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

By

Igor Kheifets and Carlos Velasco

November 2013

COWLES FOUNDATION DISCUSSION PAPER NO. 1924

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

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

Igor Kheifets† and Carlos Velasco‡

November 4, 2013

Abstract

This paper proposes new specification tests for conditional models with discrete responses. In particular, we can test the static and dynamic ordered choice model specifications, which 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. The traditional approach 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 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. 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 traditional approach 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.

*We thank Juan Mora for useful comments. Financial support from the Fundación Ramón Areces and from the Spain Plan Nacional de I+D+I (ECO2012-31748) is gratefully acknowledged.

†New Economic School, Moscow. Email: ikheifets@nes.ru

‡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. We consider both static models with covariates as well as dynamic ordered choice models, where the conditioning information set may include also past information on the discrete variable and a set of (contemporaneous) explanatory variables often appearing in biological and social sciences. These models are applied in sociology, marketing, political science, medicine, transportation planning, economics and finance, see a survey of Greene and Hensher (2010). For example, dynamic models are nowadays very 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. Apart from the specification of the conditional information relevant to the problem, the researcher typically has to specify the distribution of the latent continuous errors as well as a link function to summarize regressors information.

Before conducting inference based on such models it is needed some goodness of fit analysis of the chosen model. This is typically implemented through specification tests which establish the suitability of the fitted model by a comparison with a reference distribution, possibly complemented by some independence or uncorrelation residual tests. 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 $\Theta \subset R^m$ is the parameter space, while the alternative hypothesis H_1 for omnibus test would be the negation of H_0.

We consider a class $\mathcal{M} \equiv \mathcal{M}(\nu, \mathcal{K})$ of discrete distributions F defined on $\mathcal{K} = \{1, \ldots, K\}$, such that $F(0) = 0$, $P_F(k) := F(k) - F(k-1) \geq \nu > 0$ for $k \in \mathcal{K}$ and some ν, and $F(K) = 1$. For conditional distributions, we write $F_{t,\theta}(\cdot \mid \Omega_t) \in \mathcal{M}$ if the above definition holds a.s. with the same ν and $\mathcal{K}$ for every t, Ω_t and $\theta \in \Theta$. From now on we suppose that $F_{t,\theta}(\cdot \mid \Omega_t) \in \mathcal{M}$. See an overview of specification tests for such setup in Mora and Moro-Egido (2007) and a discussion of some alternative tests and applications in Section 6.

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),

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

are standard uniforms and independent under H_0. This serves as a basis for several specification tests of H_0, see e.g. Bai (2003) andt Kheifets (2013) 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 limitation 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. The easiest way is to interpolate the discrete values of Y_t with independent noise in $[0,1]$, cf. Kheifets and Velasco (2013), but this additional noise affects the power of the tests and may lead to different conclusion depending on the simulation outcome.

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

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

where $U_t^-(\theta_0) := F_{t,\theta_0}(Y_t - 1 \mid \Omega_t)$. This transform is nonrandom 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). As we show below, $I_{\theta_0,t}(u) - u$ constitute a martingale difference sequence (MDS) with respect to Ω_t under H_0 and can be used for testing H_0 as $I_{\theta_0,t}(u)$ loses such 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}

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

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

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

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

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

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 parameter θ_0 is unknown under the null, we use an estimate $\hat{\theta}_T$ and account for parameter estimation effect in the p-value computation with a parametric bootstrap method. It might be possible to derive, e.g. martingale, distribution free transforms but since they typically need to be programed case by case for each model, they may be impractical, therefore this task is left 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 nonrandom 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 nonrandom 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 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 random and non random transformations. Then we conclude. All proofs are contained in Appendix.

2 ALTERNATIVES TO PIT

In order to motivate the nonrandom transform (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 as well. Equivalently, 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 $Y_t^\dagger$,

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

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 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 specifications, while $U_t(\theta_0)$ and $U_t^-(\theta_0)$ are not independent nor $U[0,1]$. Then using standard discrepancy measures, the empirical cdf of $U_t^r(\theta_0)$, estimated using the *random transform* $Y_t \mapsto 1\{U_t^r(\theta_0) \leq u\}$,

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

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

$$R_{1T}(u) := T^{1/2}\left\{\hat{F}^r_{\theta_0}(u) - u\right\} = \frac{1}{T^{1/2}}\sum_{t=1}^{T}\left[1\{U^r_t(\theta_0) \leq u\} - u\right], \quad u \in [0,1].$$

We can also consider reducing the effect of the noise Z^U_t in (3) and in the random transform by taking averages over M replications of $\{Z^U_t\}_{t=1}^T$, conditional on the original data, similar to "average-jittering" of Machado and Santos Silva (2005). Suppose that for each the t we have M independent sequences of standard uniform noises $Z^U_{t,m}$, $m = 1, 2, \ldots, M$, which generate $U^r_{t,m}(\theta_0)$ according to (3). Define the *M-random* transform $Y_t \mapsto I_{\theta_0,t,M}(Y_t, u)$,

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

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

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

Note that with $M = 1$ we are back to $\hat{F}^r_{\theta_0}(u)$, and equivalently we can generalize R_{1T} to

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

In order to propose specification tests, following Handcock and Morris (1999), we define a *discrete relative distribution* of Y_t compared to F_{t,θ_0} as the cdf of $U^r_t(\theta_0)$. Under H_0, the discrete relative distribution is the standard uniform. 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: $\tilde{F}_{\theta_0}(u)$ (the most efficient), $\hat{F}^r_{\theta_0,M}(u)$ and $\hat{F}^r_{\theta_0}(u)$. This order is determined by the amount of noise introduced in the definitions of the transforms: i.e. in nonrandom, M-random and (1-)random transforms. The nonrandom transform can be equivalently obtained by integrating out the extra noise in the random transform $I_{\theta_0,t}(Y_t, u) = \int 1\{U^r_t(\theta_0) \leq u\}\, dF_Z$ or 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. Efficiency of nonrandom transform translates into the increased power of the specification tests based on this transform, whose properties we study next.

3 SPECIFICATION TESTS BASED ON THE NEW TRANSFORM

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

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

with $k = F_{t,\theta_0}^{-1}(u \mid \Omega_t)$ and $F_k := F_{t,\theta_0}(k \mid \Omega_t)$, and the conditional quantile function is defined as $F_{t,\theta_0}^{-1}(u \mid \Omega_t) := \min\{y : F_{t,\theta_0}(y \mid \Omega_t) \geq u\}$ for $u \in [0,1]$.

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

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

with conditional covariance

$$\mathrm{E}[I_{\theta_0,t}(u) I_{\theta_0,t}(v) \mid \Omega_t] = u \wedge v - uv - \delta_{F_{t,\theta_0}(\cdot|\Omega_t)}(u,v), \quad a.s.$$

Remark 1. $I_{\theta_0,t}(u)$ are not necessarily independent across t.

Remark 2. By the martingale difference property, $I_{\theta_0,t}(u)$ and $I_{\theta_0,t-j}(v)$ are uncorrelated for all $j \neq 0$ and all $u, v \in [0,1]$. On the other hand, the $I_{\theta_0,t}(u)$ 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.

Remark 3. Let $V_T(u,v) := \mathrm{Cov}[S_{1T}(u), S_{1T}(v)]$, then

$$V_T(u,v) = u \wedge v - uv - \mathrm{E}\left[\frac{1}{T}\sum_{t=1}^{T} \delta_{F_{t,\theta_0}(\cdot|\Omega_t)}(u,v)\right] \leq u \wedge v - uv,$$

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

Due to Lemma 1, $\mathrm{E}\left[\tilde{F}_{\theta_0}(u)\right] = u$ under H_0 and the natural empirical processes to perform tests on H_0 is then S_{1T}. This process, being based on a nonrandom transform, does not involve the extra noise that appears in the random 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 M-random transform. Next lemma is the key to understand the improvement of the M-random over the random; and of the nonrandom, advocated in this paper, over the M-random transform approaches.

Lemma 2. *Independently of whether* H_0 *holds or not,* $\hat{F}^r_{\theta_0,M}(u)$ *and* $\tilde{F}_{\theta_0}(u)$ *consistently and uniformly in* u *estimate the relative distribution, i.e. the cdf of* $U^r_t(\theta_0)$*.* $\tilde{F}_{\theta_0}(u)$ *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 Remark 3 and 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). Another advantages 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 nonrandom transform.

Assumption 1 Under H_0, there exists a finite $\delta_\infty(u,v)$, such that uniformly in (u,v) $\frac{1}{T}\sum_{t=1}^{T}\delta_{F_{t,\theta_0}(\cdot|\Omega_t)}(u,v) \to_p \delta_\infty(u,v)$.

Remark 4. We restrict dynamics such that the limit in probability exists, i.e. the law of large number (LLN) holds. In case of stationary and ergodic data, $\delta_\infty(u,v) = \mathrm{E}\left[\delta_{F_{1,\theta_0}(\cdot|\Omega_1)}(u,v)\right]$. Sufficient conditions for the stationarity of autoregressive Y_t appearing in our application are given in Basu and de Jong (2007). Note that the limit is also uniform, since the summands are continuous, piece-wise polynomial in u and v. This remark applies also everywhere below, where we utilize “plim” in an assumption.

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 $V(u,v) := u \wedge v - uv - \delta_\infty(u,v)$.

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*.*

The distribution of S_1 is model and parameter dependent and practical implementation of tests when θ_0 is unknown is discussed in next section. We finish this section with a discussion of the asymptotic properties of 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.

3.1 Power Analysis

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

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

Note, that $d(G, F, u) \equiv 0$ if and only if $G \equiv F$. Under any $G_t(\cdot \mid \Omega_t) \in \mathcal{M}$,

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

We consider the behavior of S_{1T} under the following class of local alternatives to H_0,

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

where

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

for some $0 < \delta < T^{1/2}$ and for all t, $H_t(\cdot \mid \Omega_t) \in \mathcal{M}$.

Assumption 2 Under H_0, there exists a finite $D(u)$, such that uniformly in u $\frac{1}{T}\sum_{t=1}^{T} d\left(H_t(\cdot \mid \Omega_t), F_{t,\theta_0}(\cdot \mid \Omega_t), u\right) \to_p D(u)$.

Remark 5. Remark 4 on the limit existence applies here. Note that under standard conditions the convergence is uniform, since the summands are piece-wise linear, because the function $d(\cdot,\cdot,\cdot)$ is piece-wise linear in u.

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

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

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 $\hat{\theta}_T$, say. We assume that we have available an estimate $\hat{\theta}_T$ so that under H_{1T}

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

and analyze the consequences of replacing θ_0 by $\hat{\theta}_T$ in S_{1T}, i.e. we consider

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

Let $\|\cdot\|$ be Euclidean norm, i.e. for matrix A, $\|A\| = \sqrt{\operatorname{tr}(AA')}$, 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. We need the following assumptions to analyze the asymptotic properties of $\hat{S}_{1T}$.

Assumption 3 (Parametric family)

(A) Parameter space Θ is a compact set in a finite-dimensional Euclidean space, $\theta \in \Theta \subset \mathbb{R}^m$.

(B) There exist $\delta > 0$, such that $F_{t,\theta}(\cdot \mid \Omega_t) \in \mathcal{M}$, in particular, $P_{F_{t,\theta}}(k \mid \Omega_t) \geq \nu > 0$ for $k = 1, \ldots, K$, 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 θ and $\max_t \mathrm{E}\left[\max_k \sup_\theta \left\|\dot{F}_{t,\theta}(k \mid \Omega_t)\right\|\right] \leq M_F < \infty$, where $\dot{F}_\theta := (\partial/\partial\theta) F_\theta$.

(D) Under H_{1T}, there exists a finite $L(u) := \operatorname{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \nabla\left(F_{t,\theta_0}(\cdot \mid \Omega_t), u\right)$, where for a cdf F_θ in $\mathcal{M}$,

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

where $\dot{P}_{F_\theta} := (\partial/\partial\theta) P_{F_\theta}$.

Remark 6. Assumption 3 is standard, see e.g. Bai (2003), we add only condition (D). Note that $\nabla(\cdot, u)$ is a piece-wise linear function in u, and therefore Remark 4 on the limit existence applies. Conditions for no effect of information truncation can be provided similar to Bai (2003).

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

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

uniformly in u.

Then no longer $\eta\left(\hat{S}_{1T}\right)$ converges to $\eta(S_1 + \delta D)$ under H_{1T}, but also the estimation effect has to be taken into account.

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

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

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

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

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

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

In particular, under H_0, $\delta\xi_0=0$ and $T^{1/2}\left(\hat{\theta}_T-\theta_0\right)$ is asymptotically $N\left(0,\Psi\right)$.

Remark 7. Assumption 4 holds for the MLE of many popular discrete models, including dynamic probit and logit and general discrete choice models. As an example consider estimates $\hat{\theta}_T$ which are asymptotically equivalent to the (conditional) maximum likelihood estimates, i.e.,

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

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

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

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

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

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

where $Z\left(\delta\xi_0,\Psi\right)$ is a normal vector with mean $\delta\xi_0$ and covariance matrix Ψ, and the asymptotic covariance function between both terms is $W(u)$.

We can state the result now on the asymptotic distribution of the empirical process $\hat{S}_{1T}$.

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

$$\hat{S}_{1T} \Rightarrow \hat{S}_{1\infty},$$

where $\hat{S}_{1\infty} := S_{1\infty} + Z(0, \Psi)' L + \delta \{D + \xi_0' L\}$ *is the Gaussian process with mean function* $\delta \{D(u) + \xi_0' L(u)\}$ *and variance function* $V(u, v) + L(u)' \Psi L(v) + W(u)' L(v) + W(v)' L(u)$.

4 DYNAMIC SPECIFICATION TESTS

Test statistics based on S_{1T}, R_{1T} and $R_{1T,M}$ check 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. For testing continuous distributions, this issue is raised in Corradi and Swanson (2006), Delgado and Stute (2008) and Kheifets (2013). However developing specification tests conditioning on infinite dimensional values of Ω_t is not possible. 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, $I_{\theta_0,t}(u_1) - u_1$ is MDS, $I_{\theta_0,t}(u_1) I_{\theta_0,t-1}(u_2) - u_1 u_2$ is centered around zero, and moreover

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

This motivates us to develop tests based on S_{2T} defined in (2). This process has also zero mean under the null and identifies not only departures from the null derived from deviations of the unconditional expectation of $I_{\theta_0,t}(u)$ from u, but also from a possible failure of the martingale property, so that $I_{\theta_0,t}(u_1)$ and $I_{\theta_0,t-1}(u_2)$ would become correlated. This idea is similar to that exploited in Kheifets' (2013) in the context of conditional distribution testing for continuous distributions, where different methods to check the independent property of the PIT are proposed. Alternative statistics exploiting the lack of correlations with any other lag could be proposed, but we might expect that low lags can typically be more useful to detect general forms of misspecification. One could consider also a biparameter analog of $R_{1T,M}$, i.e. for some $M = 1, 2, \ldots$,

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

where $u = (u_1, u_2)$ and $v = (v_1, v_2)$ are in $[0,1]^2$, i.e. $u_i, v_i \in [0,1]$. 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 random transforms, and therefore suffer from power loss compared to tests based on the nonrandom 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}.

In practice we use the process

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

Then, under H_{1T}, to study the parameter estimation effect consider

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

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

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

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

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

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.

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

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

where $S_{2\infty}$ is a Gaussian process in $[0,1]$ with zero mean and covariance function $V_2(u,v)$ defined in the Appendix. Under H_{1T}, if parameters are estimated,

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

where $\hat{S}_{2\infty} := S_{2\infty} + Z(0,\Psi)' L_2$ is the Gaussian process with zero mean and variance function $V_2(u,v) + L_2(u)' \Psi L_2(v) + W_2(u)' L_2(v) + W_2(v)' L_2(u)$.

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 $\hat{S}_{2T}$ will have 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, therefore can capture dynamic misspecification when misspecification results in such correlation, even if the unconditional expectation of d, which appears in the second term, is zero. This fact is crucial if misspecification occurs in dynamics and not in the link function.

5 BOOTSTRAP TESTS

To test H_0 we consider Cramer-von Mises, Kolmogorov-Smirnov or any other continuous functionals of $\hat{S}_{jT}$, $j = 1, 2$, $\eta\left(\hat{S}_{jT}\right)$. Then consistency properties of specification tests based on $\hat{S}_{jT}$ 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},*

$$\eta\left(\hat{S}_{jT}\right) \Rightarrow \eta\left(\hat{S}_{j\infty}\right).$$

Since the asymptotic distributions of $S_{jT}(u)$ are model dependent, and those of $\hat{S}_{jT}(u)$ 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 $F_{t,\hat{\theta}}(\cdot \mid \Omega_t)$ to mimic the null distribution of the test statistics. For parametric bootstrap see Andrews (1997), which can be adapted to complications with information truncation and initialization arising in dynamic case using discussion in Bai (2003). We describe the algorithm now.

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

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

2. Simulate Y_t^* with $F_{\hat{\theta}}(\cdot \mid \Omega_t^*)$ 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 model with simulated data Y_t^*, get $\hat{\theta}_T^*$ using the same method as for $\hat{\theta}_T$, get bootstrapped test statistics $\eta\left(\hat{S}_{jT}^*\right)$.

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

5. Reject H_0 if $\eta\left(\hat{S}_{jT}\right)$, is greater than the corresponding $(1-\alpha)$th percentile of the empirical distribution of B the bootstrap resamples $\eta\left(\hat{S}^*_{jT}\right)$, $\hat{c}^*_{jB}\left(\hat{\theta}_T\right)$.

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 (Bootstrap) 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 $\hat{\theta}_T$ of θ_T admits an asymptotic linear expansion as in Assumption 4. Moreover, assume that under the alternative H_1, there exists some θ_1 so that $\theta_1 = \text{plim}_{T\to\infty} \hat{\theta}_T$.

Then we obtain the following result.

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

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

so $\hat{c}^*_{jB}\left(\hat{\theta}_T\right) \to_p c_j(\theta_0)$, and therefore, under H_0, $\Pr\left(\eta\left(\hat{S}_{jT}\right) > \hat{c}^*_{jB}\left(\hat{\theta}_T\right)\right) \to \alpha$. Under H_1, as $B, T \to \infty$, $\hat{c}^*_{jB}\left(\hat{\theta}_T\right) = O_p(1)$.

This theorem shows that under the null we get the right asymptotic size and that under local alternatives we get non trivial power when the drifts of the stochastic processes $\hat{S}_{1T}$ and $\hat{S}_{2T}$ 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. $\lim_{T\to\infty} \Pr\left(\eta\left(\hat{S}_{jT}\right) > \hat{c}^*_{jB}\left(\hat{\theta}_T\right)\right) = 1$ if $\eta\left(\hat{S}_{jT}\right)$ is asymptotically unbounded.

6 APPLICATION AND SIMULATIONS

In this section we consider a Monte Carlo simulation exercise to investigate on 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,

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

The dynamic multinomial ordered choice model that explains y_t can be represented as

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

where V_t^* is a continuous latent variable and $\tau_1, \ldots, \tau_{K-1}$ are threshold parameters that define K intervals in $\mathbb{R}$. Then 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

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

with $\tau_0 = -\infty$ and $\tau_K = \infty$.

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 just 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 these models 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 series Y_t and conduct our tests on these (see Monte Carlo scenarios in Table 5). To speed up the simulation procedure we use the warp bootstrap algorithm of Giacomini, Politis and White (2013).

The four choices of output gap lead to Models I-IV. Output gap is constructed as the percentage deviation of actual from potential output, interpolating 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, 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 approaches can be adapted to our setup: the test of Generalized linear model (GLM) of Stute and Zhu (2002) and the Conditional Kolmogorov test of Andrews (1997), both are considered in Mora and Moro-Egido (2007). The first one, is a test based on a marked empirical process for testing the null $H_0' : \quad \mathrm{E}\left[Y \mid \tilde{X} = x\right] = m_{\tilde{\beta}_{01}}\left(x'\tilde{\beta}_{02}\right)$, where $m_{\tilde{\beta}_{01}}(\cdot)$ is a parametric link function and $\tilde{\beta}_{01}, \tilde{\beta}_{02}$ are some finite dimensional parameters. In case Y takes only two values $\{0, 1\}$, the conditional mean coincides with the conditional probability and the null is similar to our H_0 if we were considering an i.i.d setup. To test $Y_t \mid \tilde{X}_t \sim P_{\tilde{\beta}_{01}}\left(\cdot \mid \tilde{X}_t'\tilde{\beta}_{20}\right)$ define

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

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

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

and

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

which we call CvM and KS tests respectively. To apply these tests to our model, let $\tilde{X}_t = \left(X_t', Y_{t-1}\right)'$ and $\tilde{\beta} = \left(\beta', \gamma\right)'$ and take 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 $\hat{S}_{1T}$, $\hat{R}_{1T,M}$, etc, where we replace θ_0 by root-T consistent estimates $\hat{\theta}_T$, the ML estimator in our case. We are not aware of any theoretical results for bootstrap assisted tests based on $\hat{Z}_T$ in our setup, although Mora and Moro-Egido (2007) provide some simulations.

Parameter estimates for real data are reported in Tables 1 and 2. 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 3 and 4 say that all these models are rejected even at 1% significance level by biparameter nonrandom transform based tests. Note that single parameter static tests (e.g. $\hat{R}_{1T}$, $\hat{S}_{1T}$) can not reject any proposed model, with the only exception of $\hat{S}_{1T}$ which rejects at 5% Model II with Cramer – von Misses test statistics.

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). 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 apparently the choice of the output gap series does not make big differences either, nor the introduction of a lagged endogenous (discrete) variables in the information set.

The power of the tests for static Probit model is analyzed against three different alternatives: static Logit, dynamic Probit and dynamic Logit. We see that the tests without random smoothing, $\hat{S}_{1T}$ and $\hat{S}_{2T}$ always perform better than random continuous processes $\hat{R}_{1T,M}$ and $\hat{R}_{2T,M}$ which in turn dominate $\hat{R}_{1T}$ and $\hat{R}_{2T}$, thus confirming our theoretical findings. When we compare Probit and Logit specifications, while letting the dynamic aspect of the model 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 against a dynamic Logit alternatives is very high. Since the nature of misspecification is dynamic, again bivariate processes should have more power compared to single parameter counterparts, as is confirmed in our simulation results. It can also be observed that for these alternatives, Cramer-von Mises criterium provides more power than Kolmogorov-Smirnov tests. As for alternative tests based on $\hat{Z}_T$, they have power comparable to $\hat{S}_{1T}$, sometimes slightly better, and are always outperformed by any bivariate test. This is not surprising, since $\hat{Z}_T$ puts more structure, i.e. it assumes single-index model for covariates but averages across points, hence 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 closer to 100% for the dynamic alternatives.

7 CONCLUSIONS

In this paper we have proposed new specification tests for the conditional distribution of discrete time series data. The new tests are functionals of empirical processes based on a nonrandom transform that solves the implementation problem of the usual PIT for discrete distributions and achieve consistency against a wide class of alternatives. We show the validity of a bootstrap algorithm to approximate 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 nonrandom transform in the unconditional case

To stress the generality of results in this subsection, we omit subscripts t, θ_0 and use shortcuts $I_F(Y, u) = I_{\theta_0,t}(Y_t, u)$ and $I_{F,M}(Y, u) = I_{\theta_0,t,M}(Y_t, u)$. 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 $\Pr_G(Y = F^{-1}(u)) = G(F^{-1}(u)) - G(F^{-1}(u) - 1) =: P_G(F^{-1}(u))$. When $G = F$, $\Pr_F(F(Y) < u) = F(F^{-1}(u) - 1) < u$, 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 nonrandom transform can be written as

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

where

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

Note that $\delta_F(u) \in [0, 1)$. We see that $I_F(Y, u)$ is a piecewise linear (continuous) increasing in u function. In Lemma A we list the properties of this transform. They can be derived using results from Table 10, so the proof is omitted. Let

$$\begin{aligned}\delta_F(u,v) := &\ (\delta_F(u\vee v) - \delta_F(u)\,\delta_F(v))\,P_F\left(F^{-1}(u\wedge v)\right)\\ &\times 1\left\{F^{-1}(u) = F^{-1}(v)\right\} \in \left[0, u\wedge v\wedge P_F\left(F^{-1}(u\wedge v)\right)\right],\end{aligned}$$

$$\begin{aligned}d(G,F,u,v) := &\ d(G,F,u\wedge v)\\ &- (\delta_F(u\vee v) - \delta_F(u)\,\delta_F(v))\,1\left\{F^{-1}(u) = F^{-1}(v)\right\}\\ &\times \left(P_G\left(F^{-1}(u)\right) - P_F\left(F^{-1}(u)\right)\right),\end{aligned}$$

$$1-\delta_{F,H}(u) := \begin{cases} 1-\delta_F(u), & F^{-1}(u) < H^{-1}(u);\\ (1-\delta_F(u))\,(1-\delta_H(u)), & F^{-1}(u) = H^{-1}(u);\\ 1-\delta_H(u), & F^{-1}(u) > H^{-1}(u).\end{cases}$$

Lemma A. *For $0 \le v, u \le 1$*

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

(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) $|I_F(Y,u) - I_H(Y,u)| \le 3\frac{\max|F-H|}{\nu}$.

(v) $0 \le I_F(Y,u) - I_F(Y,v) \le (u-v)/\nu$.

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

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

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 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 $\{\xi_t^T\}_{t=1,2,\ldots}$ be a $\ell^\infty(\Psi)$-valued martingale difference array with respect to filtration $\mathcal{F}_t^T$, i.e. for every t, ξ_t^T maps Ω^T to $\ell^\infty(\Psi)$, 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 to study 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 $\Pi(\varepsilon) = \{\Psi(\varepsilon;k)\}_{1\le k\le N_\Pi(\varepsilon)}$ 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

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

Theorem A. *Let $\{\xi_t^T\}_{t=1,2,\ldots}$ be a $\ell^\infty(\Psi)$-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) (Lindenberg condition) $\sum_{t=1}^T \mathrm{E}\left[\left\|\xi_t^T\right\|^2 1\left\{\left\|\xi_t^T\right\| > \varepsilon\right\} \mid \mathcal{F}_t^T\right] \to_{P^T} 0$ for every $\varepsilon > 0$;
N3) (partitioning entropy condition) there exist a DFP Π of Ψ such that $\left\|\xi_t^T\right\|_{\Pi,T} = O_{P^T}(1)$ 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 Proofs

Proof of Lemma 1. Substitute $G = F = F_{\theta_0}(\cdot \mid \Omega_t)$ in Lemma A(i) to obtain that $E\left[I_{\theta_0,t}(u) \mid \Omega_t\right] = E\left[I_{\theta_0,t}(u)\right] = u$, therefore $I_{\theta_0,t}(u) - u$ 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 $I_{\theta_0,t}(u)$ are not independent in general. To show that, note that bivariate independence requires that

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

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

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

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

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

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

Proof of Lemma 2. Because $U_t^r(\theta_0)$ are continuous, $\hat{F}_{\theta_0}^r(u)$ 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 uniform consistency of $\hat{F}_{\theta_0,M}^r(u)$ and $\tilde{F}_{\theta_0}^r(u)$. 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. 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 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 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 bound in Lemma A(v) and that $(\mathrm{E}_G[\cdot] - \mathrm{E}_F[\cdot])$ applied to a.s. bounded r.v. are bounded in absolute value by $T^{-1/2}$ a.s. We obtain that $\sum_{t=1}^T \xi_t^T \Rightarrow S$, to 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) $\sqrt{T}\xi_t^T := I_{F_{\hat{\theta}_T}(\cdot|\Omega_t)}(Y_t,u) - I_{F_{\theta_0}(\cdot|\Omega_t)}(Y_t,u) - d\left(G_{T,\theta_0}(\cdot \mid \Omega_t), F_{\hat{\theta}_T}(\cdot \mid \Omega_t), u\right)$ $+d\left(G_{T,\theta_0}(\cdot \mid \Omega_t), F_{\theta_0}(\cdot \mid \Omega_t), u\right)$. Define $z_T := \sum_{t=1}^T \xi_t^T$. When it is necessary we will write explicitly arguments: $z_T(u,\hat{\theta}_T)$. We show that $\sup_u |z_T| = o_p(1)$. Since $\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) = O_P(1)$, 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,

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

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 $\forall \varepsilon > 0$

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

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

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

Take $p = 4$. The first term is small because of bounds in Lemma A(iv) and (9). Because $\left|\xi_t^T\right| \leq 1$, $T^{-p/2}\sum E\left|\xi_t^T\right|^p \leq T^{1-p/2}$. Therefore we have pointwise bound. Uniformity in u, η can be established using monotonicity of $I_{F_\theta(\cdot|\Omega_t)}(Y_t, u)$ and continuity of $d\left(G_{T,\theta_0}\left(\cdot \mid \Omega_t\right), F_{\hat{\theta}_T}\left(\cdot \mid \Omega_t\right), u\right)$ by employing bounds in Lemma A(iv) and (9). Note, that bound in Lemma A(iv) is used when $\sup_{\eta,t}\max_y\left|G_{T,t,\theta_0}\left(y \mid \Omega_t\right) - F_{t,\eta}\left(y \mid \Omega_t\right)\right| < \nu$, which holds with probability approaching to 1 as shown in (9).

Finally, use that uniformly in u

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

∎

Proof of Theorem 1. 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. We need the following

Assumption A Under H_{1T}, the following uniform limits exist

(i) $\operatorname{plim}_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\delta_{F_{t-1,\theta_0}(\cdot|\Omega_{t-1})}\left(u_2, v_2\right)\delta_{F_{t,\theta_0}(\cdot|\Omega_t)}\left(u_1, v_1\right)$,

(ii) $\operatorname{plim}_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}I_{\theta_0,t-1}\left(v_2\right)\delta_{F_{t,\theta_0}(\cdot|\Omega_t)}\left(u_1, v_1\right)$,

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

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

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

Note that

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

where

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

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

$$\begin{aligned}
V_2(u,v) :=\ & (u_1 \wedge v_1)(u_2 \wedge v_2) - 3u_1 v_1 u_2 v_2 \\
& + (u_1 \wedge v_1) \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} \delta_{F_{t-1,\theta_0}(\cdot|\Omega_{t-1})}(u_2, v_2) \\
& - \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} \delta_{F_{t,\theta_0}(\cdot|\Omega_t)}(u_1, v_1)\Big(I_{\theta_0,t-1}(u_2 \wedge v_2) - \delta_{F_{t-1,\theta_0}(\cdot|\Omega_{t-1})}(u_2, v_2)\Big) \\
& + (u_2 \wedge v_1) u_1 v_2 - u_1 \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} \delta_{F_{t,\theta_0}(\cdot|\Omega_t)}(u_1, v_1) I_{\theta_0,t-1}(v_2) \\
& + (u_1 \wedge v_2) u_2 v_1 - v_1 \operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=2}^{T} \delta_{F_{t,\theta_0}(\cdot|\Omega_t)}(u_1, v_1) I_{\theta_0,t-1}(u_2).
\end{aligned}$$

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

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

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 proof of Lemma 4.

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

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

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

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

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

[8] 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.

[9] 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.

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

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

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

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

[14] 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.

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

[16] Kheifets, I.(2013). Specification Tests for Nonlinear Time Series Models. Mimeo.

[17] 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.

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

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

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

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

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

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

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

Table 1: ML estimates and standard errors of Models I-IV with static and dynamic specifications and Probit link function applied to the original data of length $T = 204$.

	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 2: ML estimates and standard errors of Models I-IV with static and dynamic specifications and Logit link function applied to the original data of length $T = 204$.

	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 3: P-values of Cramer – von Misses tests for static Probit and Logit link function applied to the original data of length $T = 204$.

	$\hat{S}_{2T}$	$\hat{R}_{2T,50}$	$\hat{R}_{2T,25}$	$\hat{R}_{2T}$	$\hat{S}_{1T}$	$\hat{R}_{1T,50}$	$\hat{R}_{1T,25}$	$\hat{R}_{1T}$	$\hat{Z}_T$
				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 4: P-values of Kolmogorov – Smirnov tests for static Probit and Logit link function applied to the original data of length $T = 204$.

	$\hat{S}_{2T}$	$\hat{R}_{2T,25}$	$\hat{R}_{2T,25}$	$\hat{R}_{2T}$	$\hat{S}_{1T}$	$\hat{R}_{1T,50}$	$\hat{R}_{1T,25}$	$\hat{R}_{1T}$	$\hat{Z}_T$
				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 5: 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

Table 6: Empirical rejection rates of various Cramer – von Misses tests of Models I-IV with static and dynamic specifications applied to simulated data of length $T = 100$.

		$\hat{S}_{2T}$	$\hat{R}_{2T,50}$	$\hat{R}_{2T,25}$	$\hat{R}_{2T}$	$\hat{S}_{1T}$	$\hat{R}_{1T,50}$	$\hat{R}_{1T,25}$	$\hat{R}_{1T}$	$\hat{Z}_T$
Size 1 H_0 : static probit										
	10%	13.6	13.4	12.5	9.0	13.3	11.1	10.8	9.2	13.7
Model I	**5%**	**5.5**	**6.0**	**5.5**	**4.5**	**6.6**	**6.3**	**5.7**	**5.4**	**7.8**
	1%	2.0	1.5	1.1	0.4	0.8	1.1	0.4	0.5	2.2
	10%	12.8	11.9	11.2	11.0	9.9	9.9	9.7	8.2	13.6
Model II	**5%**	**5.3**	**6.7**	**5.0**	**5.5**	**6.3**	**5.2**	**4.2**	**3.3**	**6.5**
	1%	0.5	0.9	1.0	0.7	1.5	1.7	1.3	0.9	2.8
	10%	14.5	12.9	13.5	10.4	10.3	11.7	10.6	7.7	14.0
Model III	**5%**	**7.7**	**7.0**	**6.5**	**5.4**	**6.0**	**3.7**	**3.3**	**4.5**	**6.4**
	1%	1.2	1.0	1.4	1.1	1.4	0.8	0.4	0.6	0.6
	10%	11.1	11.3	10.3	7.6	9.7	9.5	8.4	7.9	13.7
Model IV	**5%**	**5.2**	**6.7**	**5.6**	**3.9**	**5.1**	**4.6**	**4.9**	**2.8**	**6.4**
	1%	0.8	0.7	0.7	0.7	1.7	0.6	1.1	0.5	1.6
Size 2 H_0 : static logit										
	10%	13.9	13.3	11.4	7.6	12.9	11.6	12.5	7.7	15.5
Model I	**5%**	**6.5**	**6.5**	**4.9**	**4.1**	**7.2**	**5.6**	**6.0**	**4.4**	**7.2**
	1%	1.2	2.0	1.0	1.0	0.9	1.3	1.6	0.9	1.7
	10%	13.7	13.8	12.8	10.9	11.9	8.7	9.6	8.1	11.1
Model II	**5%**	**5.6**	**6.7**	**7.6**	**4.0**	**4.6**	**5.3**	**4.6**	**4.8**	**5.6**
	1%	1.2	1.5	1.5	0.7	0.7	1.0	0.9	1.1	1.0
	10%	14.3	14.3	12.3	9.2	11.1	12.7	9.9	8.8	14.0
Model III	**5%**	**7.3**	**9.0**	**6.4**	**3.3**	**6.4**	**7.8**	**5.2**	**3.3**	**8.5**
	1%	1.5	1.4	1.7	0.7	1.2	2.1	2.2	0.5	2.4
	10%	10.9	10.8	10.3	9.2	12.6	9.9	9.6	8.1	16.1
Model IV	**5%**	**6.6**	**6.3**	**5.0**	**4.5**	**6.5**	**4.6**	**4.7**	**4.7**	**9.1**
	1%	2.3	2.5	0.9	0.3	1.7	1.0	0.7	0.7	1.2
Power 1 H_0 : static probit vs H_1 : static logit										
	10%	14.6	13.7	13.6	8.9	15.5	13.1	12.0	8.0	13.8
Model I	**5%**	**8.5**	**7.7**	**6.6**	**4.9**	**8.4**	**6.5**	**6.0**	**3.6**	**7.1**
	1%	1.5	1.7	1.2	0.7	2.5	2.0	1.2	0.5	2.1
	10%	10.1	9.9	9.8	9.1	11.5	10.9	11.3	9.0	14.4
Model II	**5%**	**5.1**	**5.0**	**4.4**	**4.0**	**6.4**	**6.9**	**5.3**	**4.0**	**8.7**
	1%	0.7	0.9	0.8	0.9	2.2	1.9	1.5	0.4	3.1
	10%	17.0	15.0	13.5	9.8	18.4	14.8	16.2	7.9	16.0
Model III	**5%**	**9.1**	**9.4**	**7.9**	**4.7**	**9.0**	**8.3**	**7.7**	**4.6**	**8.2**
	1%	2.2	1.7	2.3	0.3	4.1	2.6	2.0	0.4	1.9
	10%	13.8	13.0	10.7	8.9	16.5	16.9	13.5	7.8	14.8
Model IV	**5%**	**6.3**	**6.2**	**5.3**	**4.5**	**10.2**	**8.6**	**7.5**	**3.8**	**8.3**
	1%	0.7	1.3	0.8	0.8	2.5	1.7	1.2	1.0	1.2
Power 2 H_0 : static probit vs H_1 : dynamic probit										
	10%	93.0	92.0	90.3	38.4	22.9	20.9	18.4	8.7	24.7
Model I	**5%**	**89.2**	**85.2**	**82.7**	**25.7**	**13.2**	**12.0**	**11.7**	**4.6**	**18.4**
	1%	75.5	71.6	67.0	5.2	5.1	3.8	3.9	0.4	7.6
	10%	96.5	95.4	94.5	46.2	18.8	17.2	14.6	8.7	25.2
Model II	**5%**	**92.8**	**92.3**	**91.1**	**34.2**	**10.5**	**8.1**	**8.8**	**3.0**	**17.2**
	1%	86.2	82.2	81.6	11.3	1.7	1.7	2.5	0.7	3.7
	10%	93.9	91.5	92.0	35.9	14.9	16.0	14.8	8.2	15.9
Model III	**5%**	**90.7**	**88.4**	**86.1**	**22.5**	**9.2**	**9.8**	**8.5**	**5.0**	**9.4**
	1%	83.8	76.4	72.4	8.3	2.4	2.3	3.0	0.5	2.7
	10%	92.3	89.0	87.4	39.4	15.1	14.0	12.0	8.6	19.7
Model IV	**5%**	**88.1**	**84.1**	**83.0**	**27.7**	**10.3**	**7.8**	**7.4**	**4.4**	**12.5**
	1%	73.3	68.2	66.0	6.4	4.1	3.4	3.1	0.7	3.8
Power 3 H_0 : static probit vs H_1 : dynamic logit										
	10%	95.5	92.6	90.6	33.4	19.2	13.3	13.8	8.8	21.6
Model I	**5%**	**90.1**	**89.3**	**86.0**	**22.9**	**12.1**	**10.0**	**8.5**	**5.0**	**12.6**
	1%	81.1	73.8	69.9	7.5	2.2	3.9	2.5	0.5	5.1
	10%	96.7	94.8	94.4	40.9	15.9	16.5	15.1	10.2	23.1
Model II	**5%**	**94.2**	**93.0**	**90.6**	**29.8**	**9.6**	**9.1**	**7.1**	**3.9**	**14.6**
	1%	90.3	84.6	80.2	11.7	2.9	2.4	2.0	0.7	5.2
	10%	96.3	95.0	93.5	38.7	16.9	14.5	12.8	10.1	17.9
Model III	**5%**	**93.5**	**91.9**	**90.9**	**30.3**	**10.0**	**8.0**	**7.8**	**4.4**	**10.9**
	1%	85.7	83.5	80.9	11.2	1.9	1.6	2.1	0.5	2.8
	10%	94.4	91.7	89.1	37.2	19.3	19.1	18.3	10.3	22.8
Model IV	**5%**	**91.1**	**88.4**	**85.9**	**26.0**	**11.1**	**12.3**	**11.4**	**4.7**	**14.7**
	1%	80.8	80.6	76.6	10.1	4.1	4.4	3.9	0.8	4.9

Table 7: Empirical rejection rates of various Kolmogorov – Smirnov tests of Models I-IV with static and dynamic specifications applied to simulated data of length $T = 100$.

		$\hat{S}_{2T}$	$\hat{R}_{2T,50}$	$\hat{R}_{2T,25}$	$\hat{R}_{2T}$	$\hat{S}_{1T}$	$\hat{R}_{1T,50}$	$\hat{R}_{1T,25}$	$\hat{R}_{1T}$	$\hat{Z}_T$
Size 1 H_0 : static probit										
	10%	12.2	10.6	9.9	7.5	12.8	12.2	12.0	10.0	13.7
Model I	**5%**	**5.1**	**6.4**	**5.2**	**3.9**	**7.8**	**6.3**	**6.8**	**4.9**	**7.9**
	1%	1.3	0.9	1.2	0.6	1.1	1.4	0.6	0.8	1.8
	10%	12.3	11.6	8.8	10.2	10.4	9.3	11.0	8.7	12.8
Model II	**5%**	**5.5**	**6.5**	**3.9**	**4.9**	**5.9**	**5.1**	**4.1**	**4.8**	**6.2**
	1%	1.0	1.3	1.6	0.4	1.2	0.9	1.1	1.0	2.7
	10%	13.4	14.5	14.5	10.0	11.5	11.8	12.0	8.6	12.4
Model III	**5%**	**7.7**	**7.8**	**6.8**	**5.1**	**6.1**	**7.0**	**6.0**	**4.9**	**5.6**
	1%	1.1	1.2	1.7	1.0	1.7	0.5	1.0	0.5	0.7
	10%	12.7	11.1	9.8	8.1	9.9	10.3	9.4	9.2	12.6
Model IV	**5%**	**6.5**	**5.4**	**5.3**	**3.4**	**5.3**	**5.3**	**4.8**	**3.6**	**7.2**
	1%	0.4	0.6	0.4	1.0	2.1	1.8	1.5	0.4	2.7
Size 2 H_0 : static logit										
	10%	12.5	12.8	11.1	9.7	14.2	12.3	11.7	7.7	15.4
Model I	**5%**	**7.0**	**6.4**	**6.1**	**5.4**	**9.1**	**6.4**	**6.3**	**3.7**	**6.7**
	1%	0.8	1.4	1.5	0.7	2.0	1.7	2.1	1.1	1.6
	10%	10.2	9.8	10.7	9.0	12.9	8.8	9.1	9.0	11.3
Model II	**5%**	**4.7**	**4.9**	**4.6**	**3.5**	**5.6**	**3.8**	**4.0**	**4.8**	**5.8**
	1%	0.9	1.0	0.8	0.5	1.3	0.6	0.9	0.8	1.4
	10%	13.3	14.5	13.6	7.8	11.8	10.4	9.7	8.6	14.2
Model III	**5%**	**8.3**	**8.3**	**6.7**	**3.2**	**6.2**	**5.7**	**3.5**	**4.0**	**10.0**
	1%	0.9	1.0	0.7	0.6	1.2	0.9	0.9	0.2	2.3
	10%	11.8	10.9	10.2	8.9	13.0	10.3	10.4	8.9	16.4
Model IV	**5%**	**6.2**	**6.5**	**5.1**	**4.7**	**6.6**	**5.8**	**5.3**	**4.0**	**8.1**
	1%	1.3	1.4	1.3	0.7	1.7	1.1	1.1	0.9	1.3
Power 1 H_0 : static probit vs H_1 : static logit										
	10%	13.9	11.9	10.1	9.5	9.2	7.5	7.0	8.0	13.2
Model I	**5%**	**7.0**	**6.2**	**5.4**	**3.7**	**5.2**	**3.3**	**3.9**	**3.2**	**7.7**
	1%	1.5	1.2	0.8	0.2	0.7	1.1	0.4	0.9	2.2
	10%	9.6	9.6	8.3	8.3	7.6	6.9	9.7	9.7	15.2
Model II	**5%**	**4.3**	**3.8**	**4.5**	**3.7**	**4.1**	**3.9**	**3.6**	**3.9**	**8.9**
	1%	1.0	0.7	0.8	0.9	0.5	0.9	1.1	1.1	3.0
	10%	16.9	15.2	13.2	10.2	14.7	9.8	10.1	8.5	15.4
Model III	**5%**	**10.2**	**7.3**	**7.1**	**3.9**	**7.1**	**5.7**	**5.7**	**4.5**	**9.2**
	1%	1.7	1.4	1.4	0.4	1.9	1.5	1.1	0.4	1.8
	10%	13.1	11.5	10.3	9.7	11.2	12.6	9.8	8.4	11.2
Model IV	**5%**	**5.6**	**6.6**	**4.3**	**3.2**	**6.4**	**5.1**	**6.2**	**3.4**	**6.8**
	1%	0.9	1.3	1.3	0.7	2.4	1.0	1.1	0.6	1.1
Power 2 H_0 : static probit vs H_1 : dynamic probit										
	10%	88.5	85.6	83.2	24.6	16.2	14.4	13.5	9.1	23.5
Model I	**5%**	**82.8**	**79.0**	**74.5**	**13.6**	**10.3**	**9.1**	**7.1**	**3.5**	**16.9**
	1%	66.0	58.9	55.4	1.7	1.8	2.4	2.1	0.7	6.2
	10%	91.7	91.2	89.2	27.8	18.9	16.8	17.0	8.0	22.5
Model II	**5%**	**87.9**	**85.5**	**83.3**	**17.7**	**12.1**	**11.2**	**9.3**	**3.3**	**14.0**
	1%	78.4	72.4	71.9	4.2	4.5	2.9	3.2	0.6	3.3
	10%	91.1	87.7	85.2	22.4	14.8	11.8	11.7	8.3	16.8
Model III	**5%**	**85.7**	**83.2**	**79.4**	**13.8**	**7.1**	**6.4**	**7.2**	**3.9**	**9.2**
	1%	67.1	65.4	60.9	5.9	2.4	2.5	3.0	0.7	3.3
	10%	89.0	85.9	83.2	26.1	13.8	11.3	10.6	10.3	16.9
Model IV	**5%**	**81.7**	**78.5**	**74.6**	**13.8**	**7.7**	**7.8**	**6.7**	**4.9**	**11.3**
	1%	61.9	59.3	54.3	2.9	2.7	2.8	3.1	1.4	3.8
Power 3 H_0 : static probit vs H_1 : dynamic logit										
	10%	92.4	89.0	86.9	24.3	13.1	8.6	9.6	9.1	20.8
Model I	**5%**	**86.2**	**82.7**	**79.0**	**14.2**	**7.7**	**4.9**	**3.8**	**4.2**	**11.8**
	1%	68.7	64.0	57.8	4.2	1.5	1.0	1.4	0.7	4.3
	10%	93.5	90.7	89.6	27.9	17.5	13.9	13.5	9.4	20.8
Model II	**5%**	**90.0**	**86.2**	**82.2**	**15.9**	**9.3**	**7.9**	**8.1**	**4.1**	**14.2**
	1%	80.7	74.9	66.0	5.5	3.7	3.0	2.0	0.4	6.1
	10%	93.7	90.3	89.1	29.4	11.4	11.2	10.8	9.1	16.1
Model III	**5%**	**89.0**	**86.4**	**83.7**	**15.9**	**5.6**	**5.1**	**4.4**	**4.6**	**10.5**
	1%	79.5	73.3	69.9	3.3	1.8	1.0	1.7	0.7	3.2
	10%	91.1	88.1	86.2	24.4	16.0	13.1	13.5	10.7	21.0
Model IV	**5%**	**87.5**	**83.8**	**79.3**	**16.1**	**9.4**	**7.5**	**7.7**	**5.9**	**12.9**
	1%	74.0	70.6	67.0	4.8	4.0	4.1	4.1	1.0	6.1

Table 8: Empirical rejection rates of various Cramer – von Misses tests of Models I-IV with static and dynamic specifications applied to simulated data of length $T = 200$.

		$\hat{S}_{2T}$	$\hat{R}_{2T,50}$	$\hat{R}_{2T,25}$	$\hat{R}_{2T}$	$\hat{S}_{1T}$	$\hat{R}_{1T,50}$	$\hat{R}_{1T,25}$	$\hat{R}_{1T}$	$\hat{Z}_T$
Size 1 H_0 : static probit										
	10%	11.8	10.3	11.3	10.1	10.7	11.7	10.6	10.5	10.6
Model I	**5%**	**4.0**	**5.4**	**5.7**	**6.2**	**4.2**	**4.6**	**4.9**	**5.8**	**5.2**
	1%	0.7	1.0	1.0	0.8	0.7	0.9	0.7	0.6	0.8
	10%	8.9	9.4	9.0	7.1	10.5	10.5	12.2	9.7	11.4
Model II	**5%**	**4.5**	**4.4**	**3.5**	**2.4**	**6.3**	**4.7**	**5.9**	**4.4**	**7.0**
	1%	0.8	0.5	1.1	0.5	1.3	1.1	1.5	0.7	1.4
	10%	9.6	9.1	8.9	9.3	12.3	10.5	10.2	9.6	10.3
Model III	**5%**	**4.6**	**4.4**	**3.4**	**4.2**	**5.4**	**5.5**	**5.2**	**3.3**	**5.4**
	1%	1.0	0.8	1.4	0.3	0.6	1.2	1.5	0.9	1.1
	10%	10.0	11.0	11.2	9.4	10.5	10.5	11.5	11.6	14.1
Model IV	**5%**	**5.3**	**6.1**	**6.3**	**4.4**	**4.8**	**4.6**	**7.0**	**4.9**	**6.9**
	1%	1.1	0.6	1.5	0.8	1.3	0.9	1.0	1.7	1.1
Size 2 H_0 : static logit										
	10%	13.7	13.4	11.8	9.9	9.9	12.1	11.7	9.9	10.1
Model I	**5%**	**7.2**	**8.2**	**6.7**	**5.8**	**5.8**	**6.7**	**6.4**	**3.8**	**5.2**
	1%	1.4	1.3	1.3	0.9	1.7	1.7	1.0	0.6	1.1
	10%	9.9	11.4	11.2	8.5	10.2	10.4	10.4	10.0	9.6
Model II	**5%**	**5.4**	**6.4**	**6.1**	**4.7**	**4.8**	**5.6**	**5.3**	**5.2**	**6.1**
	1%	1.0	0.8	1.0	1.2	1.7	1.3	1.3	1.1	1.0
	10%	8.7	11.3	9.3	9.4	9.0	10.3	11.0	9.3	12.4
Model III	**5%**	**5.3**	**5.2**	**3.9**	**4.0**	**5.6**	**5.8**	**6.7**	**4.4**	**6.9**
	1%	0.5	0.9	1.0	0.9	0.8	0.7	1.0	1.1	1.4
	10%	11.8	12.9	11.1	10.7	11.6	11.7	10.9	8.4	13.4
Model IV	**5%**	**5.4**	**6.8**	**5.0**	**4.0**	**5.6**	**5.2**	**5.0**	**4.1**	**8.3**
	1%	2.1	1.6	1.3	0.8	1.6	1.4	2.0	1.3	2.4
Power 1 H_0 : static probit vs H_1 : static logit										
	10%	13.6	14.6	12.7	12.1	18.0	19.2	17.7	13.1	17.1
Model I	**5%**	**7.2**	**8.2**	**6.6**	**6.9**	**10.9**	**10.3**	**10.9**	**6.9**	**9.2**
	1%	1.1	0.8	2.0	1.6	2.1	1.8	2.6	2.0	1.1
	10%	8.9	10.2	12.5	10.7	12.4	12.7	14.1	11.0	10.0
Model II	**5%**	**4.5**	**4.9**	**5.6**	**6.3**	**7.5**	**6.4**	**7.3**	**6.7**	**6.5**
	1%	1.9	1.4	1.5	2.0	3.0	2.4	2.0	1.1	1.3
	10%	11.4	11.8	11.1	10.4	16.1	15.8	14.4	11.2	15.8
Model III	**5%**	**6.0**	**5.2**	**6.1**	**5.9**	**6.9**	**6.8**	**7.9**	**6.6**	**7.0**
	1%	1.2	1.7	1.6	0.9	1.6	1.8	1.9	0.8	1.5
	10%	10.7	11.8	11.9	11.0	15.8	14.1	15.4	11.3	10.1
Model IV	**5%**	**6.5**	**6.6**	**6.6**	**4.3**	**7.1**	**6.1**	**7.5**	**5.8**	**5.4**
	1%	1.7	2.5	2.3	1.1	1.3	2.1	1.3	1.2	1.5
Power 2 H_0 : static probit vs H_1 : dynamic probit										
	10%	99.0	98.2	97.6	44.5	22.6	19.4	17.7	12.4	25.8
Model I	**5%**	**98.5**	**97.3**	**95.2**	**33.2**	**13.6**	**11.6**	**9.8**	**7.5**	**16.2**
	1%	94.5	91.9	82.4	14.9	5.2	4.0	3.3	1.9	6.8
	10%	99.8	99.6	99.1	52.3	24.8	22.2	22.7	11.9	28.5
Model II	**5%**	**99.5**	**99.3**	**98.5**	**41.5**	**16.0**	**14.8**	**12.6**	**7.1**	**18.2**
	1%	98.7	96.1	92.5	21.2	7.1	5.4	3.6	2.0	5.5
	10%	99.3	98.4	97.4	41.8	19.1	18.7	17.3	12.5	24.6
Model III	**5%**	**98.5**	**97.0**	**95.9**	**30.7**	**13.0**	**11.8**	**9.8**	**7.9**	**13.9**
	1%	93.9	90.7	90.4	14.4	3.0	3.5	3.2	0.7	3.0
	10%	98.0	96.3	95.1	33.6	18.7	17.3	15.3	11.9	20.5
Model IV	**5%**	**95.8**	**93.6**	**91.6**	**22.9**	**10.2**	**9.5**	**7.6**	**5.3**	**13.7**
	1%	88.4	86.1	81.3	9.5	3.8	1.5	1.6	0.3	3.4
Power 3 H_0 : static probit vs H_1 : dynamic logit										
	10%	99.5	98.5	97.7	48.3	25.4	21.8	21.1	11.3	25.3
Model I	**5%**	**98.6**	**97.5**	**95.6**	**34.5**	**15.2**	**14.0**	**14.0**	**5.4**	**16.7**
	1%	94.5	91.2	88.7	15.5	4.7	5.5	3.7	1.7	7.1
	10%	99.8	99.5	99.3	53.6	26.1	23.8	21.5	13.1	31.1
Model II	**5%**	**99.5**	**98.9**	**98.6**	**39.1**	**16.9**	**16.1**	**13.7**	**7.4**	**20.8**
	1%	98.1	96.9	95.6	15.0	5.3	6.5	5.9	1.1	8.0
	10%	99.6	98.7	98.1	45.8	24.8	21.3	19.5	13.1	28.4
Model III	**5%**	**98.8**	**98.1**	**96.4**	**31.2**	**14.8**	**13.7**	**11.6**	**6.6**	**17.9**
	1%	96.8	92.7	91.2	15.5	5.4	3.7	3.5	1.1	4.4
	10%	98.6	97.2	95.4	34.7	18.6	18.8	16.5	11.0	19.4
Model IV	**5%**	**95.8**	**94.2**	**91.8**	**23.9**	**11.4**	**10.1**	**8.6**	**5.3**	**11.4**
	1%	89.3	86.5	82.7	11.8	3.1	1.9	1.7	1.5	1.5

Table 9: Empirical rejection rates of various Kolmogorov – Smirnov tests of Models I-IV with static and dynamic specifications applied to simulated data of length $T = 200$.

		$\hat{S}_{2T}$	$\hat{R}_{2T,50}$	$\hat{R}_{2T,25}$	$\hat{R}_{2T}$	$\hat{S}_{1T}$	$\hat{R}_{1T,50}$	$\hat{R}_{1T,25}$	$\hat{R}_{1T}$	$\hat{Z}_T$
Size 1 H_0 : static probit										
	10%	10.3	11.2	10.2	9.8	11.2	10.0	11.3	10.3	10.5
Model I	**5%**	**5.1**	**5.0**	**6.0**	**4.8**	**4.5**	**5.9**	**3.9**	**5.1**	**5.5**
	1%	1.3	1.4	1.5	0.9	0.8	0.8	1.1	1.3	0.7
	10%	7.7	8.7	7.5	7.9	11.1	11.2	12.4	10.5	11.7
Model II	**5%**	**3.7**	**3.9**	**3.9**	**2.9**	**6.3**	**5.6**	**6.4**	**4.6**	**5.7**
	1%	1.0	0.4	0.7	0.4	1.3	1.8	1.5	0.7	1.5
	10%	9.9	9.2	10.0	9.8	10.1	7.7	9.7	10.2	10.9
Model III	**5%**	**4.5**	**5.2**	**4.3**	**3.9**	**4.2**	**4.5**	**4.9**	**4.3**	**5.3**
	1%	0.6	0.6	1.1	0.8	0.9	0.6	0.3	1.1	1.3
	10%	10.2	10.7	11.1	8.6	10.0	8.0	9.7	12.1	14.7
Model IV	**5%**	**5.0**	**6.4**	**7.0**	**4.6**	**4.6**	**4.8**	**6.4**	**6.9**	**6.8**
	1%	1.2	0.8	1.1	1.2	1.7	1.9	1.4	0.9	1.5
Size 2 H_0 : static logit										
	10%	11.3	14.3	11.9	9.4	12.0	11.9	11.4	8.8	9.1
Model I	**5%**	**5.7**	**5.7**	**6.3**	**4.9**	**6.3**	**5.9**	**6.3**	**3.5**	**4.8**
	1%	1.0	1.4	1.3	0.7	1.7	1.1	0.5	0.7	1.6
	10%	9.8	10.0	9.7	9.5	10.5	9.4	12.2	10.3	9.4
Model II	**5%**	**5.5**	**5.1**	**5.9**	**3.4**	**5.4**	**4.6**	**5.9**	**4.9**	**5.3**
	1%	0.7	1.3	2.3	0.9	1.1	1.5	0.7	1.0	1.0
	10%	9.8	8.2	8.8	9.5	10.6	9.9	9.6	11.7	11.9
Model III	**5%**	**3.6**	**5.4**	**4.6**	**3.6**	**6.4**	**4.3**	**5.2**	**5.3**	**7.8**
	1%	0.7	1.4	0.8	1.0	1.2	0.5	0.3	1.2	1.6
	10%	10.7	12.2	11.9	9.7	14.3	13.3	11.3	8.0	13.0
Model IV	**5%**	**6.4**	**7.3**	**5.6**	**4.7**	**6.6**	**6.4**	**4.7**	**4.5**	**8.5**
	1%	2.0	2.0	1.1	0.8	2.0	1.0	0.6	0.9	3.3
Power 1 H_0 : static probit vs H_1 : static logit										
	10%	13.0	11.1	10.5	12.8	12.0	11.3	10.1	12.1	15.5
Model I	**5%**	**6.3**	**6.5**	**4.2**	**6.6**	**7.2**	**5.9**	**5.0**	**6.7**	**8.7**
	1%	0.8	0.9	1.0	0.9	1.8	1.1	1.4	2.1	0.6
	10%	8.6	11.1	10.3	11.8	10.7	11.3	11.1	10.8	10.2
Model II	**5%**	**4.6**	**5.0**	**6.4**	**6.3**	**5.2**	**4.9**	**5.7**	**6.5**	**6.1**
	1%	1.1	1.0	0.9	1.7	1.7	0.7	0.7	1.0	1.0
	10%	11.8	12.5	12.9	10.1	10.7	11.6	10.2	11.4	14.3
Model III	**5%**	**5.0**	**6.2**	**5.7**	**5.2**	**3.7**	**4.1**	**5.0**	**6.3**	**7.1**
	1%	0.9	0.8	1.6	1.0	0.4	0.5	0.8	1.0	1.2
	10%	12.6	11.0	12.8	9.2	10.3	8.4	10.8	11.0	9.7
Model IV	**5%**	**5.7**	**7.0**	**5.6**	**4.5**	**5.4**	**3.4**	**4.6**	**6.1**	**5.1**
	1%	1.3	0.8	1.4	1.5	0.7	1.0	1.1	1.6	1.6
Power 2 H_0 : static probit vs H_1 : dynamic probit										
	10%	97.8	95.5	93.1	34.6	22.4	19.3	17.8	12.0	24.2
Model I	**5%**	**94.3**	**92.0**	**86.5**	**22.8**	**11.4**	**10.6**	**9.7**	**5.9**	**14.0**
	1%	82.9	81.7	73.0	8.9	5.3	4.9	3.2	1.4	4.9
	10%	99.4	98.6	97.3	38.5	23.5	20.9	20.5	13.3	25.0
Model II	**5%**	**98.1**	**96.5**	**94.4**	**26.3**	**15.5**	**13.1**	**13.5**	**7.3**	**13.1**
	1%	90.7	86.8	84.1	10.7	7.9	5.9	3.8	1.7	4.7
	10%	96.8	95.3	93.3	31.4	22.1	23.1	18.6	13.8	21.4
Model III	**5%**	**94.3**	**91.0**	**87.9**	**21.0**	**14.7**	**13.1**	**12.7**	**7.3**	**13.8**
	1%	81.7	72.2	67.1	10.3	4.2	2.4	3.0	1.1	2.1
	10%	94.8	90.8	87.5	26.3	17.0	16.2	18.1	13.0	19.8
Model IV	**5%**	**90.5**	**85.0**	**82.0**	**17.9**	**11.0**	**9.5**	**9.4**	**6.3**	**11.4**
	1%	74.0	69.1	65.8	6.9	3.8	3.2	3.0	1.3	3.3
Power 3 H_0 : static probit vs H_1 : dynamic logit										
	10%	98.5	97.5	94.8	34.7	20.6	20.1	18.1	11.1	24.6
Model I	**5%**	**97.1**	**93.8**	**91.8**	**24.7**	**12.4**	**12.8**	**11.1**	**5.5**	**13.4**
	1%	85.2	81.6	78.4	8.1	6.3	4.2	3.3	1.1	5.0
	10%	99.4	98.6	98.0	38.9	26.0	22.7	21.5	13.5	26.0
Model II	**5%**	**98.9**	**97.6**	**96.5**	**29.5**	**16.9**	**17.1**	**14.6**	**7.4**	**16.9**
	1%	96.0	93.2	91.2	12.6	7.1	7.6	3.5	1.8	6.3
	10%	98.5	97.2	95.7	36.3	24.0	20.6	21.2	12.2	24.2
Model III	**5%**	**96.1**	**93.8**	**91.8**	**26.0**	**14.6**	**14.4**	**11.9**	**8.0**	**15.4**
	1%	90.4	85.2	77.6	11.6	5.3	4.1	4.5	1.0	4.9
	10%	96.3	93.7	90.3	27.5	18.0	19.8	14.7	10.3	16.2
Model IV	**5%**	**93.0**	**89.2**	**86.9**	**14.1**	**13.0**	**12.5**	**8.5**	**5.8**	**10.4**
	1%	82.0	76.2	60.0	6.5	3.8	2.0	2.4	1.3	1.0

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

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