t) = F_{t,\theta_0}(\lfloor y \rfloor \mid \Omega_t) + F_Z(y - \lfloor y \rfloor)(F_{t,\theta_0}(\lfloor y + 1 \rfloor \mid \Omega_t) - F_{t,\theta_0}(\lfloor y \rfloor \mid \Omega_t)),$$

where $\lfloor y \rfloor$ is the floor function, i.e. the maximum integer not exceeding y, and find that

$$U_t^r(\theta_0) = F_{t,\theta_0}^\dagger\left(Y_t^\dagger \mid \Omega_t\right),$$

for $Z_t^U = F_Z(Z_t)$ and any choice of F_Z, see Kheifets and Velasco (2013). Note that the cdf of $Y_t^\dagger$ conditional on Ω_t and $\{\Omega_t, Z_{t-1}, Z_{t-2}, \ldots, Z_1\}$ coincide. Under H_0, $U_t^r(\theta_0)$ are iid $U[0,1]$ variables as under any continuous distribution specification, while $U_t(\theta_0)$ and $U_t^-(\theta_0)$ are not independent nor $U[0,1]$. Using the typical discrepancy measures, the empirical cdf of $U_t^r(\theta_0)$, estimated using the *randomized transform* $Y_t \mapsto \mathbb{1}\{U_t^r(\theta_0) \leq u\}$,

$$\widehat{F}_{\theta_0}^r(u) := \frac{1}{T}\sum_{t=1}^T \mathbb{1}\{U_t^r(\theta_0) \leq u\}, \quad u \in [0,1],$$

can be compared to the uniform cdf. Kheifets and Velasco (2013) then test H_0 using empirical process based on the randomized transform

$$R_{1T}(u) := T^{1/2}\left\{\widehat{F}_{\theta_0}^r(u) - u\right\} = \frac{1}{T^{1/2}}\sum_{t=1}^T \left[\mathbb{1}\{U_t^r(\theta_0) \leq u\} - u\right], \quad u \in [0,1].$$

We can also consider reducing the dependence on a particular outcome of the noise Z_t^U in (3) and in the randomized transform by taking averages over M replications of $\{Z_t^U\}_{t=1}^T$, conditional on the original data, similar to "average-jittering" of Machado and

Santos Silva (2005). Suppose that for each t we have M independent sequences of uniform $U[0,1]$ noises $Z_{t,m}^U$, $m = 1, 2, \ldots, M$, which generate $U_{t,m}^r (\theta_0)$ according to (3). Define the *M-random* transform $Y_t \mapsto I_{t,\theta_0,M} (Y_t, u)$,

$$I_{t,\theta_0,M} (Y_t, u) := \frac{1}{M} \sum_{m=1}^M \mathbb{1}\left\{U_{t,m}^r (\theta_0) \leq u\right\},$$

which takes values on the set $\{0, 1/M, 2/M, \ldots, 1\}$ and has mean u under H_0. Then the cdf of $U_t^r (\theta_0)$ is estimated by

$$\widehat{F}_{\theta_0,M}^r (u) := \frac{1}{T} \sum_{t=1}^T I_{t,\theta_0,M} (Y_t, u), \quad u \in [0,1].$$

Note that with $M = 1$ we are back to $\widehat{F}_{\theta_0}^r (u)$, and therefore, we can generalize R_{1T} to

$$R_{1T,M} (u) := T^{1/2} \left\{\widehat{F}_{\theta_0,M}^r (u) - u\right\}, \quad u \in [0,1].$$

In order to propose specification tests, following Handcock and Morris (1999), we define the *discrete relative distribution* of Y_t compared to F_{t,θ_0} as the cdf of $U_t^r (\theta_0)$. Under H_0, the discrete relative distribution is the uniform $U[0,1]$. As we show in the next section, three consistent estimators of the discrete relative distribution of Y_t compared to F_{t,θ_0} can be ordered in terms of efficiency in the following way: $\widetilde{F}_{\theta_0} (u)$ (the most efficient), $\widehat{F}_{\theta_0,M}^r (u)$ and $\widehat{F}_{\theta_0}^r (u)$. This ordering is determined by the amount of noise introduced in the definitions of the transforms: i.e. in nonrandomized, M-randomized and (1-)randomized transforms. The nonrandomized transform can be equivalently obtained by integrating out the extra noise in the randomized transform $I_{t,\theta_0} (Y_t, u) = \int \mathbb{1}\{U_t^r (\theta_0) \leq u\} dF_Z$ or by taking the number of replications M to infinity, thus completely removing the noise from the estimate of the discrete relative distribution and other functionals of the transforms. The efficiency of the nonrandomized transform translates into the increased power of the specification tests based on this transform, whose properties we study next.

8

3 PROPERTIES OF EMPIRICAL PROCESSES BASED ON THE NONRANDOMIZED TRANSFORM

As shown in the next lemma, the building blocks of $\widetilde{F}_{\theta_0}(u)$, $I_{t,\theta_0}(u) - u$, constitute a martingale difference sequence (MDS) with respect to Ω_t, and therefore $\widetilde{F}_{\theta_0}(u)$ is an unbiased and consistent estimate of the uniform cdf under the null, a reasonable basis for developing tests of H_0. Moreover, the MDS property will allow us to establish the asymptotic properties of our test without imposing any additional restrictions. Let for $u, v \in [0, 1]$

$$\gamma_{t,\theta_0}(u, v) := \frac{(F_k - u \vee v)(u \wedge v - F_{k-1})}{F_k - F_{k-1}} \mathbb{1}\{F_{t,\theta_0}^{-1}(u \mid \Omega_t) = F_{t,\theta_0}^{-1}(v \mid \Omega_t)\},$$

where $k = k(u) = F_{t,\theta_0}^{-1}(u \mid \Omega_t)$, with $F_{t,\theta_0}^{-1}(u \mid \Omega_t) := \min\{y : F_{t,\theta_0}(y \mid \Omega_t) \geq u\}$ being the conditional quantile function and $F_k := F_{t,\theta_0}(k \mid \Omega_t)$.

Lemma 1. *Under H_0, $I_{t,\theta_0}(u) - u$ is a martingale difference sequence with respect to Ω_t, i.e.*

$$\mathrm{E}\left[I_{t,\theta_0}(u) \mid \Omega_t\right] = u, \quad a.s.,$$

with conditional covariance

$$\mathrm{E}\left[I_{t,\theta_0}(u) I_{t,\theta_0}(v) \mid \Omega_t\right] = u \wedge v - uv - \gamma_{t,\theta_0}(u, v), \quad a.s.$$

Note that $I_{t,\theta_0}(u)$ are not necessarily independent across t despite the fact that by the martingale difference property, $I_{t,\theta_0}(u)$ and $I_{t-j,\theta_0}(v)$ are serially uncorrelated for all $j \neq 0$ and all $u, v \in [0, 1]$, see the Appendix. On the other hand, the $I_{t,\theta_0}(u)$ are (conditionally) heteroskedastic, therefore the variance of S_{1T} is model and parameter dependent, but its distribution can be simulated conditional on exogenous information in Ω_t.

Let $V_{1T}(u, v) := \mathrm{Cov}\left[S_{1T}(u), S_{1T}(v)\right]$, then since $0 \leq \gamma_{t,\theta_0}(u, v) < 1$ a.s.,

$$V_{1T}(u, v) = u \wedge v - uv - \frac{1}{T} \sum_{t=1}^{T} \mathrm{E}\left[\gamma_{t,\theta_0}(u, v)\right] \leq u \wedge v - uv,$$

9

i.e. the covariance and variance of S_{1T} are not larger than those of the randomized transformation-based process R_{1T} or its weak limit, the Brownian sheet, see Corollary 4 in Kheifets and Velasco (2013).

Due to Lemma 1, $\mathrm{E}\left[\widetilde{F}_{\theta_0}(u)\right] = u$ under H_0 and the natural empirical process for performing tests on H_0 is then S_{1T}. This process, being based on a nonrandomized transform, does not involve the extra noise that appears in the randomized transform based empirical process R_{1T} for testing $U_t^r \sim U[0,1]$, proposed by Kheifets and Velasco (2013), or in its modification $R_{1T,M}$ based on the M-randomized transform. The next lemma is the key to understand the improvement of the M-randomized over the randomized and of the nonrandomized, advocated in this paper, over the M-randomized transform approaches.

Lemma 2. *Suppose that the uniform law of large numbers holds for $\widehat{F}_{\theta_0,M}^r(u)$ and $\widetilde{F}_{\theta_0}(u)$. Independently of whether H_0 holds or not, $\widehat{F}_{\theta_0,M}^r(u)$ and $\widetilde{F}_{\theta_0}(u)$ consistently and uniformly in u estimate the relative distribution, i.e. the cdf of $U_t^r(\theta_0)$. $\widetilde{F}_{\theta_0}(u)$ is more efficient, but the difference in efficiency goes to 0 as $M \to \infty$. In particular, under H_0,*

$$\mathrm{E}\left[R_{1T,M}(u) R_{1T,M}(v)\right] = \frac{1}{M} \mathrm{E}\left[R_{1T}(u) R_{1T}(v)\right] + \left(1 - \frac{1}{M}\right) \mathrm{E}\left[S_{1T}(u) S_{1T}(v)\right].$$

From Lemma 2, it follows that S_{1T} has the smallest variance, the variance of $R_{1T,M}$ is a weighted sum of those of S_{1T} and R_{1T}, see also Equation (5) in Machado and Santos Silva (2005). Other advantages of S_{1T} over $R_{1T,M}$, are 1) computational, as there is no need to simulate M paths of transformations and 2) theoretical, since the weak convergence is easier to prove for processes which are piece-wise linear in parameters. Therefore we concentrate on studying the properties of tests based on the nonrandomized transform, for which we introduce the following assumption.

Assumption 1. $F_{t,\theta_0}(\cdot \mid \Omega_t)(k) \in \mathcal{M}$ *a.s. for all t. Moreover, there exists a finite function $\gamma_\infty(u,v)$, such that uniformly in $(u,v) \in [0,1]^2$, $T^{-1} \sum_{t=1}^{T} \gamma_{t,\theta_0}(u,v) \to_p \gamma_\infty(u,v)$.*

This assumption implicitly restrict dynamics such that a uniform law of large numbers (LLN) holds for the averaged conditional covariance function. In the case of stationary

10

and ergodic data, $\gamma_\infty(u, v) = \mathrm{E}\left[\gamma_{1,\theta_0}(u, v)\right]$. Sufficient conditions for the stationarity and ergodicity of dynamic multinomial ordered choice models are given in Basu and de Jong (2007) and for autoregressive Poisson are given in Davis et al. (2003), Fokianos et al. (2009) and Doukhan et al. (2012). Then it is possible to show the uniformity of the convergence from a point-wise result, since the summands are continuous, piece-wise polynomials in u and v. As an illustration, in Section 8.5 in Appendix we discuss the assumptions for the Poisson model.

The next result describes the asymptotic distribution of S_{1T} under the null hypothesis. Let $\Rightarrow$ denote weak convergence in $\ell^\infty[0, 1]$, see e.g. van der Vaart and Wellner (1996). In fact, our empirical processes are continuous, which simplifies tightness verification. Let $V_{1\infty}(u, v) := u \wedge v - uv - \gamma_\infty(u, v)$.

Lemma 3. *Suppose Assumption 1 holds. Under H_0,*

$$S_{1T} \Rightarrow S_{1\infty},$$

where $S_{1\infty}$ is a Gaussian process in $[0, 1]$ with zero mean and covariance function $V_{1\infty}$.

The asymptotic distribution of S_{1T} is model and parameter dependent, and the practical implementation of tests when θ_0 is unknown is discussed in Section 3.2 after presenting a general class of local alternatives to the null of correct specification of the conditional distribution.

3.1 Local Alternatives

We next discuss the asymptotic properties of the empirical process S_{1T} under a class of alternative hypothesis, that will lead to consistency of the specification tests based on S_{1T} for a wide class of alternatives. We consider the following class of local alternatives to H_0,

$$H_{1T} : \quad Y_t \mid \Omega_t \sim G_{T,t,\theta_0}(\cdot \mid \Omega_t) \quad \text{for some } \theta_0 \in \Theta,$$

where

$$G_{T,t,\theta_0}(y \mid \Omega_t) = \left(1 - \frac{\delta}{T^{1/2}}\right) F_{t,\theta_0}(y \mid \Omega_t) + \frac{\delta}{T^{1/2}} H_t(y \mid \Omega_t),$$

11

for some $0 < \delta < T^{1/2}$ and for all t, $H_t\left(\cdot \mid \Omega_t\right) \in \mathcal{M}$. When $\delta = 0$ then H_{1T} nests H_0.

Following Kheifets and Velasco (2013), for any discrete distributions G and F in $\mathcal{M}$, with probability functions g and f, define

$$d\left(G, F, u\right) = G\left(F^{-1}\left(u\right)\right) - F\left(F^{-1}\left(u\right)\right)$$
$$- \frac{F\left(F^{-1}(u)\right) - u}{f\left(F^{-1}(u)\right)}\left[g\left(F^{-1}(u)\right) - f\left(F^{-1}(u)\right)\right].$$

Note, that $d\left(G, F, u\right) = E_G[I_F(Y, u)] - E_F[I_F(Y, u)] = E_G[I_F(Y, u)] - u$ and $d\left(G, F, u\right) \equiv 0$ if and only if $G \equiv F$. Under any $G_t\left(\cdot \mid \Omega_t\right) \in \mathcal{M}$,

$$\frac{1}{T^{1/2}}\, \mathrm{E}\left[S_{1T}\left(u\right)\right] = \frac{1}{T}\sum_{t=1}^{T}\mathrm{E}\left[d\left(G_t\left(\cdot \mid \Omega_t\right), F_{t,\theta_0}\left(\cdot \mid \Omega_t\right), u\right)\right].$$

The next assumption guarantees that a LLN can be applied to the empirical discrepancy between H_t and F_{t,θ_0}.

Assumption 2. Under H_{1T}, there exists a finite function $D_1\left(u\right)$, such that uniformly in $u \in [0, 1]$, $\frac{1}{T}\sum_{t=1}^{T} d\left(H_t\left(\cdot \mid \Omega_t\right), F_{t,\theta_0}\left(\cdot \mid \Omega_t\right), u\right) \to_p D_1\left(u\right)$.

Then the following lemma shows that the departure of H_0 in the direction of H_{1T} introduces a drift in the asymptotic distribution of S_{1T} that will render consistency of hypothesis tests based on functionals of H_{1T}.

Lemma 4. *Suppose Assumptions 1-2 hold. Under H_{1T},*

$$S_{1T} \Rightarrow S_{1\infty} + \delta D_1,$$

where $S_{1\infty}$ is as in Lemma 3.

3.2   Parameter Estimation Effect

In practice, tests based on S_{1T} are unfeasible since θ_0 is unknown, and has to be estimated by $\widehat{\theta}_T$, say. We assume that we have available an estimate $\widehat{\theta}_T$ so that under H_{1T}

$$T^{1/2}\left(\widehat{\theta}_T - \theta_0\right) = O_p\left(1\right),$$

and define the process with estimated parameters

$$\widehat{S}_{1T}(u) := \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left\{ I_{t,\widehat{\theta}_T}(u) - u \right\}.$$

We next analyze the consequences of replacing θ_0 by $\widehat{\theta}_T$ in $\widehat{S}_{1T}$.

Let $\| \cdot \|$ be Euclidean norm, i.e. for matrix A, $\|A\| = \sqrt{\text{tr}(AA')}$, where A' is a transpose of A. For $\varepsilon > 0$, $B(a, \varepsilon)$ is an open ball in $\mathbb{R}^m$ with the center at point a and radius ε. For a cdf F_θ in $\mathcal{M}$ define

$$\nabla(F_\theta, u) := \dot{F}_\theta\left(F_\theta^{-1}(u)\right) - \frac{F_\theta\left(F_\theta^{-1}(u)\right) - u}{f_\theta\left(F_\theta^{-1}(u)\right)} \dot{f}_\theta\left(F_\theta^{-1}(u)\right),$$

where $\dot{F}_\theta := (\partial/\partial\theta) F_\theta$ and $\dot{f}_\theta := (\partial/\partial\theta) f_\theta$. We need the following assumptions to analyze the asymptotic properties of $\widehat{S}_{1T}$.

Assumption 3 (Parametric family). (A) The parameter space Θ is a compact set in a finite-dimensional Euclidean space, $\theta \in \Theta \subset \mathbb{R}^m$.

(B) There exists $\delta > 0$, such that $F_{t,\theta}(\cdot \mid \Omega_t) \in \mathcal{M}$, for all t, Ω_t, T and $\theta \in B(\theta_0, \delta)$.

(C) $F_{t,\theta}(k \mid \Omega_t)$ is differentiable with respect to $\theta \in B(\theta_0, \delta)$ and under H_{1T}
$$\max_t \text{E}\left[\max_k \sup_{\theta \in B(\theta_0,\delta)} \left\|\dot{F}_{t,\theta}(k \mid \Omega_t)\right\|\right] \le M_F < \infty.$$

(D) Under H_{1T}, there exists a finite $L_1(u) := \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \nabla(F_{t,\theta_0}(\cdot \mid \Omega_t), u)$.

Conditions (A)-(C) about the parametric family of distribution are standard, see e.g. Bai (2003, Assumptions A1-A2). For dynamic ordered choice and Poisson models the differentiability of the conditional distribution with respect to the parameter is equivalent to the differentiability of the link function. Part (D) guarantees a nice limit behaviour of the average generalized derivative of $I_{t,\theta}$. Conditions for no effect of information truncation can be provided in a similar way to Bai (2003, Assumption A4).

The following lemma provides an expansion of the empirical process with estimated parameters as the sum of the process with known parameters and a random drift describing parameter estimation.

Lemma 5. *Suppose Assumptions 1-3 hold and $T^{1/2}\left(\widehat{\theta}_T - \theta_0\right) = O_p(1)$. Under H_{1T},*

$$\widehat{S}_{1T}(u) = S_{1T}(u) + T^{1/2}\left(\widehat{\theta}_T - \theta_0\right)' \frac{1}{T}\sum_{t=1}^{T}\nabla\left(F_{t,\theta_0}\left(\cdot\mid\Omega_t\right), u\right) + o_p(1), \qquad (4)$$

uniformly in u.

Then, continuous functionals of $\widehat{S}_{1T}$ no longer converge to those of $S_1 + \delta D_1$ under H_{1T}, but the estimation effect also has to be taken into account using the following assumption. Let $Z(\Psi)$ be a normal vector with zero mean and covariance matrix Ψ.

Assumption 4 (Parameter estimation). Under H_{1T}, the estimator $\widehat{\theta}_T$ admits the asymptotic linear expansion

$$T^{1/2}\left(\widehat{\theta}_T - \theta_0\right) = \delta\xi_0 + \frac{1}{T^{1/2}}\sum_{t=1}^{T}\ell_t(Y_t, \Omega_t) + o_p(1), \qquad (5)$$

where ξ_0 is a $m\times 1$ vector and the summands ℓ_t constitute a martingale difference sequence with respect to Ω_t, such that

(A) $\mathrm{E}\left[\ell_t(Y_t, \Omega_t)\mid\Omega_t\right] = 0$ and $T^{-1}\sum_{t=1}^{T}\mathrm{E}\left[\ell_t(Y_t, \Omega_t)\ell_t(Y_t, \Omega_t)'\mid\Omega_t\right] \overset{p}{\to} \Psi$.

(B) Lindeberg condition $T^{-1}\sum_{t=1}^{T}\mathrm{E}\left[\|\ell_t(Y_t, \Omega_t)\|^2\,\mathbb{1}\{T^{-1/2}\|\ell_t(Y_t, \Omega_t)\| > \varepsilon\}\mid\Omega_t\right] \overset{p}{\to} 0$ holds.

(C) There exists a finite function $W_1(u)$, such that $T^{-1}\sum_{t=1}^{T}\mathrm{E}\left[I_{t,\theta_0}(u)\ell_t(Y_t, \Omega_t)\mid\Omega_t\right] \to_p W_1(u)$ uniformly in u.

In particular, under H_0, $\delta\xi_0 = 0$, the estimate $\widehat{\theta}_T$ is centered and $T^{1/2}\left(\widehat{\theta}_T - \theta_0\right)$ converges in distribution to $Z(\Psi)$.

Assumption 4(A) and 4(B) hold for the MLE of many popular discrete models, including dynamic probit and logit and general discrete choice models. As an example consider

estimates $\widehat{\theta}_T$, which are asymptotically equivalent to the (conditional) maximum likelihood estimates, i.e.,

$$T^{1/2}\left(\widehat{\theta}_T - \theta_0\right) = -\frac{B_0^{-1}}{T^{1/2}}\sum_{t=1}^{T} s_t\left(Y_t, \Omega_t\right) + o_p\left(1\right),$$

where $s_t\left(k, \Omega_t\right) := \dot{f}_{t,\theta_0}\left(k \mid \Omega_t\right) / f_{t,\theta_0}\left(k \mid \Omega_t\right)$ is the score function and B_0 is a symmetric $m \times m$ positive definite matrix given by the limit of the Hessian,

$$B_0 := \plim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K} s_t\left(k, \Omega_t\right) \dot{f}_{t,\theta_0}\left(k \mid \Omega_t\right)'.$$

Under H_{1T}, $\mathrm{E}\left[s_t\left(Y_t, \Omega_t\right) \mid \Omega_t\right] = \delta T^{-1/2}\sum_{k=1}^{K} s_t\left(k, \Omega_t\right) h_t\left(k \mid \Omega_t\right)$. Then equation (5) holds with $\xi_0 = -\plim_{T\to\infty} B_0^{-1} T^{-1} \sum_{t=1}^{T}\sum_{k=1}^{K} s_t\left(k, \Omega_t\right) h_t\left(k \mid \Omega_t\right)$ and $\ell_t\left(Y_t, \Omega_t\right) = -B_0^{-1} s_t\left(Y_t, \Omega_t\right) + B_0^{-1}\sum_{k=1}^{K} s_t\left(k, \Omega_t\right) h_t\left(k \mid \Omega_t\right)$.

We can derive the covariance matrix between the process $S_{1T}\left(u\right)$ and $T^{1/2}\left(\widehat{\theta}_T - \theta_0\right)$ and obtain joint convergence results, so under H_{1T}

$$\left(S_{1T}, T^{1/2}\left(\widehat{\theta}_T - \theta_0\right)\right) \Rightarrow \left(S_{1\infty} + \delta D_1, Z\left(\Psi\right) + \delta \xi_0\right), \tag{6}$$

where the covariance function between $S_{1\infty}$ and $Z\left(\Psi\right)$ is $W_1(u)$.

We can state now the result on the asymptotic distribution of the empirical process $\widehat{S}_{1T}$ under local alternatives, whose drift is different with respect to the case without estimated parameters.

Theorem 1. *Suppose Assumptions 1-4 hold. Under H_{1T},*

$$\widehat{S}_{1T} \Rightarrow \widehat{S}_{1\infty} + \delta\left\{D_1 + \xi_0' L_1\right\},$$

where $\widehat{S}_{1\infty} := S_{1\infty} + Z\left(\Psi\right)' L_1$ is a Gaussian process with zero mean and variance function $V_1\left(u, v\right) + L_1\left(u\right)' \Psi L_1\left(v\right) + W_1\left(u\right)' L_1\left(v\right) + W_1\left(v\right)' L_1\left(u\right).$

15

4 EMPIRICAL PROCESSES FOR DYNAMIC SPEC-IFICATION

Test statistics based on S_{1T}, R_{1T} and $R_{1T,M}$ verify that the conditional distribution of Y_t is right on *average* across all possible Ω_t, so these tests might not capture all sources of misspecification. This issue is raised in Corradi and Swanson (2006), Delgado and Stute (2008) and Kheifets (2015) in relation to testing continuous distributions. However, it is not possible to develop specification tests conditioned on infinite dimensional values of Ω_t. Instead of truncating Ω_t or restricting the class of models, we consider S_{2T}, a biparameter analog of S_{1T} to control the possible dynamic misspecification. From Lemma 1, since under H_0, $I_{t,\theta_0}(u_1) - u_1$ is a MDS, $I_{t,\theta_0}(u_1) I_{t-1,\theta_0}(u_2) - u_1 u_2$ is centered around zero, and moreover

$$\mathrm{E}\left[I_{t,\theta_0}(u_1) I_{t-1,\theta_0}(u_2) \mid \Omega_{t-1}\right] = u_1 u_2, \quad a.s.$$

This motivates us to develop tests based on S_{2T} defined in (2). This process also has zero mean under the null and identifies not only departures from the null derived from deviations of the unconditional expectation of $I_{t,\theta_0}(u)$ from u, but also from a possible failure of the martingale property, so that $I_{t,\theta_0}(u_1)$ and $I_{t-1,\theta_0}(u_2)$ would become correlated. This idea is similar to that exploited in Kheifets' (2015) in the context of conditional distribution testing for continuous distributions, where different methods of checking the independence property of the PIT are proposed. Alternative statistics exploiting the lack of correlations with any other lag could be proposed, but we expect that low lags are typically more useful for detecting general forms of misspecification.

One could also consider a biparameter analog of $R_{1T,M}$, i.e. for some $M = 1, 2, \ldots,$

$$R_{2T,M}(u) := \frac{1}{(T-1)^{1/2} M} \sum_{t=2}^{T} \sum_{m=1}^{M} \left(\mathbb{1}\left\{U_{t,m}^r(\theta_0) \leq u_1\right\} \mathbb{1}\left\{U_{t-1,m}^r(\theta_0) \leq u_2\right\} - u_1 u_2 \right),$$

where $u = (u_1, u_2) \in [0,1]^2$. In particular, a bivariate analog of R_{1T}, $R_{2T}(u) := R_{2T,1}(u)$, is introduced in Kheifets and Velasco (2013). Tests based on R_{2T} and $R_{2T,M}$ involve randomized transforms and therefore suffer from power loss compared to tests based on the nonrandomized transform.

Note, that $S_{2T}(u) - u_1 S_{1T-1}(u_2)$ is a martingale. This observation will allow us to derive weak convergence of S_{2T} by employing limiting theorems for MDS. Properties of R_{2T} were established in Kheifets and Velasco (2013) and could be extended to $R_{2T,M}$. Here we discuss the properties of S_{2T} when we estimate θ_0.

In practice we use the process

$$\widehat{S}_{2T}(u) := \frac{1}{(T-1)^{1/2}} \sum_{t=2}^{T} \left\{ I_{t,\widehat{\theta}_T}(u_1) I_{t,\widehat{\theta}_T}(u_2) - u_1 u_2 \right\},$$

where we can write under H_{1T}

$$\widehat{S}_{2T}(u) = S_{2T}(u) + T^{1/2} \left(\widehat{\theta}_T - \theta_0\right)' \frac{1}{T} \sum_{t=2}^{T} \nabla_{2,t}(u) + o_p(1), \tag{7}$$

uniformly in u, where

$\nabla_{2,t}(u) := I_{t-1,\theta_0}(u_2) \nabla (F_{t,\theta_0}(\cdot \mid \Omega_t), u_1) + u_1 \nabla (F_{t-1,\theta_0}(\cdot \mid \Omega_{t-1}), u_2)$ and the asymptotic covariance function is $W_2(u) := \mathrm{ACov}\left(S_{2T}(u), T^{1/2}\left(\widehat{\theta}_T - \theta_0\right)\right)$. To study the asymptotic properties of the biparameter process we introduce the next assumption, which extends Assumption 2.

Assumption 5. Under H_{1T}, there exist finite functions $D_2(u)$ and $L_2(u)$, such that uniformly in u

(A) $T^{-1} \sum_{t=2}^{T} \{ I_{t-1,\theta_0}(u_2) d(H_t(\cdot \mid \Omega_t), F_{t,\theta_0}(\cdot \mid \Omega_t), u_1)$
$\qquad + u_1 d(H_t(\cdot \mid \Omega_t), F_{t,\theta_0}(\cdot \mid \Omega_t), u_2) \} \to_p D_2(u).$

(B) $T^{-1} \sum_{t=2}^{T} \nabla_{2,t}(u) \to_p L_2(u).$

Note that the second terms in the definitions of D_2 and L_2 correspond to $u_1 D_1(u_2)$ and $u_1 L_1(u_2)$ respectively, the equivalent for the single parameter process S_{1T}, but the first ones are new. To state the next result, we need to assume existence of probabilistic limits of several random functions. For the sake of presentation, we defer precise statements to the Appendix, see Assumption A.

17

Theorem 2. *Suppose that in addition to the conditions of Theorem 1, Assumption 5 and Assumption A from the Appendix hold. Under H_{1T},*

$$S_{2T}(u) \Rightarrow S_{2\infty} + \delta D_2,$$

where $S_{2\infty}$ is a Gaussian process in $[0,1]$ with mean zero and covariance function $V_{2\infty}(u,v)$ defined in the Appendix. Under H_{1T}, if parameters are estimated,

$$\widehat{S}_{2T} \Rightarrow \widehat{S}_{2\infty} + \delta\left\{D_2 + \xi_0' L_2\right\},$$

where $\widehat{S}_{2\infty} := S_{2\infty} + Z(\Psi)' L_2$ is a Gaussian process with zero mean and variance function $V_{2\infty}(u,v) + L_2(u)' \Psi L_2(v) + W_2(u)' L_2(v) + W_2(v)' L_2(u)$.

When $G_t(\cdot \mid \Omega_t)$ is different from $F_{t,\theta_0}(\cdot \mid \Omega_t)$ such that D_2 is non-zero, the test based on $\widehat{S}_{2T}$ has nontrivial power in the direction of H_{1T}. In contrast to the univariate case with S_{1T}, the first term in the definition of D_2 contains correlation with the past information and can therefore capture dynamic misspecification when this induces in such a correlation, even if the unconditional expectation of d, which appears in the second term $u_1 D_1(u_2)$, is zero. This fact is crucial if misspecification occurs in the dynamics and not only in the link function or other static aspects of the model.

5 BOOTSTRAP TESTS

To test H_0 we consider Cramer-von Mises, Kolmogorov-Smirnov or any other continuous functionals of $\widehat{S}_{jT}$, $j = 1,2$, $\eta\left(\widehat{S}_{jT}\right)$. Then consistency properties of specification tests based on $\widehat{S}_{jT}$ can be derived using the discussion in the previous sections by applying the continuous mapping theorem, so we omit the proof of the following result.

Theorem 3. *Suppose that conditions of Theorem 2 hold. Under H_{1T},*

$$\eta\left(\widehat{S}_{jT}\right) \to_d \eta\left(\widehat{S}_{j\infty}\right), \quad j = 1,2.$$

Since the asymptotic distributions of $S_{jT}(u)$ are model dependent, and those of $\widehat{S}_{jT}(u)$ further depend on the estimation effect, we need to resort to bootstrap methods to

implement our tests in practice. In the literature, there are several resampling methods suitable for dependent data, but since under H_0 the parametric conditional distribution is fully specified, we apply a conditional parametric bootstrap algorithm that only requires to make draws from $F_{t,\widehat{\theta}}\left(\cdot \mid \Omega_t\right)$ to mimic the null distribution of the test statistics. For a discussion of the parametric bootstrap see Stute et al. (1993) and Andrews (1997), which can be adapted to the complications with information truncation and initialization arising in the dynamic case using the discussion in Bai (2003).

To estimate the true $1 - \alpha$ quantiles $c_j\left(\theta_0\right)$ of the null asymptotic distribution of the test statistics, given by some continuous functional η applied to $\widehat{S}_{j\infty}$ with $\delta = 0$, we implement the following steps.

1. Estimate the model with data (Y_t, X_t'), $t = 1, 2, ..., T$, get parameter estimator $\widehat{\theta}_T$ and compute test statistics $\eta(\widehat{S}_{jT})$.

2. Simulate Y_t^* with $F_{\widehat{\theta}_T}\left(\cdot \mid \Omega_t^*\right)$ recursively for $t = 1, 2, ..., T$, where the bootstrap information set is $\Omega_t^* = \left(X_t, Y_{t-1}^*, X_{t-1}, Y_{t-2}^*, X_{t-2}, ...\right)$.

3. Estimate the model with simulated data Y_t^*, get $\widehat{\theta}_T^*$ using the same method as for $\widehat{\theta}_T$, get bootstrapped test statistics $\eta\left(\widehat{S}_{jT}^*\right)$.

4. Repeat 2-3 B times, compute the percentiles of the empirical distribution of the B bootstrapped test statistics.

5. Reject H_0 if $\eta\left(\widehat{S}_{jT}\right)$ is greater than the $(1 - \alpha)$th percentile of the empirical distribution of the B bootstrapped test statistics denoted by $\widehat{c}_{jB}^*\left(\widehat{\theta}_T\right)$.

To analyze the properties of our parametric bootstrap, we need to assume that the same conditions on the estimation method hold for both for original and resampled data. More formally, we have

Assumption 6. (A) The conditional distribution of Y_t conditional on Ω_t coincides with the conditional distribution of Y_t conditional on $\Omega_t \cup \{X_k'\}_{k=t+1}^T$.

(B) Suppose that the sample is generated by F_{θ_T}, for some nonrandom sequence θ_T converging to θ_0, i.e. we have a triangular array of random variables $\{Y_{Tt} : t = 1, 2, \ldots, T\}$ with (T, t) element generated by $F_{\theta_T}(\cdot \mid \Omega_{Tt})$, where $\Omega_{Tt} = \{X_t, Y_{Tt-1}, X_{t-1}, Y_{Tt-2}, X_{t-2}, \ldots\}$. Then the estimator $\widehat{\theta}_T$ of θ_T admits an asymptotic linear expansion as in Assumption 4. Moreover, assume that under the alternative H_1, there exists some $\theta_1 \in \Theta$ so that $\theta_1 = \text{plim}_{T\to\infty} \widehat{\theta}_T$.

This assumption insures that by simulating from the conditional distribution F_{θ_T} we obtain the correct joint distribution of S_{jT} and $T^{1/2}\left(\widehat{\theta}_T - \theta_T\right)$ in parallel to those required in Theorems 1-2. Assumption 6 (A) says that Y_t and future X_t are independent conditionally on past information, i.e. that there is no direct feedback effect. For example, in a latent variable form of the ordered probit model, this assumption translates to strict exogeneity, i.e. that innovations are independent of future X_t. Dependence between Y_t and future X_t is still allowed through serial dependence in X_t and Y_t. Assumption 6 (B) is similar to Condition (5.5) in Burke et al. (1979), Assumption (A1) in Stute et al. (1993) and Assumption E2 in Andrews (1997), and introduces a triangular array version of the expansion and central limit theorem for parameter estimates, see also the discussion in Section 4.1 in Andrews (1997).

We obtain the following result.

Theorem 4. *Suppose that in addition to conditions of Theorem 2, Assumption 6 holds. Under H_{1T}, as $B, T \to \infty$,*

$$\eta\left(\widehat{S}_{jT}^*\right) \to_d \eta\left(\widehat{S}_{j\infty}\right), \qquad j = 1, 2,$$

in probability, so $\widehat{c}_{jB}^*\left(\widehat{\theta}_T\right) \to_p c_j(\theta_0)$, and therefore, under H_0, $\Pr\left(\eta\left(\widehat{S}_{jT}\right) > \widetilde{c}_{jB}^*\left(\widehat{\theta}_T\right)\right) \to \alpha$. Suppose also that the conditions of Theorem 2 hold for any $\theta_0 \in \Theta$. Under H_1, as $B, T \to \infty$, $\widehat{c}_{jB}^*\left(\widehat{\theta}_T\right) = O_p(1)$.

This theorem shows that the bootstrap test statistic has the same limit distribution as the original one under local alternatives, so that under the null we get the right asymptotic size using bootstrap estimated critical values and that under local alternatives we get non

trivial power when the drifts of the stochastic processes $\widehat{S}_{1T}$ and $\widehat{S}_{2T}$ are non negligible. Similarly, under fixed alternatives we are able to get a bootstrap consistent test when the asymptotic test is consistent itself, i.e. $\lim_{T\to\infty} \Pr\left(\eta\left(\widehat{S}_{jT}\right) > \widehat{c}^*_{jB}\left(\widehat{\theta}_T\right)\right) = 1$ if $\eta\left(\widehat{S}_{jT}\right)$ diverges asymptotically.

6   APPLICATION AND SIMULATIONS

In this section we use a Monte Carlo simulation exercise to investigate the finite sample properties of the tests proposed in this paper. We take as reference the dynamic ordered discrete choice models investigated in Basu and de Jong (2007) for the modeling of the monetary policy conducted by the Federal Reserve (FED). The dependent variable uses the following codification of the changes in the reference interest rate in US, the federal funds rate i_t,

$$Y_t = \begin{cases} 1 & \text{if} & \Delta i_t < -0.25 \\ 2 & \text{if} & -0.25 \leq \Delta i_t < 0 \\ 3 & \text{if} & 0 \leq \Delta i_t < 0.25 \\ 4 & \text{if} & \Delta i_t \geq 0.25. \end{cases}$$

Data is monthly and spans January 1990 to December 2006, leading to $T = 204$ complete observations. The explanatory variables that Basu and de Jong (2007) used to explain the decisions of the FED on Δi_t are the current value and 4 lags of inflation (inf), the current value and a lag of four different measures of output gap (out) and a series of dummies that describe the decision of the FED in the previous period, $dum1_t = I(\Delta i_{t-1} < 0)$, $dum2_t = I(\Delta i_{t-1} > 0)$, $dum3_t = I(\Delta i_{t-1} < -0.25)$, $dum4_t = I(\Delta i_{t-1} > 0.25)$. Instead of these four dummies, we implement an AR(1), 'dynamic' version with one lag of the discrete Y_t as explanatory variable (and a version without lags that we refer to as 'static' to serve as a benchmark to the inclusion of lagged endogenous variables in Ω_t). We consider both the Logit and Probit versions of the models. We fit four versions of the basic model based on different definitions of the output gap and conditional on the series of inflation and output gap and on the parameter estimates obtained, we simulate

21

Table 1: Scenarios for Monte Carlo simulations.	
Scenario	Null and Alternative
Size 1	H_0 : static probit
Size 2	H_0 : static logit
Power 1	H_0 : static probit vs H_1 : static logit
Power 2	H_0 : static probit vs H_1 : dynamic probit
Power 3	H_0 : static probit vs H_1 : dynamic logit

series Y_t and conduct our tests on these (see Monte Carlo scenarios in Table 1).

The four choices of output gap lead to Models I-IV. The output gap is the percentage deviation of the actual from the potential output, which is interpolated to obtain a series of monthly frequency by replicating the GDP observation for any quarter to all the months in that quarter. Then two different measures of potential output are used: the potential output series provided by the Congressional Budget Office and a potential output series constructed in a real-time setting using the HP filter, leading to Models I and II. Apart from output gap, other measures of economic activity are used, such as unemployment rate and capacity utilization, leading to Models III and IV. Data sources are described in Basu and de Jong (2007).

We compare the performance of our tests with an alternative test which is also omnibus and does not require smoothing (and choice of smoothing parameters). Two general approaches can be adapted to our setup: the test of the Generalized Linear Model (GLM) of Stute and Zhu (2002) and the Conditional Kolmogorov test of Andrews (1997), as discussed in Mora and Moro-Egido (2007). The first one is a test based on a marked empirical process for testing the null $H_0' : \quad \mathrm{E}\left[Y \mid \widetilde{X} = x\right] = m_{\widetilde{\beta}_{01}}\left(x'\widetilde{\beta}_{02}\right)$, where $m_{\widetilde{\beta}_{01}}(\cdot)$ is a parametric link function and $\widetilde{\beta}_{01}, \widetilde{\beta}_{02}$ are finite dimensional parameters. In the cases where Y takes only two values $\{0, 1\}$, the conditional mean coincides with the conditional probability of $Y = 1$ and the null is similar to our H_0 if we were considering an i.i.d setup.

To test $Y_t \mid \widetilde{X}_t \sim P_{\widetilde{\beta}_{01}}\left(\cdot \mid \widetilde{X}_t'\widetilde{\beta}_{20}\right)$ define the process

$$Z_T(y) := \frac{1}{T^{1/2}} \sum_{t=1}^{T} \mathbb{1}\left\{\widetilde{X}_t'\widetilde{\beta}_{20} \leq y\right\} \left[Y_t - P_{\widetilde{\beta}_{01}}\left(Y_t = 1 \mid \widetilde{X}_t'\widetilde{\beta}_{20}\right)\right], \qquad y \in \mathbb{R}.$$

The second test by Andrews is obtained by substituting $\mathbb{1}\left\{\widetilde{X}_t'\widetilde{\beta} \leq y\right\}$ with $\mathbb{1}\left\{\widetilde{X}_t \leq \widetilde{x}\right\}$ (where $\widetilde{x}$ is a real vector of dimension of $\widetilde{X}_t$) in Z_T, but since it always underperforms according to simulations of Mora and Moro-Egido (2007), it is not considered here. If Y takes values $\{1, \ldots, K\}$, Mora and Moro-Egido (2007) substitute testing H_0 by K tests of the hypotheses $Y_{jt} \mid \widetilde{X}_t \sim P_{j,\widetilde{\beta}_{01}}\left(Y_t \mid \widetilde{X}_t'\widetilde{\beta}_{20}\right)$, with corresponding processes $Z_{j,T}$, where $Y_{jt} = \mathbb{1}\{Y_t = j\}$ and $j = 1, 2, \ldots, K$, then the resulting pooled test statistics are

$$\eta_Z^{CvM} = T^{-1} \sum_{j=1}^{K} \sum_{t=1}^{T} Z_{j,T}\left(\widetilde{X}_t'\widetilde{\beta}_{20}\right)^2$$

and

$$\eta_Z^{KS} = T^{-1} \max_{j=1,\ldots,K} \sum_{t=1}^{T} Z_{j,T}\left(\widetilde{X}_t'\widetilde{\beta}_{20}\right)^2,$$

which we call the CvM and KS tests respectively. To apply these tests to our model, let $\widetilde{X}_t = \left(X_t', Y_{t-1}, 1\right)'$ and $\widetilde{\beta} = \left(\beta', \rho, -\tau_1\right)'$ and take the corresponding link functions.

We analyze tests based on S_{1T}, $R_{1T,M}$, R_{1T} and S_{2T}, $R_{2T,M}$, R_{2T} and Z_T. In all cases we use Kolmogorov-Smirnov (KS) and Cramer-von Mises (CvM) measures. We only consider feasible bootstrap versions of tests based on $\widehat{S}_{1T}$, $\widehat{R}_{1T,M}$, etc, where we replace θ_0 by root-T consistent estimates $\widehat{\theta}_T$, the ML estimator in our case. We are not aware of any theoretical results for bootstrap assisted tests based on $\widehat{Z}_T$ in our setup, although Mora and Moro-Egido (2007) provide some simulations.

Parameter estimates for real data are reported in Tables 2 and 3. The main question is whether the static Probit or Logit models are appropriate for changes in the interest rates, and we check this with our tests. The p-values in Tables 4 and 5 say that all these models are rejected even at the 1% significance level by biparameter nonrandomized transform based tests. Note that single parameter static tests (e.g. $\widehat{R}_{1T}$, $\widehat{S}_{1T}$) cannot reject any proposed model with the sole exception of $\widehat{S}_{1T}$ which rejects at 5% Model II with Cramer – von Misses test statistics.

Table 2: ML estimates and standard errors of Models I-IV with static and dynamic specifications and Probit link function applied to the real US data, $T = 204$.

	I-static	I-dynamic	II-static	II-dynamic	III-static	III-dynamic	IV-static	IV-dynamic
τ_1	-4.81	-2.07	-3.31	-1.05	-3.15	-1.17	-3.41	-1.48
	(0.51)	(0.66)	(0.35)	(0.47)	(0.36)	(0.48)	(0.37)	(0.50)
τ_2	-4.05	-1.14	-2.64	-0.19	-2.34	-0.20	-2.57	-0.50
	(0.47)	(0.64)	(0.31)	(0.46)	(0.32)	(0.47)	(0.32)	(0.48)
τ_3	-1.72	1.66	-0.39	2.60	0.09	2.62	-0.11	2.29
	(0.40)	(0.63)	(0.26)	(0.48)	(0.28)	(0.48)	(0.27)	(0.49)
inf	-1.39	-1.36	-1.51	-1.60	-1.83	-1.82	-1.70	-1.70
	(0.68)	(0.72)	(0.67)	(0.71)	(0.69)	(0.73)	(0.69)	(0.73)
inf_{-1}	1.86	2.90	1.94	3.05	2.05	3.07	2.14	3.01
	(0.99)	(1.06)	(0.98)	(1.06)	(1.00)	(1.07)	(1.01)	(1.07)
inf_{-2}	-1.30	-2.81	-1.27	-2.80	-1.60	-2.92	-2.12	-3.11
	(0.98)	(1.07)	(0.97)	(1.06)	(0.99)	(1.07)	(1.02)	(1.09)
inf_{-3}	1.39	2.44	1.60	2.74	1.79	2.79	1.27	2.33
	(0.99)	(1.06)	(0.98)	(1.06)	(1.00)	(1.08)	(1.03)	(1.09)
inf_{-4}	0.43	-0.53	-0.23	-1.05	-0.00	-0.85	0.88	-0.20
	(0.68)	(0.73)	(0.66)	(0.71)	(0.67)	(0.73)	(0.71)	(0.76)
out	-1.02	-1.02	0.36	0.40	3.35	2.54	-0.98	-0.62
	(0.30)	(0.33)	(0.59)	(0.63)	(0.68)	(0.74)	(0.22)	(0.23)
out_{-1}	0.81	0.90	0.84	0.65	2.48	0.95	-1.03	-0.65
	(0.29)	(0.32)	(0.59)	(0.64)	(0.67)	(0.73)	(0.22)	(0.23)
Y_{-1}	—	-1.08	—	-1.12	—	-1.03	—	-0.94
		(0.15)		(0.15)		(0.16)		(0.16)

Table 3: ML estimates and standard errors of Models I-IV with static and dynamic specifications and Logit link function applied to the real US data, $T = 204$.

	I-static	I-dynamic	II-static	II-dynamic	III-static	III-dynamic	IV-static	IV-dynamic
τ_1	-8.46	-3.77	-6.01	-2.12	-5.61	-2.15	-6.15	-2.82
	(0.98)	(1.20)	(0.68)	(0.83)	(0.69)	(0.85)	(0.72)	(0.89)
τ_2	-7.03	-1.96	-4.71	-0.46	-4.12	-0.31	-4.56	-0.90
	(0.90)	(1.17)	(0.60)	(0.81)	(0.59)	(0.83)	(0.61)	(0.86)
τ_3	-3.00	3.02	-0.85	4.52	0.07	4.60	-0.24	4.04
	(0.72)	(1.12)	(0.47)	(0.84)	(0.49)	(0.86)	(0.49)	(0.87)
inf	-2.44	-2.29	-2.53	-2.89	-3.17	-3.28	-2.81	-3.06
	(1.21)	(1.30)	(1.21)	(1.29)	(1.21)	(1.32)	(1.22)	(1.32)
inf_{-1}	3.28	4.95	3.22	5.46	3.59	5.43	3.41	5.31
	(1.78)	(1.92)	(1.77)	(1.92)	(1.76)	(1.93)	(1.82)	(1.95)
inf_{-2}	-2.48	-5.02	-2.17	-5.22	-2.97	-5.21	-3.52	-5.40
	(1.74)	(1.95)	(1.73)	(1.94)	(1.76)	(1.95)	(1.86)	(1.99)
inf_{-3}	2.42	4.36	2.61	5.20	2.94	5.11	1.65	4.02
	(1.75)	(1.92)	(1.75)	(1.93)	(1.77)	(1.95)	(1.86)	(1.99)
inf_{-4}	0.93	-0.87	-0.17	-1.88	0.32	-1.54	2.11	-0.28
	(1.20)	(1.32)	(1.18)	(1.28)	(1.19)	(1.30)	(1.27)	(1.36)
out	-1.78	-1.79	0.43	0.63	5.87	4.12	-1.83	-1.15
	(0.54)	(0.60)	(1.04)	(1.14)	(1.24)	(1.34)	(0.40)	(0.42)
out_{-1}	1.43	1.59	1.61	1.29	4.21	1.50	-1.88	-1.14
	(0.52)	(0.59)	(1.04)	(1.15)	(1.20)	(1.33)	(0.40)	(0.42)
Y_{-1}	—	-1.98	—	-2.04	—	-1.86	—	-1.71
		(0.28)		(0.27)		(0.28)		(0.28)

Table 4: P-values of Cramer – von Misses tests for static Probit and Logit link function applied to the real US data, $T = 204$.

	$\widehat{S}_{2T}$	$\widehat{R}_{2T,50}$	$\widehat{R}_{2T,25}$	$\widehat{R}_{2T}$	$\widehat{S}_{1T}$	$\widehat{R}_{1T,50}$	$\widehat{R}_{1T,25}$	$\widehat{R}_{1T}$	$\widehat{Z}_{T}$
			H_0 : static probit						
Model I	0.001	0.001	0.001	0.237	0.009	0.026	0.078	0.516	0.244
Model II	0.001	0.001	0.001	0.166	0.077	0.057	0.229	0.167	0.022
Model III	0.001	0.001	0.001	0.307	0.492	0.632	0.616	0.731	0.109
Model IV	0.001	0.002	0.002	0.496	0.721	0.509	0.582	0.668	0.268
			H_0 : static logit						
Model I	0.001	0.001	0.001	0.152	0.021	0.079	0.221	0.793	0.199
Model II	0.001	0.001	0.001	0.112	0.113	0.155	0.459	0.240	0.032
Model III	0.001	0.001	0.001	0.360	0.314	0.493	0.541	0.745	0.171
Model IV	0.001	0.001	0.001	0.448	0.890	0.804	0.899	0.634	0.272

Table 5: P-values of Kolmogorov – Smirnov tests for static Probit and Logit link function applied to the real US data, $T = 204$.

	$\widehat{S}_{2T}$	$\widehat{R}_{2T,50}$	$\widehat{R}_{2T,25}$	$\widehat{R}_{2T}$	$\widehat{S}_{1T}$	$\widehat{R}_{1T,50}$	$\widehat{R}_{1T,25}$	$\widehat{R}_{1T}$	$\widehat{Z}_{T}$
			H_0 : static probit						
Model I	0.003	0.002	0.002	0.082	0.047	0.193	0.372	0.354	0.392
Model II	0.001	0.001	0.002	0.586	0.351	0.426	0.626	0.450	0.107
Model III	0.001	0.001	0.001	0.155	0.454	0.435	0.244	0.742	0.124
Model IV	0.001	0.002	0.002	0.799	0.936	0.913	0.801	0.355	0.230
			H_0 : static logit						
Model I	0.001	0.001	0.001	0.133	0.010	0.050	0.212	0.684	0.220
Model II	0.001	0.001	0.001	0.354	0.114	0.201	0.319	0.416	0.058
Model III	0.001	0.001	0.001	0.149	0.511	0.472	0.350	0.642	0.173
Model IV	0.002	0.002	0.001	0.769	0.975	0.968	0.867	0.411	0.207

Table 6: Simulated size/power rates for the nominal 5% level of Cramer – von Misses tests of Models I-IV with static and dynamic specifications applied to simulated data, $T = 100$.

	$\widehat{S}_{2T}$	$\widehat{R}_{2T,50}$	$\widehat{R}_{2T,25}$	$\widehat{R}_{2T}$	$\widehat{S}_{1T}$	$\widehat{R}_{1T,50}$	$\widehat{R}_{1T,25}$	$\widehat{R}_{1T}$	$\widehat{Z}_T$
Size 1 H_0 : static probit									
Model I	5.5	6.0	5.5	4.5	6.6	6.3	5.7	5.4	7.8
Model II	5.3	6.7	5.0	5.5	6.3	5.2	4.2	3.3	6.5
Model III	7.7	7.0	6.5	5.4	6.0	3.7	3.3	4.5	6.4
Model IV	5.2	6.7	5.6	3.9	5.1	4.6	4.9	2.8	6.4
Size 2 H_0 : static logit									
Model I	6.5	6.5	4.9	4.1	7.2	5.6	6.0	4.4	7.2
Model II	5.6	6.7	7.6	4.0	4.6	5.3	4.6	4.8	5.6
Model III	7.3	9.0	6.4	3.3	6.4	7.8	5.2	3.3	8.5
Model IV	6.6	6.3	5.0	4.5	6.5	4.6	4.7	4.7	9.1
Power 1 H_0 : static probit vs H_1 : static logit									
Model I	8.5	7.7	6.6	4.9	8.4	6.5	6.0	3.6	7.1
Model II	5.1	5.0	4.4	4.0	6.4	6.9	5.3	4.0	8.7
Model III	9.1	9.4	7.9	4.7	9.0	8.3	7.7	4.6	8.2
Model IV	6.3	6.2	5.3	4.5	10.2	8.6	7.5	3.8	8.3
Power 2 H_0 : static probit vs H_1 : dynamic probit									
Model I	89.2	85.2	82.7	25.7	13.2	12.0	11.7	4.6	18.4
Model II	92.8	92.3	91.1	34.2	10.5	8.1	8.8	3.0	17.2
Model III	90.7	88.4	86.1	22.5	9.2	9.8	8.5	5.0	9.4
Model IV	88.1	84.1	83.0	27.7	10.3	7.8	7.4	4.4	12.5
Power 3 H_0 : static probit vs H_1 : dynamic logit									
Model I	90.1	89.3	86.0	22.9	12.1	10.0	8.5	5.0	12.6
Model II	94.2	93.0	90.6	29.8	9.6	9.1	7.1	3.9	14.6
Model III	93.5	91.9	90.9	30.3	10.0	8.0	7.8	4.4	10.9
Model IV	91.1	88.4	85.9	26.0	11.1	12.3	11.4	4.7	14.7

Table 7: Simulated size/power rates for the nominal 5% level of Kolmogorov – Smirnov tests of Models I-IV with static and dynamic specifications applied to simulated data, $T = 100$.

	$\widehat{S}_{2T}$	$\widehat{R}_{2T,50}$	$\widehat{R}_{2T,25}$	$\widehat{R}_{2T}$	$\widehat{S}_{1T}$	$\widehat{R}_{1T,50}$	$\widehat{R}_{1T,25}$	$\widehat{R}_{1T}$	$\widehat{Z}_T$
Size 1 H_0 : static probit									
Model I	5.1	6.4	5.2	3.9	7.8	6.3	6.8	4.9	7.9
Model II	5.5	6.5	3.9	4.9	5.9	5.1	4.1	4.8	6.2
Model III	7.7	7.8	6.8	5.1	6.1	7.0	6.0	4.9	5.6
Model IV	6.5	5.4	5.3	3.4	5.3	5.3	4.8	3.6	7.2
Size 2 H_0 : static logit									
Model I	7.0	6.4	6.1	5.4	9.1	6.4	6.3	3.7	6.7
Model II	4.7	4.9	4.6	3.5	5.6	3.8	4.0	4.8	5.8
Model III	8.3	8.3	6.7	3.2	6.2	5.7	3.5	4.0	10.0
Model IV	6.2	6.5	5.1	4.7	6.6	5.8	5.3	4.0	8.1
Power 1 H_0 : static probit vs H_1 : static logit									
Model I	7.0	6.2	5.4	3.7	5.2	3.3	3.9	3.2	7.7
Model II	4.3	3.8	4.5	3.7	4.1	3.9	3.6	3.9	8.9
Model III	10.2	7.3	7.1	3.9	7.1	5.7	5.7	4.5	9.2
Model IV	5.6	6.6	4.3	3.2	6.4	5.1	6.2	3.4	6.8
Power 2 H_0 : static probit vs H_1 : dynamic probit									
Model I	82.8	79.0	74.5	13.6	10.3	9.1	7.1	3.5	16.9
Model II	87.9	85.5	83.3	17.7	12.1	11.2	9.3	3.3	14.0
Model III	85.7	83.2	79.4	13.8	7.1	6.4	7.2	3.9	9.2
Model IV	81.7	78.5	74.6	13.8	7.7	7.8	6.7	4.9	11.3
Power 3 H_0 : static probit vs H_1 : dynamic logit									
Model I	86.2	82.7	79.0	14.2	7.7	4.9	3.8	4.2	11.8
Model II	90.0	86.2	82.2	15.9	9.3	7.9	8.1	4.1	14.2
Model III	89.0	86.4	83.7	15.9	5.6	5.1	4.4	4.6	10.5
Model IV	87.5	83.8	79.3	16.1	9.4	7.5	7.7	5.9	12.9

Table 8: Simulated size/power rates for the nominal 5% level of Cramer – von Misses tests of Models I-IV with static and dynamic specifications applied to simulated data, $T = 200$.

	$\widehat{S}_{2T}$	$\widehat{R}_{2T,50}$	$\widehat{R}_{2T,25}$	$\widehat{R}_{2T}$	$\widehat{S}_{1T}$	$\widehat{R}_{1T,50}$	$\widehat{R}_{1T,25}$	$\widehat{R}_{1T}$	$\widehat{Z}_{T}$
			Size 1 H_0 : static probit						
Model I	4.0	5.4	5.7	6.2	4.2	4.6	4.9	5.8	5.2
Model II	4.5	4.4	3.5	2.4	6.3	4.7	5.9	4.4	7.0
Model III	4.6	4.4	3.4	4.2	5.4	5.5	5.2	3.3	5.4
Model IV	5.3	6.1	6.3	4.4	4.8	4.6	7.0	4.9	6.9
			Size 2 H_0 : static logit						
Model I	7.2	8.2	6.7	5.8	5.8	6.7	6.4	3.8	5.2
Model II	5.4	6.4	6.1	4.7	4.8	5.6	5.3	5.2	6.1
Model III	5.3	5.2	3.9	4.0	5.6	5.8	6.7	4.4	6.9
Model IV	5.4	6.8	5.0	4.0	5.6	5.2	5.0	4.1	8.3
		Power 1 H_0 : static probit vs H_1 : static logit							
Model I	7.2	8.2	6.6	6.9	10.9	10.3	10.9	6.9	9.2
Model II	4.5	4.9	5.6	6.3	7.5	6.4	7.3	6.7	6.5
Model III	6.0	5.2	6.1	5.9	6.9	6.8	7.9	6.6	7.0
Model IV	6.5	6.6	6.6	4.3	7.1	6.1	7.5	5.8	5.4
		Power 2 H_0 : static probit vs H_1 : dynamic probit							
Model I	98.5	97.3	95.2	33.2	13.6	11.6	9.8	7.5	16.2
Model II	99.5	99.3	98.5	41.5	16.0	14.8	12.6	7.1	18.2
Model III	98.5	97.0	95.9	30.7	13.0	11.8	9.8	7.9	13.9
Model IV	95.8	93.6	91.6	22.9	10.2	9.5	7.6	5.3	13.7
		Power 3 H_0 : static probit vs H_1 : dynamic logit							
Model I	98.6	97.5	95.6	34.5	15.2	14.0	14.0	5.4	16.7
Model II	99.5	98.9	98.6	39.1	16.9	16.1	13.7	7.4	20.8
Model III	98.8	98.1	96.4	31.2	14.8	13.7	11.6	6.6	17.9
Model IV	95.8	94.2	91.8	23.9	11.4	10.1	8.6	5.3	11.4

Table 9: Simulated size/power rates for the nominal 5% level of Kolmogorov – Smirnov tests of Models I-IV with static and dynamic specifications applied to simulated data, $T = 200$.

	$\widehat{S}_{2T}$	$\widehat{R}_{2T,50}$	$\widehat{R}_{2T,25}$	$\widehat{R}_{2T}$	$\widehat{S}_{1T}$	$\widehat{R}_{1T,50}$	$\widehat{R}_{1T,25}$	$\widehat{R}_{1T}$	$\widehat{Z}_{T}$
			Size 1 H_0 : static probit						
Model I	5.1	5.0	6.0	4.8	4.5	5.9	3.9	5.1	5.5
Model II	3.7	3.9	3.9	2.9	6.3	5.6	6.4	4.6	5.7
Model III	4.5	5.2	4.3	3.9	4.2	4.5	4.9	4.3	5.3
Model IV	5.0	6.4	7.0	4.6	4.6	4.8	6.4	6.9	6.8
			Size 2 H_0 : static logit						
Model I	5.7	5.7	6.3	4.9	6.3	5.9	6.3	3.5	4.8
Model II	5.5	5.1	5.9	3.4	5.4	4.6	5.9	4.9	5.3
Model III	3.6	5.4	4.6	3.6	6.4	4.3	5.2	5.3	7.8
Model IV	6.4	7.3	5.6	4.7	6.6	6.4	4.7	4.5	8.5
		Power 1 H_0 : static probit vs H_1 : static logit							
Model I	6.3	6.5	4.2	6.6	7.2	5.9	5.0	6.7	8.7
Model II	4.6	5.0	6.4	6.3	5.2	4.9	5.7	6.5	6.1
Model III	5.0	6.2	5.7	5.2	3.7	4.1	5.0	6.3	7.1
Model IV	5.7	7.0	5.6	4.5	5.4	3.4	4.6	6.1	5.1
		Power 2 H_0 : static probit vs H_1 : dynamic probit							
Model I	94.3	92.0	86.5	22.8	11.4	10.6	9.7	5.9	14.0
Model II	98.1	96.5	94.4	26.3	15.5	13.1	13.5	7.3	13.1
Model III	94.3	91.0	87.9	21.0	14.7	13.1	12.7	7.3	13.8
Model IV	90.5	85.0	82.0	17.9	11.0	9.5	9.4	6.3	11.4
		Power 3 H_0 : static probit vs H_1 : dynamic logit							
Model I	97.1	93.8	91.8	24.7	12.4	12.8	11.1	5.5	13.4
Model II	98.9	97.6	96.5	29.5	16.9	17.1	14.6	7.4	16.9
Model III	96.1	93.8	91.8	26.0	14.6	14.4	11.9	8.0	15.4
Model IV	93.0	89.2	86.9	14.1	13.0	12.5	8.5	5.8	10.4

To study the reliability of these results we conduct a Monte Carlo experiment using the estimated models with the real data as data generating processes and obtain the simulations for the discrete response conditional on the covariates time series. In Tables 6 and 7 we provide the empirical size and power results of our tests across simulations for sample size $T = 100$ and static Probit and Logit and output gap choices (Models I to IV). To speed up the simulation procedure, we use the warp bootstrap algorithm of Giacomini, Politis and White (2013). We see that all bootstrap tests provide reasonable size accuracy, tests based on single parameter empirical processes underrejecting slightly, while ones based on bivariate processes tend to overreject moderately. Kolmogorov-Smirnov and Cramer-von Mises tests perform similarly in all cases, and the choice of the output gap series does not make large differences either, nor does the introduction of lagged endogenous (discrete) variables in the information set.

The power of the tests for the static Probit model is analyzed against three different alternatives: static Logit, dynamic Probit and dynamic Logit. We see that the tests without randomization, $\widehat{S}_{1T}$ and $\widehat{S}_{2T}$ always perform better than random continuous processes $\widehat{R}_{1T,M}$ and $\widehat{R}_{2T,M}$, which in turn dominate $\widehat{R}_{1T}$ and $\widehat{R}_{2T}$, thus confirming our theoretical findings. When we compare Probit and Logit specifications while letting the dynamic aspect of the model be well specified, static in both cases, we observe that with this sample size and these specifications, it is almost impossible to distinguish Probit from Logit models. The power against a dynamic Probit and Logit alternatives is very high. Since the nature of misspecification is dynamic, once again bivariate processes should have more power compared to single parameter counterparts, as it is confirmed in our simulation results. It can also be observed that for these alternatives, the Cramer-von Mises criterium provides more power than Kolmogorov-Smirnov tests. As for alternative tests based on $\widehat{Z}_T$, they have power comparable to $\widehat{S}_{1T}$, sometimes slightly better, and are always outperformed by any bivariate test. This is not surprising, since $\widehat{Z}_T$ has more structure, i.e. it assumes a single-index model for covariates, but averages across points, thus suffering the same problems as other single parameter tests considered here.

In Tables 8 and 9 we provide the empirical size and power results of our tests for the

larger sample size $T = 200$. Here the size properties are similar, while power rejections rates are noticeably higher for the dynamic alternatives.

7 CONCLUSIONS

In this paper we have proposed new specification tests for the conditional distribution of discrete data with possibly infinite support. The new tests are functionals of empirical processes based on a nonrandomized transform that solves the implementation problem of the usual PIT for discrete distributions and achieves consistency against a wide class of alternatives. We show the validity of a bootstrap algorithm for approximating the null distribution of the test statistics, which are model and parameter dependent. In our simulation study, we show that our method compares favorably in many relevant situations with other methods available in the literature and have illustrated the new method in a small application.

8 APPENDIX

8.1 Properties of the nonrandomized transform

In this section we derive the basic properties of the nonrandomized transform, which are required prior to proving the weak convergence results for our empirical process. Without loss of generality and in order to make the exposition more transparent, we omit subscripts t, θ_0 and conditioning set Ω_t, and use shortcuts $I_F(Y, u) = I_{t,\theta_0}(Y_t, u)$ and $I_{F,M}(Y, u) = I_{t,\theta_0,M}(Y_t, u)$.

For $F \in \mathcal{M}$, $F(F^{-1}(u)) \geq u > F(F^{-1}(u) - 1)$ and equality holds iff $u = F(k)$ for some integer k. For a random variable $Y \sim G \in \mathcal{M}$ we find $\Pr_G(F(Y) < u) = G(F^{-1}(u) - 1)$ and $g(F^{-1}(u)) := \Pr_G(Y = F^{-1}(u)) = G(F^{-1}(u)) - G(F^{-1}(u) - 1)$. For $G = F$, we have that $\Pr_F(F(Y) < u) = F(F^{-1}(u) - 1) < u$, i.e. $F(Y)$ is not uniform and the expectation of the indicator function $I(F(Y) < u)$ is never u as it is for

continuous F.

The nonrandomized transform can be written as

$$I_F(Y, u) = (1 - \delta_F(u)) \mathbb{1}\{Y = F^{-1}(u)\} + \mathbb{1}\{Y < F^{-1}(u)\},$$

where

$$\delta_F(u) := \frac{F(F^{-1}(u)) - u}{f(F^{-1}(u))}.$$

Note that $\delta_F(u) \in [0, 1)$. We see that $I_F(Y, u)$ is a piecewise linear (continuous) function increasing in u. Let

$$\delta_F(u, v) := (\delta_F(u \vee v) - \delta_F(u)\delta_F(v)) f(F^{-1}(u \wedge v))$$

$$\times \mathbb{1}\{F^{-1}(u) = F^{-1}(v)\} \in [0, u \wedge v \wedge f(F^{-1}(u \wedge v))],$$

$$d(G, F, u, v) := d(G, F, u \wedge v)$$

$$- (\delta_F(u \vee v) - \delta_F(u)\delta_F(v)) \mathbb{1}\{F^{-1}(u) = F^{-1}(v)\}$$

$$\times (g(F^{-1}(u)) - f(F^{-1}(u))),$$

In Table 10 and Lemma A we list the properties of this transform.

Lemma A. *For $0 \le v \le u \le 1$ and $F, G, H \in \mathcal{M}$,*

(i) $\mathrm{E}_G[I_F(Y, u)] = u + d(G, F, u)$, *where* $\mathrm{E}_G[\cdot] = \int(\cdot)dG$ *and* $d(G, F, u) \in [-u, 1-u]$. *When $G = F$, the expectation is u.*

(ii) $I_F(Y, u) I_F(Y, v) = I_F(Y, u \wedge v) -$
$(\delta_F(u \vee v) - \delta_F(u)\delta_F(v)) \times \mathbb{1}\{Y = F^{-1}(u) = F^{-1}(v)\}.$

(iii) $\mathrm{E}_G[I_F(Y, u) I_F(Y, v)] = u \wedge v - \delta_F(u, v) + d(G, F, u, v)$.

(iv) $|I_F(Y, u) - I_H(Y, u)| \le 1 \wedge \frac{|F(Y) - H(Y)| \vee |F(Y-1) - H(Y-1)|}{f(Y) \vee h(Y)}$
Moreover, $\mathrm{E}_F\left[|I_F(Y, u) - I_H(Y, u)|^2\right] \le 9 \sup_k |F(k) - H(k)|$.

(v) $|I_F(Y, u) - u - I_F(Y, v) + v| \le |u - v| \vee (1 - f(Y))$ *and* $|I_F(Y, u) - u - I_F(Y, v) + v| = |u - v|$ *if* $u, v \le F(Y - 1)$ *or* $u, v \ge F(Y)$.
Moreover, $\mathrm{E}_F\left[\sup_{u,v \in \Psi(\varepsilon)} |I_F(Y, u) - u - I_F(Y, v) + v|^2\right] \le 4\varepsilon^2$, *for any interval* $\Psi(\varepsilon) \subset [0, 1]$ *of length ε^2.*

Table 10: Values of functionals of the new nonrandomized transform $I\left(\cdot,\cdot\right)$ for all possible values of Y relative to inverted cdfs at points u and v. For instance, $I_F\left(Y,u\right)-I_F\left(Y,v\right)=0$ if $Y<F^{-1}\left(u\right)$ and $Y<F^{-1}\left(v\right)$, while $I_F\left(Y,u\right)-I_F\left(Y,v\right)=-\delta_F\left(u\right)$ if $Y=F^{-1}\left(u\right)<F^{-1}\left(v\right)$.

	$Y<F^{-1}\left(u\right)$	$Y=F^{-1}\left(u\right)$	$Y>F^{-1}\left(u\right)$
	The value of $I_F\left(Y,u\right)$		
	1	$1-\delta_F\left(u\right)$	0
	The value of $\mathbb{1}\{I_F\left(Y,u\right)\leq v\}$		
$v=0$	0	0	1
$v\in(0,1)$	0	$\mathbb{1}\{1-\delta_F\left(u\right)\leq v\}$	1
$v=1$	1	1	1
	The value of $I_F\left(Y,u\right)-I_F\left(Y,v\right)$		
$Y<F^{-1}\left(v\right)$	0	$-\delta_F\left(u\right)$	-1
$Y=F^{-1}\left(v\right)$	$\delta_F\left(v\right)$	$\delta_F\left(v\right)-\delta_F\left(u\right)$	$-1+\delta_F\left(v\right)$
$Y>F^{-1}\left(v\right)$	1	$1-\delta_F\left(u\right)$	0
	The value of $I_F\left(Y,u\right)I_F\left(Y,v\right)$		
$Y<F^{-1}\left(v\right)$	1	$1-\delta_F\left(u\right)$	0
$Y=F^{-1}\left(v\right)$	$1-\delta_F\left(v\right)$	$\left(1-\delta_F\left(u\right)\right)\left(1-\delta_F\left(v\right)\right)$	0
$Y>F^{-1}\left(v\right)$	0	0	0
	The value of $I_F\left(Y,u\right)-I_H\left(Y,u\right)$		
$Y<H^{-1}\left(u\right)$	0	$-\delta_F\left(u\right)$	-1
$Y=H^{-1}\left(u\right)$	$\delta_H\left(u\right)$	$\delta_H\left(u\right)-\delta_F\left(u\right)$	$-1+\delta_H\left(u\right)$
$Y>H^{-1}\left(u\right)$	1	$1-\delta_F\left(u\right)$	0
	The value of $I_F\left(Y,u\right)I_H\left(Y,u\right)$		
$Y<H^{-1}\left(u\right)$	1	$1-\delta_F\left(u\right)$	0
$Y=H^{-1}\left(u\right)$	$1-\delta_H\left(u\right)$	$\left(1-\delta_F\left(u\right)\right)\left(1-\delta_H\left(u\right)\right)$	0
$Y>H^{-1}\left(u\right)$	0	0	0

(vi) $\mathrm{E}_{F_z}\left[\mathbb{1}\left\{F^{\dagger}\left(Y^{\dagger}\right)<u\right\}\right]=I_F\left(Y,u\right).$

(vii) $\mathrm{E}_{F_z}\left[I_{F,M}\left(Y,u\right)I_{F,M}\left(Y,v\right)\right]=\frac{1}{M}I_F\left(Y,u\wedge v\right)+\left(1-\frac{1}{M}\right)I_F\left(Y,u\right)I_F\left(Y,v\right).$

8.2 Functional weak convergence of discrete martingales

In this section we present Lindeberg-Feller-type sufficient conditions for functional weak convergence of discrete martingales. In general, to establish the weak convergence one needs to check tightness and finite-dimensional convergence. In case of martingales, both parts can be verified without imposing restrictive conditions. Here we state a result of Nishiyama (2000), which extends Theorem 2.11.9 of van der Vaart and Wellner (1996) to martingales, see also Theorem A.1 in Delgado and Escanciano (2007). Further details on notation and definitions can be found in books Van der Vaart and Wellner (1996) for empirical processes and row-independent triangular arrays and in Jacod and Shiryaev (2003) for finite-dimensional semimartingales. For every T, let $\left(\Omega^T,\mathcal{F}^T,\{\mathcal{F}_t^T\},P^T\right)$ be a discrete stochastic basis, where $\left(\Omega^T,\mathcal{F}^T,P^T\right)$ is a probability space equipped with a filtration $\left\{\mathcal{F}_t^T\right\}$. For nonempty set Ψ, let $\{\xi_t^T\}_{t=1,2,\dots}$ be a $\ell^\infty\left(\Psi\right)$-valued martingale difference array with respect to filtration $\mathcal{F}_t^T$, i.e. for every t, ξ_t^T maps Ω^T to $\ell^\infty\left(\Psi\right)$, the space of bounded, $\mathbb{R}$-valued functions on Ψ with sup-norm $\|\cdot\|=\|\cdot\|_\infty$ and for each $u\in\Psi$, $\xi_t^T(u)$ is a $\mathbb{R}$-valued martingale difference array: $\xi_t^T(u)$ is $\mathcal{F}_t^T$-measurable and $\mathrm{E}\left[\xi_t^T(u)\mid\mathcal{F}_t^T\right]=0$. We are interested in studying the weak convergence of discrete martingales $\sum_{t=1}^T\xi_t^T$. Denote a decreasing series of finite partitions (DFP) of Ψ as $\Pi=\{\Pi(\varepsilon)\}_{\varepsilon\in(0,1)\cap\mathbb{Q}}$, where $\Pi(\varepsilon)=\{\Psi(\varepsilon;k)\}_{1\leq k\leq N_\Pi(\varepsilon)}$ such that $\Psi=\bigcup_{k=1}^{N_\Pi(\varepsilon)}\Psi(\varepsilon;k)$, $N_\Pi(1)=1$ and $\lim_{\varepsilon\to 0}N_\Pi(\varepsilon)=\infty$ monotonically in ε. The ε-entropy of the DFP Π is $H_\Pi(\varepsilon)=\sqrt{\log N_\Pi(\varepsilon)}$. The quadratic Π-modulus of ξ_t^T is $\mathbb{R}_+\cup\{\infty\}$-valued process

$$\left\|\xi_t^T\right\|_{\Pi,T}=\sup_{\varepsilon\in(0,1)\cap\mathbb{Q}}\frac{1}{\varepsilon}\max_{1\leq k\leq N_\Pi(\varepsilon)}\sqrt{\sum_{t=1}^T\mathrm{E}\left[\sup_{u,v\in\Psi(\varepsilon;k)}\left|\xi_t^T(u)-\xi_t^T(v)\right|^2\mid\mathcal{F}_t^T\right]}. \tag{8}$$

Theorem A. *Let* $\{\xi_t^T\}_{t=1,2,\dots}$ *be a* $\ell^\infty\left(\Psi\right)$*-valued martingale difference array and*

N1) (conditional variance convergence) $\sum_{t=1}^T\mathrm{E}\left[\xi_t^T(u)\xi_t^T(v)\mid\mathcal{F}_t^T\right]\to_{P^T}V(u,v)$ *for every* $u,v\in\Psi;$

N2) (Lindeberg condition) $\sum_{t=1}^{T} \mathrm{E}\left[\left\|\xi_t^T\right\|^2 \mathbf{1}\left\{\left\|\xi_t^T\right\| > \varepsilon\right\} \mid \mathcal{F}_t^T\right] \to_{PT} 0$ *for every* $\varepsilon > 0$;

N3) (partitioning entropy condition) there exist a DFP Π *of* Ψ *such that* $\left\|\xi_t^T\right\|_{\Pi,T} = O_{PT}(1)$ *and* $\int_0^1 H_\Pi(\varepsilon)d\varepsilon < \infty$.

Then $\sum_{t=1}^{T} \xi_t^T \Rightarrow S$, *where* S *has normal marginals* $(S(v_1), S(v_2), \ldots, S(v_d))) \sim_d N(0, \Sigma)$ *with covariance* $\Sigma = \{V(v_i, v_j)\}_{ij}$.

8.3   Additional technical assumptions

To establish the asymptotic properties of the biparameter process S_{2T} we need the following assumption for uniform convergence of different empirical quantities.

Assumption A. Under H_{1T}, the following uniform limits to continuous functions exist

1. $\mathrm{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} \gamma_{t-1,\theta_0}(u_2, v_2) \gamma_{t,\theta_0}(u_1, v_1)$,

2. $\mathrm{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} I_{t-1,\theta_0}(v_2) \gamma_{t,\theta_0}(u_1, v_1)$,

3. $\mathrm{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} I_{t-1,\theta_0}(u_2) d\left(H_t(\cdot \mid \Omega_t), F_{t,\theta_0}(\cdot \mid \Omega_t), u_1\right)$,

4. $\mathrm{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} I_{t-1,\theta_0}(u_2) \mathrm{E}\left[I_{t,\theta_0}(u_1) \ell_t(Y_t, \Omega_t) \mid \Omega_t\right]$,

5. $\mathrm{plim}_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} I_{t-1,\theta_0}(u_2) \nabla\left(F_{t,\theta_0}(\cdot \mid \Omega_t), u_1\right)$.

As it is discussed in the text, these conditions restrict the dynamics of the data process such that some LLN holds, which is the case, e.g., for stationary and ergodic processes.

8.4   Proofs

Proof of Lemma A. **(i)** By definition of $I_F(Y, u)$, $\mathrm{E}_G[I_F(Y, u)] = (1-\delta_F(u))g(F^{-1}(u)) + G(F^{-1}(u)) - g(F^{-1}(u)) = d(G, F, u) - \delta_F(u)f(F^{-1}(u)) + F(F^{-1}(u)) = d(G, F, u) + u$. Similarly, by direct calculation we obtain (ii), (iii), (vi) and (vii). We now provide a detailed proof of (iv) and (v).

(iv) We prove a stronger result that for $G \in \mathcal{M}$, such that $\sup_k |F(k) - G(k)| \vee |H(k) - G(k)| \leq \sup_k |F(k) - H(k)|$ the expectation with respect to G is bounded: $\mathrm{E}_G \left[(I_F(Y, u) - I_H(Y, u))^2 \right] \leq 9 \sup_k |F(k) - H(k)|$. Then, the required bound is obtained by setting $G \equiv F$.

Since $|I_F(Y, u) - I_H(Y, u)|$ never exceeds 1, we have that $\mathrm{E}_G \left[(I_F(Y, u) - I_H(Y, u))^2 \right] \leq \mathrm{E}_G \left[|I_F(Y, u) - I_H(Y, u)| \right]$, therefore we bound the latter expectation.

Suppose that $F^{-1}(u) = H^{-1}(u)$. Then $I_F(Y, u) - I_H(Y, u) = \delta_H(u) - \delta_F(u)$ for $Y = F^{-1}(u)$, i.e. with probability $g(F^{-1}(u))$, and is zero for other Y. Therefore,

$$
\begin{aligned}
&\mathrm{E}_G \left[|I_F(Y, u) - I_H(Y, u)| \right] = |\delta_H(u) - \delta_F(u)| \, g(F^{-1}(u)) \\
&\leq \left| F(F^{-1}(u)) - H(F^{-1}(u)) \right| + \left| f(F^{-1}(u)) - g(F^{-1}(u)) \right| \delta_F(u) + \left| h(F^{-1}(u)) - g(F^{-1}(u)) \right| \delta_H(\\
&\leq \sup_k |F(k) - H(k)| + \sup_k |f(k) - g(k)| + \sup_k |h(k) - g(k)| \\
&\leq 5 \sup_k |F(k) - H(k)|,
\end{aligned}
$$

since $\delta_F(u), \delta_H(u) \in [0, 1)$ and $\sup_k |h(k) - g(k)| \leq 2 \sup_k |F(k) - G(k)|$.

Suppose that $F^{-1}(u) < H^{-1}(u)$. Note that $I_F(Y, u) - I_H(Y, u) = 0$ for $Y \notin [F^{-1}(u), H^{-1}(u)]$. We separately bound each term in

$$
\begin{aligned}
\mathrm{E}_G \left[|I_F(Y, u) - I_H(Y, u)| \right] =\ & \mathrm{E}_G \left[|I_F(Y, u) - I_H(Y, u)| \, \mathbb{1}\{Y = F^{-1}(u)\} \right] \\
& + \mathrm{E}_G \left[|I_F(Y, u) - I_H(Y, u)| \, \mathbb{1}\{Y = H^{-1}(u)\} \right] \\
& + \mathrm{E}_G \left[|I_F(Y, u) - I_H(Y, u)| \, \mathbb{1}\{F^{-1}(u) < Y < H^{-1}(u)\} \right].
\end{aligned}
$$

For $Y = F^{-1}(u)$, $I_F(Y, u) - I_H(Y, u) = -\delta_F(u)$. Then

$$
\begin{aligned}
&\mathrm{E}_G \left[(I_F(Y, u) - I_H(Y, u)) \, \mathbb{1}\{Y = F^{-1}(u)\} \right] = \left| I_F(F^{-1}(u), u) - I_H(F^{-1}(u), u) \right| g(F^{-1}(u)) \\
&= \delta_F(u) g(F^{-1}(u)) \\
&= F(F^{-1}(u)) - u + \delta_F(u) \left(g(F^{-1}(u)) - f(F^{-1}(u)) \right) \\
&\leq \sup_k |F(k) - H(k)| + \sup_k |f(k) - g(k)| \leq 3 \sup_k |F(k) - H(k)|,
\end{aligned}
$$

since $\delta_F(u) \in [0, 1)$ and for $u \in [H(F^{-1}(u)), F(F^{-1}(u))]$ we have that $F(F^{-1}(u)) - u \leq F(F^{-1}(u)) - H(F^{-1}(u))$.

For $Y = H^{-1}(u)$, $I_F(Y, u) - I_H(Y, u) = -1 + \delta_H(u)$.

Then

$$\mathrm{E}_G\left[|I_F(Y, u) - I_H(Y, u)| \, \mathbb{1}\{Y = H^{-1}(u)\}\right]$$

$$= \left|I_F\left(H^{-1}(u), u\right) - I_H\left(H^{-1}(u), u\right)\right| g(H^{-1}(u))$$

$$= (1 - \delta_H(u)) \, g\left(H^{-1}(u)\right)$$

$$= u - H\left(H^{-1}(u) - 1\right) + (1 - \delta_H(u))\left(g\left(H^{-1}(u)\right) - h\left(H^{-1}(u)\right)\right)$$

$$\leq \sup_k |F(k) - H(k)| + \sup_k |h(k) - g(k)| \leq 3\sup_k |F(k) - H(k)|,$$

since $\delta_H(u) \in [0, 1)$ and for $u \in [H(H^{-1}(u) - 1), F(H^{-1}(u) - 1)]$ we have that $u - H(H^{-1}(u) - 1) \leq F(H^{-1}(u) - 1) - H(H^{-1}(u) - 1)$.

For $F^{-1}(u) < Y < H^{-1}(u)$, $I_F(Y, u) - I_H(Y, u) = -1$. Then

$$\mathrm{E}_F |I_F(Y, u) - I_H(Y, u)| \, \mathbb{1}\{F^{-1}(u) < Y < H^{-1}(u)\}$$

$$= \sum_{k=F^{-1}(u)+1}^{H^{-1}(u)-1} g(k) = G(H^{-1}(u) - 1) - G(F^{-1}(u))$$

$$\leq F(H^{-1}(u) - 1) - F(F^{-1}(u)) + 2\sup_k |G(k) - F(k)|$$

$$\leq F(H^{-1}(u) - 1) - H(H^{-1}(u) - 1) + 2\sup_k |G(k) - F(k)| \leq 3\sup_k |F(k) - H(k)|$$

since $H(H^{-1}(u) - 1) < u < F(F^{-1}(u)) < F(H^{-1}(u) - 1)$.

Adding everything together, get that $\mathrm{E}_G\left[|I_F(Y, u) - I_H(Y, u)|^2\right] \leq 9\sup_k |F(k) - H(k)|$ for $F^{-1}(u) < H^{-1}(u)$. This equation is symmetric with respect to F and H; therefore, it holds also for $F^{-1}(u) > H^{-1}(u)$.

(v) Let $[a, b]$ denote the interval $\Psi(\varepsilon)$ of length ε^2, $\sup \xi^2$ denote the supremum of ξ^2 over $u, v \in [a, b]$, where $\xi := I_F(Y, u) - u - I_F(Y, v) + v$.

Note that $|\xi| \leq 1$; moreover, if $[F(Y - 1), F(Y)] \cap [a, b] = \emptyset$, then $\sup |\xi| = \varepsilon^2$ and if $[a, b] \subset [F(Y - 1), F(Y)]$, then $\sup |\xi| = \frac{1 - f(Y)}{f(Y)}\varepsilon^2$.

Suppose that $F^{-1}(a) = F^{-1}(b)$, i.e. $[a, b] \subset [F(F^{-1}(a) - 1), F(F^{-1}(a))]$. Then $\mathrm{E}_F\left[\sup \xi^2\right] \leq \mathrm{E}_F\left[\sup |\xi|\right] = \varepsilon^2 \sum_{k \neq F^{-1}(a)} f(k) + \left(\frac{1 - f(F^{-1}(a))}{f(F^{-1}(a))}\right)\varepsilon^2 f(F^{-1}(a)) = 2(1 - f(F^{-1}(a)))\varepsilon^2 \leq 2\varepsilon^2$.

Suppose that $F^{-1}(a) < F^{-1}(b)$, i.e. $[a, b]$ contains at least one point $F(k)$ or even intervals $[F(k-1), F(k)] \subset [a, b]$. On such intervals, $|\xi|$ goes up to $1 - f(k)$, but the probability of Y taking all such k is bounded by $b - a$. More precisely,

$$
\begin{aligned}
\mathrm{E}_F\left[\sup \xi^2\right] \ \leq \ & \mathrm{E}_F\left[\sup |\xi|\right] \\
= \ & \varepsilon^2 \sum_{k < F^{-1}(a)-1} f(k) + \left(\frac{1 - f(F^{-1}(a))}{f(F^{-1}(a))}\right) \varepsilon^2 f(F^{-1}(a)) \\
& + \sum_{k \in [F^{-1}(a)+1, F^{-1}(b)-1]} f(k) + \left(\frac{1 - f(F^{-1}(b))}{f(F^{-1}(b))}\right) \varepsilon^2 f(F^{-1}(b)) + \varepsilon^2 \sum_{k > F^{-1}(b)} f(k) \\
< \ & 4\varepsilon^2,
\end{aligned}
$$

since the sum of the first and the last terms is below ε^2, the second and the fourth terms each is bounded by ε^2 and the third term is $\sum_{k \in [F^{-1}(a)+1, F^{-1}(b)-1]} f(k) = F\left(F^{-1}(b) - 1\right) - F\left(F^{-1}(a) + 1\right) \leq b - a = \varepsilon^2$. ∎

Proof of Lemma 1. Substitute $G = F = F_{\theta_0}(\cdot \mid \Omega_t)$ in Lemma A(i) to demonstrate that $E\left[I_{t,\theta_0}(u) \mid \Omega_t\right] = E\left[I_{t,\theta_0}(u)\right] = u$, therefore $I_{t,\theta_0}(u) - u$ is a martingale difference sequence for every $u \in [0, 1]$. The conditional variance expression follows from Lemma A(iii) by taking $G = F = F_{\theta_0}(\cdot \mid \Omega_t)$. ∎

However the $I_{t,\theta_0}(u)$ are not independent in general. To show that, note that bivariate independence requires that

$$
\Pr\left(I_{t,\theta_0}(u) \leq u_1, I_{t-1,\theta_0}(u) \leq u_2\right) = \Pr\left(I_{t,\theta_0}(u) \leq u_1\right) \Pr\left(I_{t-1,\theta_0}(u) \leq u_2\right)
$$

for all u, u_1 and $u_2 \in [0, 1]$. Now we see that the lhs is

$$
\begin{aligned}
\mathrm{E}\left[\mathbb{1}\{I_{t,\theta_0}(u) \leq u_1\} \mathbb{1}\{I_{t-1,\theta_0}(u) \leq u_2\}\right] = \ & \mathrm{E}\left[\mathrm{E}\left[\mathbb{1}\{I_{t,\theta_0}(u) \leq u_1\} \mathbb{1}\{I_{t-1,\theta_0}(u) \leq u_2\} \mid \Omega_t\right]\right] \\
= \ & \mathrm{E}\left[\mathbb{1}\{I_{t-1,\theta_0}(u) \leq u_2\} \mathrm{E}\left[\mathbb{1}\{I_{t,\theta_0}(u) \leq u_1\} \mid \Omega_t\right]\right]
\end{aligned}
$$

and now, for $u_1, u \in (0, 1)$ and under H_0,

$$
\begin{aligned}
\mathrm{E}\left[\mathbb{1}\{I_{t,\theta_0}(u) \leq u_1\} \mid \Omega_t\right] = \ & 1 - F_{\theta_0}\left(F_{\theta_0}^{-1}(u \mid \Omega_t) \mid \Omega_t\right) \\
& + \mathbb{1}\left\{1 - \delta_{F_{\theta_0}(\cdot \mid \Omega_t)}(u) \leq u_1\right\} f_{\theta_0}(\cdot \mid \Omega_t)\left(F_{\theta_0}^{-1}(u \mid \Omega_t)\right),
\end{aligned}
$$

which depends on Ω_t, and therefore $E\left(\mathbb{1}\{I_{t,\theta_0}(u) \leq u_1\} \mid \Omega_t\right) \neq E\left(\mathbb{1}\{I_{t,\theta_0}(u) \leq u_1\}\right)$ with positive probability, and independence does not follow in general.

37

Proof of Lemma 2. Because $U_t^r(\theta_0)$ are continuous, $\widehat{F}_{\theta_0}^r(u)$ is a (uniform) consistent estimate of cdf of $U_t^r(\theta_0)$. Then by Lemma A(vi) and A(vii) and ULLN we get the uniform consistency of $\widehat{F}_{\theta_0,M}^r(u)$ and $\widetilde{F}_{\theta_0}^r(u)$. The efficiency gain comes from Lemma A(ii). ∎

Proof of Lemma 3. We need to verify conditions N1-N3 of Theorem A. Fix $\varepsilon > 0$ and take $\Psi = [0,1]$ with usual norm and equidistant partition $0 = u_0 < u_1 < \ldots < u_{N_\Pi(\varepsilon)} = 1$, i.e. partition of $[0,1]$ in $N_\Pi(\varepsilon) = [\varepsilon^{-2}] + 1$ equal intervals of length ε^2 (the last interval maybe even smaller), $\Psi(\varepsilon; k) = [u_{k-1}, u_k]$ and $\xi_t^T = (I_F(Y_t, u) - u)/\sqrt{T}$, which is a square integrable martingale difference by Lemma 1. Then Condition N1 follows from Lemma 1 and Assumption 1. Condition N2 is satisfied because for $T > 1 + [\varepsilon^{-2}]$, the indicator $1\left\{\sup_{u \in [0,1]} |I_F(Y_t, u) - u|/\sqrt{T} > \varepsilon\right\} = 0$. Condition N3 follows from the bound in Lemma A(v). Indeed, $\int_0^1 H_\Pi(\varepsilon) d\varepsilon < \infty$ and

$$\left\| \xi_t^T \right\|_{\Pi,k} \leq \sup_{\varepsilon \in (0,1) \cap \mathbb{Q}} \frac{1}{\varepsilon} \max_{1 \leq k \leq N_\Pi(\varepsilon)} \sqrt{\varepsilon^2} \leq 1 \quad \text{a.s.}$$

∎

Proof of Lemma 4. Apply weak convergence result from Lemma 3 under $G_{T,\theta_0}(\cdot \mid \Omega_t)$ with $\xi_t^T := \left(I_{F_{\theta_0}(\cdot|\Omega_t)}(Y_t, u) - u - d\left(G_{T,\theta_0}(\cdot \mid \Omega_t), F_{\theta_0}(\cdot \mid \Omega_t), u\right) \right)/\sqrt{T}$, which is a square integrable martingale difference because of Lemma A(i) with $G = G_{T,\theta_0}(\cdot \mid \Omega_t)$ and $F = F_{\theta_0}(\cdot \mid \Omega_t)$. Then Condition N1 follows from Lemma A(iii) and the fact that $d(G, F, u, v)$ are bounded in absolute value by $T^{-1/2}$ a.s. Condition N2 is satisfied because for $T > 1 + [\varepsilon^{-2}]$, the indicator is 0. Condition N3 follows from the bound in Lemma A(v) and the fact that $(\mathbb{E}_G[\cdot] - \mathbb{E}_F[\cdot])$ applied to a.s. bounded r.v. are bounded in absolute value by $T^{-1/2}$ a.s. We obtain that $\sum_{t=1}^T \xi_t^T \Rightarrow S$, the same limit as in Lemma 3. Finally, use additivity of $d(\cdot, \cdot, \cdot)$ in the first argument and apply ULLN to $S_T - \sum_{t=1}^T \xi_t^T = \sum_{t=1}^T d\left(G_{T,\theta_0}(\cdot \mid \Omega_t), F_{\theta_0}(\cdot \mid \Omega_t), u\right)/\sqrt{T} = \delta \sum_{t=1}^T d\left(H(\cdot \mid \Omega_t), F_{\theta_0}(\cdot \mid \Omega_t), u\right)/T$. ∎

Proof of Lemma 5. Under H_{1T}, i.e. under G_{T,θ_0}, Equation (4) can be established using standard methods, applying Doob and Rosenthal inequalities for MDS (Hall and Heyde, 1980) $\sqrt{T} \xi_t^T := I_{F_{\widehat{\theta}_T}(\cdot|\Omega_t)}(Y_t, u) - I_{F_{\theta_0}(\cdot|\Omega_t)}(Y_t, u) - d\left(G_{T,\theta_0}(\cdot \mid \Omega_t), F_{\widehat{\theta}_T}(\cdot \mid \Omega_t), u\right) + d\left(G_{T,\theta_0}(\cdot \mid \Omega_t), F_{\theta_0}(\cdot \mid \Omega_t), u\right)$. Define $z_T := \sum_{t=1}^T \xi_t^T$. When it is necessary, we will

write explicitly arguments: $z_T(u, \widehat{\theta}_T)$. We show that $\sup_u |z_T| = o_p(1)$. Since $\sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) = O_P(1)$, it is sufficient to establish that for some $\gamma < 1/2$

$$\sup_{u, \|\eta - \theta_0\| \leq T^{-\gamma}} |z_T(u, \eta)| = o_p(1).$$

Note that for $T > \delta^2 / \nu_1^2$, by Assumption 3C,

$$\Pr\left(\sup_{\eta, t} \max_y |G_{T, t, \theta_0}(y \mid \Omega_t) - F_{t, \eta}(y \mid \Omega_t)| > \nu_1\right) \leq M_F T^{-\gamma} / \nu_1. \tag{9}$$

First, we will show that $\forall \eta, u \ |z_T| = o_p(1)$. Since ξ_t^T are bounded by 2 in absolute value and form a martingale difference sequence with respect to Ω_t, by the Doob inequality $\forall p \geq 1$ and $\forall \varepsilon > 0$

$$P\left(\max_{t=1, \ldots, T} |z_t| > \varepsilon\right) \leq E |z_T|^p / \varepsilon^p,$$

and by Rosenthal inequality, $\forall p \geq 2 \ \exists C_1$

$$E |z_T|^p \leq C_1 \left[E\left\{\sum E\left((\xi_t^T)^2 \mid \Omega_t\right)\right\}^{p/2} + \sum E |\xi_t^T|^p\right].$$

Take $p = 4$. The first term is small because of bounds in Lemma A(iv) and (9). Because $|\xi_t^T| \leq 2/\sqrt{T}$, $\sum E |\xi_t^T|^p \leq 2T^{1-p/2}$. Therefore we have a pointwise bound. Uniformity in u, η can be established using monotonicity of $I_{F_\theta(\cdot | \Omega_t)}(Y_t, u)$ and continuity of $d\left(G_{T, \theta_0}(\cdot \mid \Omega_t), F_{\widehat{\theta}_T}(\cdot \mid \Omega_t), u\right)$ by employing bounds in Lemma A(iv) and (9).

Finally, use that uniformly in u

$$\frac{1}{\sqrt{T}} \sum_t \left(d\left(G_{T, \theta_0}(\cdot \mid \Omega_t), F_{\widehat{\theta}_T}(\cdot \mid \Omega_t), u\right) - d\left(G_{T, \theta_0}(\cdot \mid \Omega_t), F_{\theta_0}(\cdot \mid \Omega_t), u\right)\right)$$

$$= \sqrt{T}\left(\widehat{\theta}_T - \theta_0\right) \frac{1}{T} \sum_t \nabla\left(F_{\theta_0}(\cdot \mid \Omega_t), u\right) + o_p(1).$$

■

Proof of Theorem 1. The joint weak convergence (6) follows from finite-dimensional convergence by CLT for MDS, while tightness was established in the proof of Lemma 4. ■

Proof of Theorem 2. Note that

$$S_{2T} = \sum_{t=2}^{T} \xi_t^T + \frac{1}{(T-1)^{1/2}} \left\{(I_{\theta_0, T}(u_1) - u_1) I_{\theta_0, T-1}(u_2) + u_1(I_{\theta_0, 1}(u_2) - u_2)\right\},$$

where

$$\xi_t^T := \frac{1}{(T-1)^{1/2}} \{(I_{t,\theta_0}(u_1) - u_1) I_{t-1,\theta_0}(u_2) + u_1 (I_{t,\theta_0}(u_2) - u_2)\}$$

is a square integrable martingale difference by Lemma 1. The rest is similar to the proof of Theorem 1. To obtain $S_{2T}(u) \Rightarrow S_{2\infty}(u)$ under H_0, verify conditions N1-N3 of Theorem A for ξ_t^T as it is done in the proof of Lemma 3. The covariance function of $S_{2\infty}(u)$ is

$$V_2(u,v) := (u_1 \wedge v_1)(u_2 \wedge v_2) - 3u_1 v_1 u_2 v_2$$

$$+ (u_1 \wedge v_1) \plim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} \gamma_{t-1,\theta_0}(u_2,v_2)$$

$$- \plim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} \gamma_{t,\theta_0}(u_1,v_1) \left(I_{t-1,\theta_0}(u_2 \wedge v_2) - \gamma_{t-1,\theta_0}(u_2,v_2)\right)$$

$$+ (u_2 \wedge v_1) u_1 v_2 - u_1 \plim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} \gamma_{t,\theta_0}(u_1,v_1) I_{t-1,\theta_0}(v_2)$$

$$+ (u_1 \wedge v_2) u_2 v_1 - v_1 \plim_{T\to\infty} \frac{1}{T} \sum_{t=2}^{T} \gamma_{t,\theta_0}(u_1,v_1) I_{t-1,\theta_0}(u_2).$$

Under H_{1T}, apply the same weak convergence result under $G_{T,t,\theta_0}(\cdot \mid \Omega_t)$ with

$$\zeta_t^T := \xi_t^T - I_{t-1,\theta_0}(u_2) d\left(G_{T,t,\theta_0}(\cdot \mid \Omega_t), F_{t,\theta_0}(\cdot \mid \Omega_t), u_1\right)/\sqrt{T-1}$$

$$+ u_1 d\left(G_{T,t,\theta_0}(\cdot \mid \Omega_t), F_{t,\theta_0}(\cdot \mid \Omega_t), u_2\right)/\sqrt{T-1},$$

which is a square integrable martingale difference because of Lemma A(i) with $G = G_{T,t,\theta_0}(\cdot \mid \Omega_t)$ and $F = F_{\theta_0}(\cdot \mid \Omega_t)$. Then proceed as in the proof of Lemma 4.

In order to establish (7), repeat the steps of the proof of Lemma 5 for $\widetilde{\zeta}_t^T := \zeta_t^T - \widehat{\zeta}_t^T$, where $\widehat{\zeta}_t^T$ is ζ_t^T with $F_{t,\widehat{\theta}_T}$ in place of F_{t,θ_0}. ∎

Proof of Theorem 4. Repeat the arguments of the proofs of Theorems 1 and 2 for sample generated by F_{θ_T}, defined in Assumption 6, to obtain conditional convergence. Then follow as in Andrews (1997) proof of Corollary 1. ∎

8.5 Checking assumptions for the Poisson model

Here we write Y_t for $Y_t^\star$. For Poisson model $Y_t \mid \Omega_t \sim \text{Poisson}(\lambda_t)$ the probability distribution is $\Pr(Y_t = k \mid \Omega_t) = P_{\lambda_t}(k) = \frac{\lambda_t^k \exp(-\lambda_t)}{k!}$ and the cumulative distribution function is

$$F_{t,\theta}(k \mid \Omega_t) = \sum_{j=0}^{k} \Pr(Y_t = j \mid \Omega_t) = \sum_{j=0}^{k} \frac{\lambda_t^j \exp(-\lambda_t)}{j!} = Q(k, \lambda_t),$$

where $Q(\cdot, \cdot)$ is the regularized gamma function, and $\lambda_t = \lambda_t(\beta) = \exp(X_t'\beta)$, $t = 1, 2, \ldots$. If covariates X_t are iid or stationary and ergodic, and Ω_t omits lags of the dependent variable Y_t, then the LLN applies both under the null and local alternatives (like, e.g., the local alternative considered in Eq. (2.12) in Cameron and Trivedi, 1990) to justify Assumptions 2-6 and Assumption A, which involve functions of Ω_t that are uniformly continuous in u. However, it can also be interesting to allow the intensity to depend on lags of the dependent variable. For simplicity we consider $AR(1)$ dynamics. $AR(p)$ can be treated similarly but is more lengthy. The parameters enter through $\lambda_t = \lambda_t(\theta) = \alpha_0 + \alpha_1 \lambda_{t-1} + \rho Y_{t-1}$, $t = 1, 2, \ldots$, and are gathered in $\theta = (\alpha_0, \alpha_1, \rho)'$. We assume that α_0, α_1, ρ are positive, λ_0 and Y_0 are fixed and $\alpha_1 + \rho < 1$. Under these conditions, there exist a unique stationary and ergodic solution to this model (Fokianos et al., 2009). Such data generating processes allow to use results on (generic, uniform) LLN, which facilitate the checking of assumptions in the paper. Conditions for stationarity and ergodicity for nonlinear $\lambda_t(\theta)$ can be found in Neumann (2011) and are directly applicable to the analysis under the null hypothesis. However, we are not aware of LLN results for these models under local alternatives despite Fokianos and Neumann (2013, Proposition 2.3(ii)) use related arguments.

Let $\lambda_{t,0} = \lambda_t(\theta_0)$ and the null hypothesis is $Y_t \mid \Omega_t \sim \text{Poisson}(\lambda_{t,0})$ for some $\theta_0 \in \Theta$. Then $U_t = Q(Y_t + 1, \lambda_{t,0})$ and $U_t^- = Q(Y_t, \lambda_{t,0})$, and the *nonrandomized transform $Y_t \mapsto$*

$I_{t,\theta_0}(u)$ for $u \in [0,1]$ is

$$
I_{t,\theta_0}(u) = \begin{cases} 0, & u \leq Q(Y_t, \lambda_{t,0}); \\ \dfrac{u - Q(Y_t, \lambda_{t,0})}{\lambda_{t,0}^{Y_t} \exp(-\lambda_{t,0})} Y_t!, & Q(Y_t, \lambda_{t,0}) \leq u \leq Q(Y_t + 1, \lambda_{t,0}); \\ 1, & Q(Y_t + 1, \lambda_{t,0}) \leq u, \end{cases}
$$

from where one obtains the empirical processes and the test statistics defined in Sections 1-2.

Now consider Assumption 1. For Poisson model

$$
\gamma_{t,\theta_0}(u,v) = \frac{(Q(k+1, \lambda_{t,0}) - u \vee v)(u \wedge v - Q(k, \lambda_{t,0}))}{\lambda_{t,0}^k \exp(-\lambda_{t,0})} k! \mathbb{1}\{k(u) = k(v)\},
$$

where $k = k(u) = \min\{y : Q(y, \lambda_{t,0}) \geq u\}$. For the Poisson DGP described above, Y_t is stationary and ergodic, $\gamma_\infty(u,v) := \mathrm{E}\left[\gamma_{1,\theta_0}(u,v)\right]$ satisfies Assumption 1. By the same argument Assumptions 2, 3D, 4C, 5 are fulfilled.

Assumption 3A and 3B are trivial. For Assumption 3C note that

$$
\dot{F}_{t,\theta}(k \mid \Omega_t) = \left(\sum_{j=0}^{k-1} \frac{\lambda_t^j}{j!} - \sum_{j=0}^{k} \frac{\lambda_t^j}{j!} \right) \exp(-\lambda_t) \dot{\lambda}_t = -\frac{\lambda_t^k}{k!} \exp(-\lambda_t) \dot{\lambda}_t,
$$

where

$$
\dot{\lambda}_t = \left(1 + \alpha_1 \frac{\partial \lambda_{t-1}}{\partial \alpha_0}, \lambda_{t-1} + \alpha_1 \frac{\partial \lambda_{t-1}}{\partial \alpha_1}, Y_{t-1} + \alpha_1 \frac{\partial \lambda_{t-1}}{\partial \rho} \right)'.
$$

The last expression can be iterated from $t-1$ to $t=1$ and because $\alpha_1 < 1$ the arithmetic progression sum of mean squares is bounded, as in the proof of Lemma 3.2 of Fokianos et al. (2009).

Assumption 4A, 4B and 6B are standard, see e.g. Andrews (1997) which adapts to Poisson model using Theorem 3.1 of Fokianos et al. (2009).

Assumption 6A is trivial, because there is no explanatory variables other than own past values.

9   Acknowledgements

References

[1] Andrews, D.W.K. (1997) A conditional Kolmogorov test, *Econometrica* 65, 1097-1128.

[2] Bai, J. (2003) Testing Parametric Conditional Distributions of Dynamic Models, *Review of Economics and Statistics* 85, 531-549.

[3] Basu, D. and R. de Jong (2007). Dynamic Multinomial Ordered Choice with an Application to the Estimation of Monetary Policy Rules. *Studies in Nonlinear Dynamics and Econometrics*, 11, 1507-1507.

[4] Burke, M. D., Csorgo M., Csorgo S. and P. Revesz (1978). Approximaiton of the empirical process whe parameters are estimated. *Annals of Probability*, 7, 790-810.

[5] Cameron A.C. and P.K. Trivedi (1990) Regression-based tests for overdispersion in the Poisson model, *Journal of Econometrics* 46, 347-364.

[6] Corradi, V. and R. Swanson (2006) Bootstrap conditional distribution test in the presence of dynamic misspecification, *Journal of Econometrics* 133, 779-806.

[7] Czado, C., T. Gneiting and L. Held (2009). Predictive model assessment for count data. *Biometrics*, 65, 1254-1261.

[8] Davis R. A., W. T. M. Dunsmuir and S. B. Streett (2003) Observation-Driven Models for Poisson Counts. *Biometrika* 90, 777-790.

[9] Delgado, M. and Escanciano, J. C. (2007) Nonparametric tests for conditional symmetry in dynamic models. *Journal of Econometrics* 141, 652-682.

[10] Delgado, M. and W. Stute (2008) Distribution-free specification tests of conditional models. *Journal of Econometrics* 143, 37-55.

[11] Dolado, J. J. and Maria-Dolores, R. (2002). Evaluating Changes in the Bank of Spain's Interest Rate Target: An Alternative Approach Using Marked Point Processes, *Oxford Bulletin of Economics and Statistics*, 64, 159-182.

[12] Doukhan P., K. Fokianos and D. Tjostheim (2012) On weak dependence conditions for Poisson autoregressions, *Statistics and Probability Letters* 82, 942-948.

[13] Fokianos K. and M. Neumann (2013) A goodness-of-fit test for Poisson count processes, *Electronic Journal of Statistics* 7, 793-819.

[14] Fokianos K., A. Rahbek and D. Tjostheim (2009) Poisson Autoregression, *Journal of the American Statistical Association* 104, 1430-1439.

[15] Genest, C., J. Nešlehová and B. Rémillard (2014). On the empirical multilinear copula process for count data. *Bernoulli,* 20, 1344-1371.

[16] Giacomini R., Politis D. and H. White (2013) A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators, *Econometric Theory* 29, 567-589.

[17] Greene, W. and Hensher D., (2010). *Modeling Ordered Choices. A Primer.* Cambridge University Press.

[18] Hamilton, J. and O. Jordá (2002). A model of the Federal Funds rate target. *The Journal of Political Economy* 110, 1135-1167.

[19] Handcock, M. S. and Morris, M. (1999). Relative Distribution Methods in the Social Sciences. Springer, New York.

[20] Jacod, J. and Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes.* Springer, 2d Ed. Berlin.

[21] Jung, R.C., M. Kukuk and R. Liesenfeld (2006). Time series of count data: modeling, estimation and diagnostics. *Computational Statistics and Data Analysis*, 51, 2350-2364.

[22] Kauppi, H. and Saikkonen, P. (2008). Predicting U.S. recessions with dynamic binary response models. *Review of Economics and Statistics* 90, 777-791.

[23] Kedem and Fokianos (2002). *Regression Models for Time Series Analysis.* Willey, New Jersey.

[24] Kheifets, I. (2015). Specification Tests for Nonlinear Time Series Models. *Econometrics Journal* 18, 67-94.

[25] Kheifets, I., and C. Velasco (2013). Model Adequacy Checks for Discrete Choice Dynamic Models. In Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis Essays in Honor of Halbert L. White Jr. Chen, Xiaohong; Swanson, Norman R. (Eds.), 363-382.

[26] Lee, S. (2014). Goodness of fit test for discrete random variables, *Computational Statistics and Data Analysis* 69, 92-100.

[27] Machado, J. A. F., and Santos Silva, J. M. C. (2005). Quantiles for Counts, *Journal of the American Statistical Association* 100, 1226-1237.

[28] Mora, J. and A.I. Moro-Egido (2007). On specification testing of ordered discrete choice models. *Journal of Econometrics* 143, 191-205.

[29] Neumann M. (2011) Absolute regularity and ergodicity of Poisson count processes, *Bernoulli* 17, 1268-1284.

[30] Nishiyama Y. (2000). Weak convergence of some classes of martingales with jumps. *Annals of Probability* 28, 685-712.

[31] Rydberg and Shephard (2003). Dynamics of Trade-By-Trade Price Movements: Decomposition and Models. *Journal of Financial Econometrics* 1, 2-25.

[32] Rosenblatt M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics* 23, 470-72.

[33] Startz R. (2008). Binomial Autoregressive Moving Average Models with an Application to U.S. Recessions. *Journal of Business and Economic Statistics* 26, 1-8.

[34] Stute W., Gonzalez Manteiga W., and M. Presedo Quindimil (2002). Bootstrap Based Goodness-Of-Fit-Tests. *Metrika* 40, 243-256.

[35] Stute W. and Zhu, L.-X. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics* 29, 535-545.

[36] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer, New York.