Semi-Parametric Independence Testing for Time Series of Counts and the Role of the Support

David Harris* and Brendan McCabe†

September 18, 2018

*Department of Economics, University of Melbourne, Australia. harrisd@unimelb.edu.au

†Management School, University of Liverpool, UK. Brendan.Mccabe@liverpool.ac.uk

Proposed running head: Independence Tests for Count Time Series

Corresponding author for proofs: David Harris, harrisd@unimelb.edu.au

Abstract

This paper considers testing for independence in a time series of small counts within an Integer Autoregressive (INAR) model, taking a semi-parametric approach that avoids any distributional assumption on the arrivals process of the model. The nature of the testing problem is shown to differ depending on whether or not the support of the arrivals distribution is the full set of natural numbers (as would be the case for Poisson or Negative Binomial distributions for example) or some strict subset of the natural numbers (such as for a Binomial or Uniform distribution). The theory for these two cases is studied separately.

For the case where the arrivals have support on the natural numbers, a new asymptotically efficient semi-parametric test, the effective score (Neyman-Rao) test, is derived. The semi-parametric Likelihood-Ratio, Wald and score tests are shown to be asymptotically equivalent to the effective score test, and hence also asymptotically efficient. Asymptotic relative efficiency calculations demonstrate that the semi-parametric effective score test can provide substantial power advantages over the first order autocorrelation coefficient, which is most commonly applied in practice.

For the case where the arrivals have support that is a strict subset of the natural numbers, the theory is considerably altered because the support of the observations becomes different under the null and alternative hypotheses. The semi-parametric Likelihood-Ratio, Wald and score tests become asymptotically degenerate in this case, while the effective score test remains valid. Remarkably, in this case the effective score test is also found to have power against local alternatives that shrink to the null at the rate T^{-1}. In rare cases where the arrival support is partly or totally known, additional tests exploiting this information are considered.

Finite sample properties of the tests in these various cases demonstrate the semi-parametric effective score test can provide substantial power advantages over the first order autocorrelation test implied by a parametric Poisson specification. The simulations also reveal situations in which the first order autocorrelation is preferable in finite samples, so a hybrid of the effective score and autocorrelation tests is proposed to capture most of the benefits of each test.

1 Introduction

Consider the $INAR(1)$ model (originally proposed by Al-Osh and Alzaid (1987) and McKenzie (1985)) for count data $\{y_t, t = 1, \ldots, T\}$, which has the form

$$y_t = \beta \circ y_{t-1} + u_t, \tag{1}$$

where the arrivals process (disturbance) u_t is i.i.d. with some distribution Π on a support $\mathcal{U} \subseteq \mathbb{N}$ ($\mathbb{N}$ is the set of non-negative integers). Here $\beta\circ$ is the usual binomial thinning operator whereby $\Pr(\beta \circ n = k) = \mathrm{Bi}(n, k; \beta)$ for $k = 0, 1, \ldots, n$ and $0 \le \beta < 1$. This model has a well defined physical interpretation as it may be thought of as a queue, a birth and death process or even a branching process with immigration. It is a natural model for any series of low counts that may be thought of as a "stock" variable. The $INAR$ model has found many applications in an increasing number of fields: for example in medicine one can consult Franke and Seligmann (1993), Pickands and Stine (1997) and Cardinal et al. (1999), in environmental studies Thyregod et al. (1999) and Pavlopoulos and Karlis (2008), in commerce Bockenholt (1999) and Gourieroux and Jasiak (2004), and in economics Brännäs and Hellstrom (2001) and Rudholm (2001).

Testing for independence in y_t across time involves testing

$$H_0 : \beta = 0 \text{ versus } H_1 : \beta > 0.$$

Testing for independence in counts has already been discussed by, for example, Venkataraman (1982) and Mills and Seneta (1989), and in the context of the $INAR$ model, Freeland (1998), Jung and Tremayne (2003) and Sun and McCabe (2013) have postulated some parametric forms for u_t.

In this paper we do not assume that u_t has a known distribution such as a Poisson or negative binomial or even a member of a family such as that of Katz (1965). Instead a semi-parametric approach is adopted, in which the arrivals distribution Π is not parameterised. Estimation theory in this setting has been developed by Drost *et al.* (2009) when estimating $\beta \in (0, 1)$, and extended by McCabe *et al.* (2011) to theory for forecast distributions. The current work makes the non-trivial extension of this theory to include the boundary case $\beta = 0$.

Without a known parametric form for Π, the support $\mathcal{U}$ is also unknown. It may be that u_t is supported on $\mathcal{U} = \mathbb{N}$ (such as for a Poisson or Negative Binomial distribution) or that u_t is supported on a finite set (such as a Binomial, Uniform or truncated distribution). The support may also contain gaps if the arrivals are composed of a mixture distribution, such as may be formed from several unobserved streams or if a censoring mechanism is at work, which may also in extreme cases mimic outlier behaviour. In this paper we show that the semi-parametric hypothesis testing theory differs substantially depending on whether $\mathcal{U} = \mathbb{N}$ or $\mathcal{U} \subset \mathbb{N}$, and therefore focus separately on these two cases. The role of the structure of the support is novel here and does not arise in this form in standard time series tests of dependence (eg Box-Pierce (1970), Whang (1998) and Shao (2011) among others). Semi- and non-parametric specification tests (eg Rodríguez-Póo, Sperlich and Vieu (2015)) or (conditional) independence tests (eg Huang, Sun and White (2016)) routinely invoke finite support assumptions on observable variables, but such regularity conditions are not as fundamental to the structure of the optimal testing problem as the arrivals support is in the semi-parametric INAR model.

In section 2 we focus on the case where $\mathcal{U} = \mathbb{N}$. We extend the result of Drost et al. (2009) on the local asymptotic normality (LAN) of the likelihood ratios of (1) to include $\beta = 0$ and hence, in the manner of Choi et al. (1996, CHS), derive an efficient (i.e. asymptotically uniformly most powerful) test of $H_0 : \beta = 0$ versus $H_1 : \beta > 0$, with Π being considered as an infinite-dimensional nuisance parameter. The resulting effective score test has asymptotically normal distribution theory under the null and local alternatives. The classical score, Wald (W) and likelihood ratio (LR) tests, modified to account for the null lying on the boundary of the parameter space for β, are shown to be equivalent to the effective score test and hence also efficient. A commonly used test is based on the lag-1 autocorrelation, which is asymptotically efficient for Poisson arrivals but otherwise inefficient. Under negative binomial arrivals, a numerical comparison of the Pitman asymptotic relative efficiency (ARE) reveals potentially very large efficiency gains for the new effective score test relative to the lag-1 autocorrelation test.

In section 3 we focus on the case where $\mathcal{U}$ is a strict subset of $\mathbb{N}$, so that it is finite and/or contains gaps. The structure of $\mathcal{U}$ in this case is made specific in Assumption 3 below. Non-standard theory applies in this case because the support for y_t is dependent on the parameter under test, being $\mathcal{U}$ under $\beta = 0$ and a super-set of $\mathcal{U}$ when $\beta > 0$. Under the null of independence, the three classical statistics are shown to exhibit degenerate asymptotic behaviour whilst the effective score remains asymptotically normal. Remarkably, the effective score test is also shown to have power against local alternatives that converge to the null at rate T^{-1}, much faster than the usual $T^{-1/2}$ rate, implying a greater than usual ability to distinguish $\beta > 0$ from $\beta = 0$. While the details are technical, intuitively the reason is that when the support of H_0 is a strict subset of H_1, observations that lie outside of the support of H_0 are highly informative about the hypotheses, *additional* to any correlation present in the data. It is noteworthy that *it is not necessary to know* the structure of $\mathcal{U}$ for this power to obtain.

In section 4 we report on simulation studies that confirm the asymptotic local power findings, showing the superiority of the effective score over the correlation test is especially pronounced for larger sample sizes when the arrivals support is restricted. Nevertheless in some cases the correlation coefficient may still perform better in small samples. This motivates a hybrid test, which combines the small sample advantages of the parametric correlation test with the large sample adaptiveness of the semi-parametric effective score. This hybrid test rejects if either the correlation or the effective score test individually rejects, with the sizes of the individual tests modified so that the overall size of the hybrid is controlled. Simulations indicate that the hybrid test does indeed retain the advantages of the individual components at little cost and hence provides the best available independence test for practical count data time series analysis.

In section 5, we consider the rare cases where the structure of $\mathcal{U}$ is totally or partially known. Tests exploiting such information are proposed, and shown in some special cases to have even greater power than the effective score test. However they are also shown to be practically inapplicable in general.

Proofs of the main mathematical results are provided in Appendix A, and some supplementary results and proofs are collected in the online Appendix B, available at Cambridge Journals Online (`journals.cambridge.org/ect`).

2 Efficient testing with standard support

Assume we have an integer valued time series generated by the Markovian process (1). The following conditions are used throughout the paper.

Assumption 1

- *(a) The arrivals u_t form an i.i.d. sequence whose support is denoted $\mathcal{U}$, where $\mathcal{U} \subseteq \mathbb{N}$.*
- *(b) For every t, u_t has distribution $\Pi = \{\pi_k\}_{k=0}^{\infty}$ where $\pi_k = \Pr(u_t = k)$, and Π is such that $E(|u_t|^5) < \infty$.*
- *(c) $\beta \in [0, 1)$ and the thinning operator sequence $\{\beta \circ y_{t-1}\}$ is independent of the arrivals sequence $\{u_t\}$.*
- *(d) The initial value y_0 is drawn from the distribution Π and is independent of the thinning and arrivals sequences.*

The mean and variance of the arrival process u_t can then be defined respectively as $\mu_u = E(u_t) = \sum_{k=1}^{\infty} k\pi_k$ and $\sigma_u^2 = \text{var}(u_t) = \sum_{k=1}^{\infty} k^2\pi_k - \mu_u^2$.

This section derives semi-parametric likelihood-based tests for $H_0 : \beta = 0$ against $H_1 : \beta > 0$ in equation (1) and proves their efficiency under the following standard support assumption.

Assumption 2 *The support of u_t satisfies $\mathcal{U} = \mathbb{N}$.*

An implication of this support assumption is that π_k is strictly positive for every $k \in \mathbb{N}$, as would be the case for the Poisson or negative binomial distributions for example.

2.1 The effective score test

The semi-parametric log-likelihood is

$$\log L_T(\beta, \Pi) = \sum_{t=1}^{T} \log \sum_{k=0}^{y_t \wedge y_{t-1}} \mathrm{Bi}(y_{t-1}, k; \beta)\, \pi_{y_t - k}, \tag{2}$$

and the MLE of Π defined after imposing the null hypothesis on the log-likelihood is

$$\hat{\Pi} = \arg \max_{\Pi \in \mathcal{P}_{\mathcal{Y}_T}} \log L_T(0, \Pi) \tag{3}$$

where $\mathcal{P}_{\mathcal{Y}_T} = \{\Pi : 0 < \pi_j \leq 1 \text{ for } j \in \mathcal{Y}_T \text{ and } \pi_j = 0 \text{ for } j \notin \mathcal{Y}_T\}$ and $\mathcal{Y}_T$ is the empirical support of $(y_1, \ldots, y_T)$. This estimator simply consists of the sample probabilities

$$\hat{\pi}_j = T^{-1} \sum_{t=1}^{T} 1_j(y_t)$$

for $j \in \mathcal{Y}_T$, where $1_j(y_t)$ is the indicator function for the event $y_t = j$. By definition these sample probabilities will satisfy $\hat{\pi}_j > 0$ for every $j \in \mathcal{Y}_T$, while the definition of the estimator $\hat{\Pi}$ also implies that $\hat{\pi}_j = 0$ for any $j \notin \mathcal{Y}_T$ (which includes $j = -1$, which will be relevant when $y_t = 0$).

Following the definitions of CHS, the standardised effective score test statistic can be shown to be

$$\hat{\xi}_T = \frac{\hat{S}^*_{T,\beta}}{\hat{\omega}} \tag{4}$$

where the numerator of (4) consists of

$$\hat{S}^*_{T,\beta} = T^{-1/2} \sum_{t=1}^{T} (y_{t-1} - \hat{\mu}_u)(\hat{g}_t - 1),$$

with $\hat{g}_t = \hat{\pi}_{y_t - 1}/\hat{\pi}_{y_t}$ and $\hat{\mu}_u = \bar{y}$. The denominator of (4) is

$$\hat{\omega}^2 = \hat{\sigma}_u^2 \cdot \hat{\sigma}_g^2, \tag{5}$$

with estimators $\hat{\sigma}_u^2 = T^{-1} \sum_{t=1}^{T} (y_{t-1} - \hat{\mu}_u)^2$, $\hat{\sigma}_g^2 = T^{-1} \sum_{t=1}^{T} (\hat{g}_t - \bar{g})^2$ and $\bar{g} = T^{-1} \sum_{t=1}^{T} \hat{g}_t$. The derivation of this effective score test statistic and its asymptotic properties are summarised in the following theorem.

Theorem 1 *Under Assumptions 1 and 2 and the local sequence* $\beta = \beta_T(h_\beta) = T^{-1/2} h_\beta$, $h_\beta \geq 0$,

$$\hat{\xi}_T \rightsquigarrow N(\omega h_\beta, 1),$$

where $\omega^2 = \sigma_u^2 (\sum_{k=1}^{\infty} \pi_{k-1}^2/\pi_k - 1)$ *and "*$\rightsquigarrow$*" denotes weak convergence. The effective score test that rejects* H_0 *for* $\hat{\xi}_T > z_\alpha$*, where* z_α *is the* $100(1-\alpha)\%$ *percentile of the standard normal distribution, is asymptotically efficient.*

The proof of the Theorem is given in section A.1, and involves the following steps.

1. First, it is shown that the log-likelihood ratios of the semi-parametric INAR model have the LAN property under Assumption 2. This extends the LAN result of Drost *et al.* (2009) for $0 < \beta < 1$ to include the case $\beta = 0$.

2. The LAN property permits standard asymptotic inference to be carried out. In this case the effective score test defined by CHS is derived based on knowledge of the nuisance parameters, i.e. the arrivals probabilities $\{\pi_k\}$. The effective score statistic is then shown to have an asymptotically standard normal null distribution, and to provide an asymptotically uniformly most powerful test against alternatives of the form $\beta > 0$.

3. The effective score test is infeasible because it depends on nuisance parameters, specifically $\{\pi_k\}$ and the associated means and variances μ_u, σ_u^2 and σ_g^2. It is then shown that replacing these parameters with MLE's defined under H_0 gives the feasible statistic $\hat{\xi}_T$ in (4) and that $\hat{\xi}_T - \xi_T \xrightarrow{p} 0$ under both the null and local alteratives. Hence the feasible effective score test is shown to have the asymptotic properties claimed in the statement of Theorem.

The theorem shows that the asymptotic local power of the effective score test is $1 - \Phi\left(z_\alpha - \omega h_\beta\right)$, Φ being the distribution function of the standard normal distribution. The parameter ω^2 depends on the arrivals variance σ_u^2 and the sum $\sum_{k=1}^{\infty} \pi_{k-1}^2/\pi_k$. The latter is equal to $E(\pi_{y_t-1}^2/\pi_{y_t}^2)$, which arises as part of the limit of the sample variance $\hat{\sigma}_g^2$ of $\hat{g}_t = \hat{\pi}_{y_t-1}/\hat{\pi}_{y_t}$.

2.2 The Wald, LR and score tests

The asymptotic distributions of the classical W, LR and score tests can also be derived with appropriate modifications for the fact that $\beta = 0$ is on the boundary of the parameter space. The Wald and LR tests are shown to be asymptotically equivalent to the effective score test. They require estimation of the model (1) under the alternative and in this situation estimators for β must lie in the parameter space $B = [0, 1)$. The MLE is

$$\left(\tilde{\beta}, \tilde{\Pi}\right) = \arg \max_{\beta \in B, \Pi \in \mathcal{P}_{\mathcal{Y}_T}} \log L_T\left(\beta, \Pi\right).$$

The statistics are the Wald, $W_T = T\hat{\omega}^2\tilde{\beta}^2$, the likelihood ratio $\Lambda_T = 2(\log L_T(\tilde{\beta}, \tilde{\Pi}) - \log L_T(0, \hat{\Pi}))$ and the score $\Psi_T = \hat{S}_{T,\beta}^2/\hat{\omega}^2$. Here $\hat{S}_{T,\beta} = T^{-1/2}\sum_{t=1}^{T} y_{t-1}\left(\hat{g}_t - 1\right)$ and this differs from the effective score by the lack of a centering factor for y_{t-1}. Also there is the one-sided score statistic $\Psi_T^+ = \hat{S}_{T,\beta}/\hat{\omega}$, that differs from $\hat{\xi}_T$ by using the raw score of β, not the effective score. The following theorem provides the asymptotic distribution, under local alternatives, of the MLE of $\tilde{\beta}$ and of the various test statistics.

Theorem 2 *Under the conditions of Theorem 1 :*
(i) $\sqrt{T}\tilde{\beta} \rightsquigarrow (Z_\beta \vee 0)$ *where* $Z_\beta \sim N\left(h_\beta, \omega^{-2}\right)$,
(ii) $\Lambda_T, W_T \rightsquigarrow (Z_\Lambda \vee 0)^2$ *where* $Z_\Lambda \sim N\left(\omega h_\beta, 1\right)$,
(iii) $\Psi_T \rightsquigarrow Z_\Lambda^2$ *and* $\Psi_T^+ - \hat{\xi}_T \xrightarrow{p} 0$.

The proof of Theorem 2, given in section A.2, follows standard arguments once the LAN property in item 1 of the proof of Theorem 1 above has been established.

Theorem 2 implies that tests of asymptotic level α are provided by rejecting the null for $\Lambda_T, W_T > z_\alpha^2$ (z_α^2 being the usual α level χ_1^2 critical value), $\Psi_T > z_{\alpha/2}^2$ and $\Psi_T^+ > z_\alpha$. Part *(iii)* shows that the one-sided score test, Ψ_T^+, is efficient as it is asymptotically equivalent to the effective score $\hat{\xi}_T$. An inspection of the proof shows that this is also true for Λ_T and W_T as they are based directly on the MLE $\tilde{\beta}$ and hence the Wald and LR tests are efficient as well. The test Ψ_T based on the square of the score (i.e. the textbook two-sided score test) is not efficient because it does not allow for the one-sided nature of the hypotheses.

2.3 Comparison with an autocorrelation test

The effective score test can be compared with that based on the standardized first order correlation coefficient

$$\hat{\rho}_T = \frac{T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1}-\bar{y}\right)\left(y_t-\bar{y}\right)}{T^{-1}\sum_{t=1}^{T}\left(y_{t-1}-\bar{y}\right)^2},$$

which has asymptotic distribution $\hat{\rho}_T \rightsquigarrow N\left(h_\beta, 1\right)$ under the conditions of Theorem 1. This test is efficient when u_t is Poisson, although the mean-variance equality is not imposed in the denominator of $\hat{\rho}_T$ so that the test remains asymptotically correctly sized for other arrivals distributions. (Obviously $\hat{\rho}_T$ also emerges from a normally distributed AR(1) model.)

Since, from Theorem 1, we know that $\hat{\xi}_T \rightsquigarrow N\left(\omega h_\beta, 1\right)$, it follows that the Pitman ARE of $\hat{\xi}_T$ relative to $\hat{\rho}_T$ is given by ω. To illustrate the role of ω, Table 1 shows the ARE of $\hat{\xi}_T$ relative to $\hat{\rho}_T$ when u_t has a Negative Binomial distribution specified by

$$\pi_k = \binom{k+r-1}{r-1}(1-p)^r p^k,\ k = 0, 1, 2, \ldots$$

for $0 < p < 1$ and $r = 1, 2, \ldots$, which satisfies Assumption 2. When $r = 1$ this corresponds to the geometric distribution and as r increases the Negative Binomial distribution approaches the Poisson.

Compared to $\hat{\rho}_T$, the effective score test shows moderate power gains for $p = 0.25$, rising to very large power gains for $p = 0.75$ and small r. Notice ω is different from the over-dispersion, which is commonly used to describe deviations from the canonical Poisson distribution. The over-dispersion (variance relative to the mean σ_u^2/μ_u) for the Negative Binomial distribution depends

Table 1: ARE of $\hat{\rho}_T$ relative to $\hat{\xi}_T$, NegBin(r,p) arrivals

r	$p = 0.25$	$p = 0.50$	$p = 0.75$
1	1.15	1.41	2.00
2	1.10	1.24	1.50
5	1.05	1.10	1.18
10	1.02	1.05	1.08

only on p and is equal to 1.33 for $p = 0.25$, 2 for $p = 0.50$ and 4 for $p = 0.75$. Therefore the asymptotic local power of the effective score test tends to increase with the over-dispersion, although this is not the only factor that affects the power of the test, since r is also an important determinant as it influences the probabilities of adjacent members of the support. The power of the $\hat{\xi}_T$ test approaches that of the $\hat{\rho}_T$ test as r increases.

3 Extension to General Support

The behaviour of the statistics for testing $H_0 : \beta = 0$ changes once $\mathcal{U} \neq \mathbb{N}$. This occurs when $\mathcal{U}$ is a finite set and/or when $\mathcal{U}$ contains "gaps". In this section, gaps are formally defined along with some other classes of sets that prove useful when analysing local power. The behaviour of the optimal tests derived under $\mathcal{U} = \mathbb{N}$ is analysed when in fact $\mathcal{U}$ does contain gaps, and it turns out that only the effective score test $\hat{\xi}_T$ has non-degenerate asymptotic behaviour in the presence of gaps. The effective score test remains asymptotically normal but surprisingly has local power against alternatives that converge to the null at rate T^{-1}.

The concept of a finite support and/or support with gaps is formalised in the following assumption.

Assumption 3 *The support $\mathcal{U}$ of u_t is such that*

(a) there exists at least one integer $k \geq 1$ such that $k - 1 \in \mathcal{U}$ but $k \notin \mathcal{U}$,

(b) there exists at least one integer $k \geq 1$ such that $k - 1 \in \mathcal{U}$ and $k \in \mathcal{U}$.

Part (a) of the Assumption is the defining feature — a support with this structure will be finite and/or have at least one gap. In this case the support of y_t in (1), denoted Y_t, is dependent on β. In particular $Y_t = \mathcal{U}$ for all t when $\beta = 0$, but $Y_t \supset \mathcal{U}$ when $\beta > 0$, so that the support of y_t differs between the null and alternative hypotheses under Assumption 3(a). As shown in this section, the asymptotic theory of the testing problem becomes non-standard as a consequence.

A variety of possible structures for $\mathcal{U}$ are included in Assumption 3(a). A leading example is where $\mathcal{U}$ is finite.

Example 1 *(Finite support) Let* $\mathcal{U} = \{0, 1, \ldots, M\}$ *for* $M < \infty$*. Then* $Y_t = \mathcal{U}$ *when* $\beta = 0$ *and* $Y_t = \{0, 1, \ldots, M + tM\}$ *when* $\beta > 0$*, where the support of* y_0 *is assumed to be* $\mathcal{U}$ *in Assumption 1(d).*

This example includes arrivals distributions such as the binomial and the uniform, and M will generally be unknown.

Example 2 *(Support with a gap) Let* $\mathcal{U} = \{0, 1, \ldots, M_1, M_1 + 1 + g, \ldots, M_2\}$ *where* g *is a strictly positive integer. In this case there is a gap in the support in which at least one integer following* M_1 *has probability zero. It is possible that* M_2 *may be either finite or infinite.*

Illustrations of supports with this structure can be found in the discussion following Theorem 4 below.

Part (b) of Assumption 3 requires that there be at least one pair of consecutive integers contained in $\mathcal{U}$. Example 1 satisfies this for $M \geq 1$ and Example 2 satisfies it provided $M_1 \geq 1$ and/or $M_2 \geq M_1 + 1 + g$. The following example, in which arrivals occur in pairs, satisfies Assumption 3(a) but not 3(b).

Example 3 *(Infinite support with gaps – "Noah's Ark arrivals") Let* $\mathcal{U} = \{0, 2, 4, \ldots\}$*. Then* $Y_t = \mathcal{U}$ *when* $\beta = 0$ *and* $Y_t = \mathbb{N}$ *when* $\beta > 0$*.*

A support such as this one without any consecutive integers leads to degenerate behaviour under the null in the effective score test, as discussed following Theorem 3 below.

It is convenient to have notation for various subsets of integers defined from $\mathcal{U}$ and its gaps. First define $\mathcal{U}^{(0)} = \mathcal{U}$. We now wish to consider those integers that constitute the first element(s) of each of the gap(s) in $\mathcal{U}^{(0)}$ i.e. those i that satisfy

$$\mathcal{U}^{(1)} = \left\{ i \notin \mathcal{U}^{(0)} : i - 1 \in \mathcal{U}^{(0)} \right\}.$$

Thus in Example 1 we have $\mathcal{U}^{(1)} = \{M + 1\}$, in Example 2 we have $\mathcal{U}^{(1)} = \{M_1 + 1, M_2 + 1\}$ (the second element being omitted if M_2 is infinite), while in Example 3 we have $\mathcal{U}^{(1)} = \{1, 3, 5, \ldots\}$. The set $\mathcal{U}^{(1)}$ turns out to be particularly influential for the asymptotic local power of the effective score test under Assumption 3.

The constants $\pi^{(0)} = \sum_{j \in \mathcal{U}^{(0)}} \pi_{j-1}$ and $\pi^{(1)} = \sum_{j \in \mathcal{U}^{(1)}} \pi_{j-1}$ are of particular relevance to the asymptotic analysis under Assumption 3, and satisfy $\pi^{(0)} + \pi^{(1)} = 1$. So, in Example 1, $\pi^{(0)} = 1 - \pi_M$ and $\pi^{(1)} = \pi_M$ while in Example 3, $\pi^{(0)} = 0$ and $\pi^{(1)} = 1$.

3.1 Asymptotic Properties under the Null when $\mathcal{U} \neq \mathbb{N}$

Theorems 1 and 2 showed that, under Assumption 2, the effective score test is asymptotically equivalent to the usual one-sided score, Wald and likelihood ratio tests. This equivalence breaks down when Assumption 2 is replaced by Assumption 3. The following theorem shows this under the null hypothesis.

Theorem 3 *Under Assumptions 1 and 3 and* $H_0 : \beta = 0$
(i) $\hat{\xi}_T \rightsquigarrow N(0,1)$,
(ii) $\Psi_T \xrightarrow{p} +\infty$, $\Psi_T^+ \xrightarrow{p} -\infty$,
(iii) $\Lambda_T, W_T \xrightarrow{p} 0$.

Part *(i)* of the Theorem shows that the asymptotic null distribution of $\hat{\xi}_T$ remains standard normal under Assumption 3. A test of asymptotic level α is therefore specified by rejecting the null for $\hat{\xi}_T > z_\alpha$, without knowledge of whether or not Assumption 2 or 3 holds. The three classical statistics, however, have degenerate behaviour under Assumption 3. The score tests based on Ψ_T and Ψ_T^+ have asymptotic size of one if applied with the usual rejection criterion, while the Wald and likelihood ratio tests have asymptotic size of zero.

The source of the difference between $\hat{\xi}_T$ and the other statistics lies in the difference between the properties of the score $\hat{S}_{T,\beta}$ and the effective score $\hat{S}^*_{T,\beta}$. In particular it can be shown using the arguments in the proof of Theorem 3 that

$$T^{-1/2}\hat{S}_{T,\beta} = E\left(y_{t-1}\left(\frac{\pi_{y_t-1}}{\pi_{y_t}} - 1\right)\right) + O_p(T^{-1/2}),$$

and then it follows that

$$E\left(y_{t-1}\left(\frac{\pi_{y_t-1}}{\pi_{y_t}} - 1\right)\right) = \mu_u \sum_{j \in \mathcal{U}^{(0)}} \left(\frac{\pi_{j-1}}{\pi_j} - 1\right)\pi_j = \mu_u\left(\pi^{(0)} - 1\right) = -\mu_u \pi^{(1)}.$$

This expectation is not zero under Assumption 3 since $\pi^{(0)} = \sum_{j \in \mathcal{U}^{(0)}} \pi_{j-1} < 1$. This implies that

$$\hat{S}_{T,\beta} = -T^{1/2}\mu_u \pi^{(1)} + O_p(1) \xrightarrow{p} -\infty$$

under Assumption 3, which explains the properties of the score tests given in part *(ii)* of Theorem 3. The effective score, however, satisfies

$$T^{-1/2}\hat{S}^*_{T,\beta} = E\left((y_{t-1} - \mu_u)\left(\frac{\pi_{y_t-1}}{\pi_{y_t}} - 1\right)\right) + O_p(T^{-1/2}),$$

in which

$$E\left((y_{t-1} - \mu_u)\left(\frac{\pi_{y_t-1}}{\pi_{y_t}} - 1\right)\right) = E\left((y_{t-1} - \mu_u)\right)\pi^{(1)} = 0.$$

The centering of y_{t-1}, the effect of which is asymptotically negligible under Assumption 2, is vital to the construction of a non-degenerate test under Assumption 3.

More generally, following on from the score $\hat{S}_{T,\beta}$ not having zero mean (even asymptotically), the usual LAN (Local Asymptotic Normality) property of the log-likelihood ratio does not hold in this model under Assumption 3. As is well-known (and discussed in detail in CHS for example), this LAN property provides the fundamental underpinning for standard likelihood based asymptotic inference. In the proof of Theorem 1 the LAN property was shown to hold under Assumption 2 (see section A.1.1 in the Appendix for details), but this no longer applies under Assumption 3. As a result the usual properties and implications of scores and likelihood ratio expansions can no longer be relied upon under Assumption 3.

Theorem 3 shows that only the effective score statistic $\hat{\xi}_T$ provides a test with non-degenerate behaviour under Assumption 3. There is an important role for Assumption 3(b) in this result, because if this does not hold (such as if the support is given in Example 3) then it must be the case that $\hat{\pi}_{y_t-1} = 0$ for every t, resulting in $\hat{g}_t = 0$ for every t and hence $\hat{S}^*_{T,\beta} = 0$. It is for this reason that we exclude this situation from Assumption 3, but if ever it is found in an application that the sample support $\mathcal{Y}_T$ contains no consecutive integers then we define $\hat{\xi}_T = 0$ and do not reject H_0.

3.2 Asymptotic Local Power $\mathcal{U} \neq \mathbb{N}$

The asymptotic power of the effective score test of $\beta = 0$ under Assumption 3 can be shown to be determined by those observations, if any, that fall outside the arrivals support $\mathcal{U}^{(0)}$. Even though it is not assumed that $\mathcal{U}^{(0)}$ is known, such observations carry a lot of information in this hypothesis testing problem since $y_t \notin \mathcal{U}^{(0)}$ is impossible under the null hypothesis. In fact the *possibility* of such observations permits the test to have non-degenerate asymptotic local power against the local sequence

$$\beta_T = T^{-1} h_\beta, \;\; h_\beta > 0, \tag{6}$$

which approaches the null hypothesis faster than the usual $T^{-1/2}$ rate. The reason is that the behaviour of the effective score depends importantly on $\hat{g}_t = \hat{\pi}_{y_t-1}/\hat{\pi}_{y_t}$, in which the denominator $\hat{\pi}_{y_t}$ has different asymptotic properties depending on whether $y_t \in \mathcal{U}^{(0)}$ or not. The magnitude of the T^{-1} rate will be better appreciated after the results of Lemma 1 below are presented.

The main determinant of the asymptotic local power of the test is the number of observations that remain in $\mathcal{U}^{(1)}$ asymptotically. To quantify this, define the counting process which counts the number of times that the number i occurs in the sample by

$$N_{T,i} = \sum_{t=1}^{T} 1_i \left(y_t \right)$$

for any i. The following lemma gives a law of small numbers for these processes.

Lemma 1 *Under Assumptions 1 and 3 and* $\beta_T = T^{-1}h_\beta$, $h_\beta > 0$, *as* $T \to \infty$

$$\left\{N_{T,i}, i \in \mathcal{U}^{(1)}\right\} \rightsquigarrow \left\{N_i, i \in \mathcal{U}^{(1)}\right\},$$

where $\left\{N_i, i \in \mathcal{U}^{(1)}\right\}$ *is a set of independent Poisson random variables with respective parameters* $h_\beta \mu_u \pi_{i-1}$. *In addition,*

$$\Pr\left(y_t \in \mathcal{U}^{(0)} \text{ for all } t = 1, \ldots, T\right) \to \exp\left(-h_\beta \mu_u \pi^{(1)}\right), \tag{7}$$

$$\Pr\left(y_t \in \mathcal{U}^{(1)} \text{ for any } t = 1, \ldots, T\right) \to 1 - \exp\left(-h_\beta \mu_u \pi^{(1)}\right). \tag{8}$$

Consider, for example, $\mathcal{U}^{(0)} = \{1, 2, ..., M\}$ and let $N_{T,M+1}$ be the number of y_t's in the sample of size T which equal $M+1$. Then, under these local alternatives, $\Pr\left[N_{T,M+1} = k\right]$, $k = 0, 1, 2...$ may be computed asymptotically from the Poisson distribution $e^{-\lambda}\lambda^k/k!$ with mean $\lambda = h_\beta \mu_u \pi_M$.

The limit probabilities in (7) and (8) show that there are just two possibilities under (6): (a) asymptotically the sample $\{y_1, \ldots, y_T\}$ is restricted to $\mathcal{U}^{(0)}$, which happens with probability $\exp\left(-h_\beta \mu_u \pi^{(1)}\right)$, or (b) the sample contains at least one element of $\mathcal{U}^{(1)}$, which happens with probability $1 - \exp\left(-h_\beta \mu_u \pi^{(1)}\right)$. No other outcome is possible in the limit. The lemma includes Assumption 2 as a degenerate case since under that assumption $\Pr\left(\mathcal{Y}_T = \mathcal{U}^{(0)}\right) = 1$ for all t, which is consistent with (7) since $\pi^{(1)} = 0$ under Assumption 2.

The intuition for the T^{-1} rate in (6) comes from (7). It can be deduced from the proof of Lemma 1 that, approximately,

$$\Pr\left(y_t \in \mathcal{U}^{(0)} \text{ for all } t = 1, \ldots, T\right) \approx \left(1 - \beta_T \mu_u \pi^{(1)}\right)^T$$

which converges to the exponential function when $\beta_T = O\left(T^{-1}\right)$.

The asymptotic local power of the effective score test under (6) depends on the limiting counting process $\left\{N_i, i \in \mathcal{U}^{(1)}\right\}$, and in particular whether or not $N_i > 0$ for any $i \in \mathcal{U}^{(1)}$. We define the indicator random variable

$$Q = 1\left(N_i > 0 \text{ for any } i \in \mathcal{U}^{(1)}\right) \tag{9}$$

for this event. If $Q = 0$ then all observations lie in $\mathcal{U}^{(0)}$ asymptotically while if $Q = 1$ at least one observation lies in $\mathcal{U}^{(1)}$.

Theorem 4 *Under Assumptions 1 and 3 and* $\beta_T = T^{-1}h_\beta$, $h_\beta > 0$, *as* $T \to \infty$

$$\hat{\xi}_T \rightsquigarrow Z_Q + X_Q$$

where $Z_Q \sim N(0,1)$ *and* $X_Q = 0$ *if* $Q = 0$*, and* $Z_Q = 0$ *and* $X_Q = X$ *if* $Q = 1$*, where* Q *is given in (9) and* X *is a random variable which depends on the underlying law of the observations (i.e.* Π *and* h_β*).*

The theorem shows that if the sample is asymptotically restricted to $\mathcal{U}^{(0)}$ (i.e. $Q = 0$), which occurs with probability $\exp\left(-h_\beta \mu_u \pi^{(1)}\right)$, then $\hat{\xi}_T$ is asymptotically standard normally distributed, implying the effective score test has asymptotic power equal to size under (6). If some observations remain in $\mathcal{U}^{(1)}$ asymptotically (i.e. $Q = 1$), the asymptotic distribution of $\hat{\xi}_T$ is provided by the distribution of X, which is non-standard and depends on Π and h_β in a complicated way. A representation of X is provided in the proof of the Theorem, but the extent of the power arising from this non-standard distribution needs to be approximated by simulation, which is investigated in Section 4.2. Note the effect of increasing the local alternative parameter h_β is to increase the probability that $Q = 1$ (given by $1 - \exp\left(-h_\beta \mu_u \pi^{(1)}\right)$), so that larger deviations from the null increase the potential for non-trivial power against $\beta_T = T^{-1} h_\beta$.

In addition to h_β, the parameters $\pi^{(1)}$ and μ_u determine the probability of $Q = 1$ and hence the asymptotic local power of the test. The parameter $\pi^{(1)}$ is of particular interest since it reflects the role of the finiteness/gaps in the arrivals distribution. To illustrate, consider a Binomial distribution with parameters $(6, 0.5)$ which has finite support $\mathcal{U}^{(0)} = \{0,1,2,3,4,5,6\}$ and a mean of $\mu_u = 3$. In this case $\mathcal{U}^{(1)} = \{7\}$ and $\pi^{(1)} = \Pr(u_t = 6) = \mathrm{Bi}(k = 6; n = 6, p = 0.5)$, which is 0.015625. Taking $h_\beta = 1$ for this illustration, this produces $\Pr(Q = 1) = 1 - \exp(-1 \times 3 \times 0.015625) \approx 0.04579$, implying relatively low asymptotic local power.

Suppose, however, we keep the same binomial probabilities as before but relabel the support to $\mathcal{U}^{(0)} = \{0,1,2,3,6,7,8\}$. There is now a gap in the distribution at $\{4,5\}$, implying that $\mathcal{U}^{(1)} = \{4,9\}$ and hence

$$\begin{aligned}\pi^{(1)} &= \pi_3 + \pi_8 \\ &= \mathrm{Bi}(6,3;0.5) + \mathrm{Bi}(6,6;0.5) \\ &= 0.3125 + 0.015625 \\ &= 0.328125.\end{aligned}$$

The mean of this relabelled distribution is now $\mu_u = 3.6875$ and the resulting probability of $Q = 1$ is $1 - \exp\left(-h_\beta \mu_u \pi^{(1)}\right) \approx 0.7018$, implying the potential for much higher asymptotic local power in this case. This is largely due to the introduction of the gap at $\{4,5\}$ into the distribution producing a larger value of $\pi^{(1)}$. However there is evidently also a change in μ_u in this example,[1] from 3 to 3.6875. As an alternative illustration to control for μ_u, suppose we maintain the support

[1] We are grateful to a referee for this observation.

$\mathcal{U}^{(0)} = \{0,1,2,3,6,7,8\}$, but now use the probabilities from a $\text{Bi}(6,k;0.4271)$ distribution for $\{\pi_k\}$ instead of $\text{Bi}(6,k;0.5)$. This has the effect of producing $\mu_u = 3$ (i.e. the same as the $\text{Bi}(6,k;0.5)$ distribution supported on $\{0,1,2,3,4,5,6\}$) and now

$$\begin{aligned} \pi^{(1)} &= \pi_3 + \pi_8 \\ &= \text{Bi}\,(6,3;0.4271) + \text{Bi}\,(6,6;0.4271) \\ &\approx 0.2991. \end{aligned}$$

The probability of $Q=1$ in this case is now $1-\exp\left(-h_\beta \mu_u \pi^{(1)}\right) \approx 0.5923$, which is again very much higher than the standard Binomial case (0.04579). This illustrates the importance of the structure of the gaps in the arrivals distribution in determining the asymptotic local power of the test.

It is not whether the arrivals support is finite or infinite that matters, as the following example illustrates. Consider a Poisson distribution with parameter 4, but supported on $\mathcal{U}^{(0)} = \{0,1,3,4,6,7,\ldots\}$, so that gaps appear at $\mathcal{U}^{(1)} = \{2,5,8,\ldots\}$. This results in a relatively large value of $\pi^{(1)}$ given by

$$\pi^{(1)} = \Pr\left(u_t = 1\right) + \Pr\left(u_t = 4\right) + \Pr\left(u_t = 7\right) + \ldots \approx 0.3337$$

and illustrates that it is the locations of any gaps relative to points of high probability mass in the support that determine $\pi^{(1)}$ and hence test power.

Finally, Assumption 3(b) requires that $\mathcal{U}$ contains at least one pair of consecutive integers, which is likely to be most empirically relevant. If, however, this part of the Assumption does not hold, we have defined above that $\hat{\xi}_T = 0$ under H_0, which provides a test with size equal to zero. This unusual size property is not a problem here (in fact it can be viewed as a nice feature not to be risking Type I errors), since it can be seen in the proof of Theorem 4 that $\hat{\xi}_T \rightsquigarrow X_Q$ under the local alternatives $\beta_T = T^{-1}h_\beta$. That is, if all observations are asymptotically restricted to $\mathcal{U}^{(0)}$ then the statistic $\hat{\xi}_T$ remains exactly zero asymptotically, while if any observations remain in $\mathcal{U}^{(1)}$ asymptotically then $\hat{\xi}_T$ is no longer degenerate and in fact, because its limit is given by X in this case, will exhibit qualitatively similar asymptotic power properties to the situation where Assumption 3(b) holds. Therefore the effective score test is valid to apply under any *unknown* structure of the arrivals distribution.

4 Finite Sample Properties

4.1 $\mathcal{U} = \mathbb{N}$

A Monte Carlo experiment was carried out to illustrate some of the features of the finite sample properties of the effective score test $\hat{\xi}_T$ and the first order autocorrelation test ($\hat{\rho}_T$). In each case the data generating process is (1) with $T = 100, 200, 400, 800$ and u_t i.i.d. with a distribution satisfying Assumption 2. In each experiment the sizes of the two tests are simulated under $\beta = 0$, and power for a suitable range of values of β corresponding to the sample size and distribution. The tables below are a selection from a larger set of results available from the authors.

Table 2 shows results when the arrivals distribution is Poisson with parameter 5. The finite sample size properties of both tests are good at each sample size, with only small deviations below the nominal 0.05 level in smaller samples that are reduced by $T = 800$. In this case both of the $\hat{\xi}_T$ and $\hat{\rho}_T$ tests are asymptotically efficient. For smaller samples and larger deviations from the null hypothesis, the parametric nature of the $\hat{\rho}_T$ test results in higher power than the $\hat{\xi}_T$ test. Correspondingly, large deviations from the null and large Poisson parameters result in high counts and hence a large number of probabilities are required to be estimated by the semi-parametric test, leading to a possible loss of power relative to the parametric test.

The arrivals distribution may deviate from standard distributions such as Poisson and Negative Binomial if it takes a mixture form. This may occur if the arrivals are drawn randomly from two different sources with different distributions. For example, Table 3 shows the results when the arrivals are an equally weighted mixture of Poisson with parameter 1 and Binomial with parameters $(10, 0.75)$. In this case the finite sample size properties of both tests remain good, while the $\hat{\xi}_T$ test is clearly superior in power, local to the null hypothesis, for all sample sizes, and for all parameter values for larger sample sizes. This illustrates the capacity of the $\hat{\xi}_T$ to adapt to unknown and non-standard arrivals distributions.

In general the $\hat{\xi}_T$ test tends to perform better in the locality of the null hypothesis, most likely because the sample support of $(y_1, \ldots, y_T)$ is smallest there. The $\hat{\rho}_T$ test can obtain finite sample power higher than the $\hat{\xi}_T$ test for smaller sample sizes, but the $\hat{\xi}_T$ test tends to dominate in larger samples when the arrivals are not Poisson, as predicted by the asymptotic theory.

4.2 $\mathcal{U} \neq \mathbb{N}$

A Monte Carlo experiment was carried out to illustrate the effect of having an arrivals support that satisfies Assumption 3.

Table 4 gives results for the $\hat{\xi}_T$ and $\hat{\rho}_T$ tests when the arrivals distribution is uniform on finite support $\mathcal{U} = \{0, 1, 2, 3\}$. In all cases the test sizes are close to the nominal level of 0.05. Other than for some large deviations from the null for $T = 100$, the $\hat{\xi}_T$ test exhibits substantially greater power than the $\hat{\rho}_T$ test for all parameter values. The power difference is especially large (over 50%) for the larger sample sizes and for smaller values of β. This particularly large divergence in the power properties of the two tests under Assumption 3 is predicted by the asymptotic theory, in which $\hat{\xi}_T$ has power against $O\left(T^{-1}\right)$ alternatives while $\hat{\rho}_T$ only has power against $O\left(T^{-1/2}\right)$

alternatives. For larger values of β (closer to one) the powers of the two tests are very similar, reflecting the consistency of both tests against fixed alternatives.

The influence of the arrivals' distributional shape and the gaps in its support can also be illustrated with the binomial distribution. Table 5 shows results for the Binomial$(6, 0.5)$ distribution which falls under Assumption 3 and the parameter $\pi^{(1)}$ is 0.015625. For comparison, Table 6 gives results for the previously discussed Binomial$(6, 0.5)$ supported on $\{0, 1, 2, 3, 6, 7, 8\}$ with a gap at $\{4, 5\}$ with $\pi^{(1)} = 0.328125$. This larger value of $\pi^{(1)}$ results in massive power increases for the $\hat{\xi}_T$ test relative to the Binomial$(6, 0.5)$ distribution with no gaps, while the power of the $\hat{\rho}_T$ test is essentially unchanged. The size of these power gains is attributable to the superior efficiency of the $\hat{\xi}_T$ test becoming overwhelmingly apparent in finite samples when $\pi^{(1)}$ is large. More extensive simulation results that further confirm these findings are available on request.

4.3 A Hybrid Test

The results of the Monte Carlo experiments reveal that the $\hat{\xi}_T$ and $\hat{\rho}_T$ tests each have situations in which their finite sample power is substantially superior to the other. The $\hat{\rho}_T$ test is often superior for small sample sizes due to its simpler parametric form, while the $\hat{\xi}_T$ test is often superior in larger samples when the arrivals distribution is clearly non-Poisson and especially when the arrivals support takes non-standard forms under Assumption 3. This suggests that it may be of practical interest to explore whether a combined test could be constructed. The combined test would attempt to capture the good power performance of each component without incurring the large power losses that an individual test may incur. The general form of the combined test is

$$\text{reject } H_0 \text{ if } \hat{\xi}_T > c_\alpha \text{ and/or } \hat{\rho}_T > c_\alpha,$$

for some critical value c_α.

Under H_0 and either Assumptions 2 or 3(a) these statistics can easily be seen to be jointly asymptotically distributed as

$$\begin{pmatrix} \hat{\xi}_T \\ \hat{\rho}_T \end{pmatrix} \rightsquigarrow \begin{pmatrix} Z_\xi \\ Z_\zeta \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix} \right),$$

where in practice λ can be consistently estimated by

$$\hat{\lambda} = \frac{T^{-1} \sum_{t=1}^{T} (\hat{g}_t - \bar{g})(y_t - \bar{y})}{\hat{\omega}},$$

and the joint critical value obtained by solving

$$1 - \Phi_2\left(\hat{c}_\alpha, \hat{c}_\alpha; \hat{\lambda}\right) = \alpha,$$

for $\hat{c}_\alpha$. Since $\hat{g}_t$ and y_t are generally found to be positively correlated, this procedure of estimating λ and the subsequent critical value will provide some power gain relative to a standard Bonferroni inequality.

Some Monte Carlo results are shown in Table 7 that illustrate the comparative advantages of each of the $\hat{\xi}_T$ and $\hat{\rho}_T$ and the practical contribution of the hybrid. The arrivals distribution is an equally weighted mixture of Poisson(1) and Negative Binomial$(1, 0.75)$ distributions. For the small sample size ($T = 100$) the simple structure of $\hat{\rho}_T$ delivers a more powerful test, while for the larger sample size ($T = 800$) the asymptotic efficiency of the $\hat{\xi}_T$ test delivers superior power. The hybrid test has good finite sample size properties and comparable power to whichever is the better test of $\hat{\xi}_T$ and $\hat{\rho}_T$ in any situation. The combined test is not asymptotically efficient, but has the appealing practical property of retaining most of the power of the individual tests while not suffering from the substantial power deficiencies that can beset *either of* them. Similar results are found in unreported simulations for many other distributions.

5 Testing with known or partially known support

It is unlikely in practice that the structure of the support $\mathcal{U}$ will be known. However in this section we briefly discuss the potential value of knowing either (a) the exact form of $\mathcal{U}$, or (b) that $\mathcal{U}$ satisfies Assumption 3, but not its exact form.

5.1 Known support

First suppose that $\mathcal{U}$ is known and satisfies Assumption 3. Thus, for every t, y_t has support $\mathcal{U}$ when $\beta = 0$, but has support strictly larger than $\mathcal{U}$ when $\beta > 0$. The implication is that any observation y_t that falls outside $\mathcal{U}$ implies that $H_0 : \beta = 0$ must be false. We therefore define the test

$$\Gamma_T : \text{reject } H_0 \text{ if } y_t \notin \mathcal{U} \text{ for any } t.$$

which has size of exactly zero since $y_t \notin \mathcal{U}$ is impossible under H_0. The Γ_T test has interesting asymptotic power properties since, from (7), the power of Γ_T against $\beta_T = T^{-1}h_\beta$ is asymptotically $1 - \exp\left(-h_\beta \mu_u \pi^{(1)}\right)$. This is $\Pr(Q = 1)$, see equation (9). Both the Γ_T and $\hat{\xi}_T$ tests derive their asymptotic power against $\beta_T = T^{-1}h_\beta$ from observations that fall outside $\mathcal{U}^{(0)}$ asymptotically (i.e. when $Q = 1$). Given $Q = 1$, the power of the Γ_T test is one while the power of the $\hat{\xi}_T$ test is $\Pr\left(X > z_\alpha\right)$, which is less than one. Given $Q = 0$ the power of the Γ_T test is zero while the power of the $\hat{\xi}_T$ test is α, so in this case both tests have power equal to size. Knowledge of $\mathcal{U}$ therefore permits the construction of a test with very good asymptotic properties, superior to those of the effective score test.

5.2 Partially known support

Now suppose that $\mathcal{U}$ is partially known, in the sense that Assumption 3 is known to apply but the exact form of $\mathcal{U}$ is not known. In this case the Γ_T test is infeasible, but the following feasible procedure can be applied.[2] Consider two estimators of $\mathcal{U}$ — the first is the usual *unconditional* sample support

$$\widehat{\mathcal{U}} = \{j : y_t = j \text{ for any } t = 1, \ldots, T\},$$

and the second is the sample support *conditional* on $y_{t-1} = 0$ for any t:

$$\widetilde{\mathcal{U}} = \{j : y_t = j \text{ and } y_{t-1} = 0 \text{ for any } t = 2, \ldots, T\}.$$

A feasible test of $H_0 : \beta = 0$ against $H_1 : \beta > 0$ will reject H_0 if $\widetilde{\mathcal{U}} \subset \widehat{\mathcal{U}}$, or, in a form analogous to that of Γ_T above,

$$\widehat{\Gamma}_T : \text{reject } H_0 \text{ if } y_t \notin \widetilde{\mathcal{U}} \text{ for any } t.$$

Any observation that falls outside the *conditional* support $\widetilde{\mathcal{U}}$ is evidence against H_0.

If $\beta = 0$ then it clearly must hold that $\widehat{\mathcal{U}} \subseteq \mathcal{U}$. Furthermore $\Pr(\widehat{\mathcal{U}} = \mathcal{U}) \to 1$ as $T \to \infty$, since for any $j \in \mathcal{U}$, $\Pr(j \in \widehat{\mathcal{U}}) = 1 - (1 - \pi_j)^T \to 1$. Similarly $\widetilde{\mathcal{U}} \subseteq \mathcal{U}$ and $\Pr(\widetilde{\mathcal{U}} = \mathcal{U}) \to 1$, which implies the unconditional and conditional supports are asymptotically equivalent under H_0 and hence that the size of $\widehat{\Gamma}_T$ is asymptotically zero (although not exactly zero as it is for Γ_T).

If $\beta > 0$ the construction of the support $\widetilde{\mathcal{U}}$ conditional on $y_{t-1} = 0$ ensures that $\widetilde{\mathcal{U}} \subseteq \mathcal{U}$ and $\Pr(\widetilde{\mathcal{U}} = \mathcal{U}) \to 1$ continue to hold. However now there is some non-zero probability that $y_t \notin \mathcal{U}$ for some t, and such observation(s) outside $\mathcal{U}$ must also be outside $\widetilde{\mathcal{U}}$, which causes the $\widehat{\Gamma}_T$ test to reject H_0. Specifically, for $\beta = \beta_T = T^{-1}h_\beta$, Lemma 1 implies that $\Pr(y_t \notin \mathcal{U} \text{ for some } t) \to 1 - \exp(-h_\beta \mu_u \pi^{(1)})$, and this probability is the asymptotic local power of the $\widehat{\Gamma}_T$ test. Therefore this $\widehat{\Gamma}_T$ test matches the asymptotic properties of the Γ_T test when $\mathcal{U}$ is fully known.

Clearly the Γ_T and $\widehat{\Gamma}_T$ tests are inapplicable if it is not known that Assumption 3 applies.

5.3 Finite sample properties

Table 8 gives some simulation results for a data generating process with Uniform arrivals on $\mathcal{U} = \{0, 1, 2, 3\}$ for the tests introduced in section 5 that incorporate full or partial knowledge of $\mathcal{U}$. The tests Γ_T (based on knowing $\mathcal{U}$ fully) and $\hat{\Gamma}_T$ (based on knowing that $\mathcal{U}$ has the form $\{0, 1, \ldots, M\}$ with M unknown) are included, along with randomised versions of both that are mixed with an independent Bernoulli draw to produce size of (approximately) 5%. The Hybrid test is also included for comparison purposes. The results shown for $T = 100$ and $T = 200$ show the potentially very large power gains available from knowledge (full or partial) of the structure

[2] We are grateful to a referee for this insightful suggestion.

of $\mathcal{U}$. Both Γ_T and $\widehat{\Gamma}_T$ achieve power far above that of the Hybrid test, while maintaining sizes of (essentially) zero. It is also notable that the performances of Γ_T and $\widehat{\Gamma}_T$ are almost identical in these sample sizes, showing that partial knowledge of the structure of the support can be as good as full knowledge in some circumstances.

However the finite sample properties of $\widehat{\Gamma}_T$ can vary dramatically with the arrivals distribution. Table 9 presents some simulation results from a data generating process with $\beta = 0$ and Binomial$(6, 0.5)$ arrivals distribution. To clarify the discussion we have assumed knowledge that the structure of the support is $\{0, 1, \ldots, M\}$ and defined

$$\widehat{\mathcal{U}} = \{0, 1, \ldots, \widehat{M}\}, \quad \widehat{M} = \max_t y_t,$$

and

$$\widetilde{\mathcal{U}} = \{0, 1, \ldots, \widetilde{M}\}, \quad \widetilde{M} = \max_{t: y_{t-1}=0} y_t.$$

The $\widehat{\Gamma}_T$ test therefore rejects H_0 if $\widetilde{M} < \widehat{M}$. It must always be true that $\widetilde{M} \leq \widehat{M}$ and $\widetilde{M} \leq M$. If H_0 is true then $\widehat{M} \leq M$ must also hold, but if H_0 is false then $\widehat{M} > M$ is possible. The columns of Table 9 report the proportion of replications for which $\widehat{M} = M$, $\widetilde{M} = M$ and the resulting $\widehat{\Gamma}_T$ test rejects H_0 (i.e. $\widetilde{M} < \widehat{M}$). The results reveal that very large sample sizes (T greater than 10,000) are required for the asymptotic theory to provide a reasonable approximation for the size properties of $\widehat{\Gamma}_T$. For practical sample sizes the test has sizes of over 90% and is therefore unusable. Unreported simulations, available on request, document similar behaviour for other arrivals distributions and illustrate the infeasibility of the attempt to control the finite sample size of $\widehat{\Gamma}_T$.

The conclusion from these simulations is that if the support $\mathcal{U}$ is fully known then Γ_T should be applied. However, if the support is only partially known in the sense of only knowing that Assumption 3 applies, then the $\widehat{\Gamma}_T$ is theoretically beneficial but in finite samples cannot be size-controlled across the range of arrivals distributions that satisfy Assumption 3. Instead, in the likely absence of full knowledge of the arrivals support, the effective score test and especially the hybrid test have been demonstrated to provide a theoretically sound and practically useful semi-parametric independence test in the INAR model.

6 Conclusion

In this paper we have investigated the asymptotic theory of semi-parametric likelihood-based independence tests in INAR(1) models, and shown that the structure of the support of the arrivals process plays a crucial role in the distribution theory under both null and local alternative hypotheses. If the arrivals are supported on the non-negative integers then the classical likelihood-based tests were shown to have standard asymptotic distribution theory and to be asymptotically efficient in the sense of Choi *et al* (1996). If the support of the arrivals is finite and/or contains gaps, it was shown that only the Neyman-Rao (effective score) test retains a non-degenerate asymptotic null distribution, and also that it has non-trivial power against local alternatives that approach the null at the unusually fast rate of $1/T$. Moreover this power was shown to derive from observations that fall outside the support of the arrivals, further emphasising the importance of the structure of the support in this problem. Finite sample simulations illustrated the practical importance of these asymptotic findings and were also used to motivate the construction of a hybrid of the effective score and first order autocorrelation tests that provides robust performance across a wide range of sample sizes and *unknown* arrivals distributions and supports, and can be recommended for use in practical data analysis.

Within the framework of count data modelling, certain interesting extensions of our results exist. Dependence testing in a higher order INAR model with non-parametric arrivals process would give rise to the same support issues as investigated here, while also involving more complicated methods to handle the null hypothesis lying on the boundary of a higher dimensional parameter space. Tests involving other thinning operators (see Weiss (2008)) could also be considered, along with potentially extended versions of the standard INAR model involving covariates. In all cases the possibility of restricted supports implied by low counts may complicate standard inference.

While beyond our scope, the results of this paper suggest the possibility that there exist other situations, not necessarily count data, in which statistical inference may depend in important ways on the support of the observations. For example, other testing problems involving latent processes (e.g. factor or state space models) with restricted supports, perhaps subject to censoring or truncation, may exhibit similar non-standard features to those demonstrated here.

References

Al-Osh, M.A. and Alzaid, A.A. (1987) First-order integer valued autoregressive (INAR(1)) process. *Journal of Time Series Analysis,* 8, 261-75.

Billingsley, P. (1995) *Probability and Measure*, 3rd edition. John Wiley & Sons, New York, U.S.A..

Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd edition. John Wiley & Sons, New York, U.S.A..

Bockenholt, U. (1999). Mixed INAR(1) Poisson regression models: analyzing heterogeneity and serial dependencies in longitudinal count data, *Journal of Econometrics*, 89, 317-338.

Box, G.E.P and D.A. Pierce (1970) Distribution of the residual autocorrelations in ARIMA time series models. *Journal of the American Statistical Association* 91, 1331–1342.

Brännäs, K. and Hellstrom, J. (2001). Generalized integer valued autoregression, *Econometric Reviews,* 20, 425-443.

Cardinal, M., Roy, R. and Lambert, J. (1999). On the application of integer-valued time series models for the analysis of disease incidence, *Statistics in Medicine,* 18, 2025-2039.

Choi, S, Hall, W.J. & Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models, *Annals of Statistics*, Vol 24, No 2, pp 841-861.

Drost, F.C., Van den Akker, R. and Werker, B.J.M. (2009). Efficient estimation of autoregression parameters and innovation distributions for semiparametric integer-valued AR(p) models, *Journal of the Royal Statistical Society (B),* 71, 467-485.

Franke, J. and Seligmann, T. (1993). Conditional maximum-likelihood estimates for INAR(1) processes and their applications to modelling epileptic seizure counts. In *Developments in Time Series*, 310-330. Subba Rao, T. (ed.). Chapman & Hall, London.

Freeland, R.K. (1998) *Statistical Analysis of Discrete Time Series with Applications to the Analysis of Workers Compensation Claims Data*, PhD thesis, The University of British Columbia, British Columbia, Canada.

Gourieroux, C. and Jasiak, J. (2004). Heterogeneous INAR(1) model with application to car insurance, *Insurance Mathematics and Economics,* 34, 177-192.

Huang, M., Y. Sun and H. White (2016) A flexible nonparametric test for conditional independence. *Econometric Theory* 32, 1434–1482.

Jung, R.C. & Tremayne, A.R. (2003) Testing for serial dependence in time series models of counts, *Journal of Time Series Analysis*, vol. 24, no. 1, pp. 65-84

Katz, L. (1965) Unified treatment of a broad class of discrete probability distributions. In: *Classical and Contagious Discrete Distribution*, G. P. Patil, ed., Calcutta: Statistical Publishing Society, Pergamon, Oxford, pp. 175-182.

McCabe, B.P.M., Martin, G.M. & Harris, D. (2011) Efficient probabilistic forecasts for counts, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 2, pp. 253-272.

McKenzie, E. (1985) Some simple models for discrete variate time series, *Journal of the American Water Resources Association*, vol. 21, no. 4, pp. 645-650.

Mills, T.M. & Seneta, E. (1989) Goodness-of-fit for a branching process with immigration using sample partial autocorrelations, *Stochastic Processes and their Applications*, vol. 33, no. 1, pp. 151-161.

Pavlopoulos, H. and Karlis, D. (2008). INAR(1) modelling of overdispersed count series with an environmental application. *Environmetrics*, 19, 369-393.

Pickands, J. and Stine, R. (1997). Estimation for an M/G/1 queue with incomplete information, *Biometrika*, 84, 295-308.

Rodríguez-Póo, J.M., S. Sperlich and P. Vieu (2015) Specification testing when the null is nonparametric or semiparametric. *Econometric Theory* 31, 1281–1309.

Rudholm, N. (2001). Entry and the number of firms in the Swedish pharmaceuticals market, *Review of Industrial Organization,* 19, 351-364.

Shao, X. (2011) Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory* 27, 312–343.

Sun, J. and McCabe B.P.M. (2013) Score Statistics for Testing the Serial Dependence in Count Data. *Journal of Time Series Analysis* vol. 34, no. 3, pp. 315–329.

Thyregod, P., Carstensen, J., Madsen, H. and Arnbjerg-Nielsen, K. (1999). Integer valued autoregressive models for tipping bucket rainfall measurements, *Environmetrics*, 10, 395-411.

Venkataraman, K.N. (1982) A time series approach to the study of the simple subcritical Galton-Watson process with immigration, *Advances in Applied Probability*, vol. 14, no. 1, pp. 1-20.

Weiss, C. (2008), "Thinning operations for modeling time series of counts - a survey", *AStA Advances in Statistical Analysis*, vol. 92, no. 3, pp. 319-341.

Whang, Y.J (1998) A test of autocorrelation in the presence of heteroskedasticity of unknown form. *Econometric Theory* 14, 87–122.

Table 2: Simulated size and power, Poisson(5) arrivals

$T=100$			$T=200$			$T=400$			$T=800$		
β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$
0.000	0.048	0.040	0.000	0.044	0.046	0.000	0.044	0.042	0.000	0.049	0.049
0.080	0.108	0.163	0.050	0.115	0.156	0.030	0.110	0.133	0.015	0.098	0.107
0.160	0.206	0.431	0.100	0.235	0.374	0.060	0.228	0.303	0.030	0.180	0.202
0.240	0.315	0.732	0.150	0.387	0.636	0.090	0.392	0.532	0.045	0.289	0.347
0.320	0.441	0.919	0.200	0.551	0.845	0.120	0.569	0.752	0.060	0.412	0.497
0.400	0.548	0.984	0.250	0.697	0.960	0.150	0.729	0.895	0.075	0.555	0.662
0.480	0.619	0.999	0.300	0.807	0.993	0.180	0.845	0.969	0.090	0.690	0.799
0.560	0.660	1.000	0.350	0.886	1.000	0.210	0.919	0.994	0.105	0.799	0.897
0.640	0.676	1.000	0.400	0.935	1.000	0.240	0.962	0.999	0.120	0.887	0.955
0.720	0.672	1.000	0.450	0.957	1.000	0.270	0.983	1.000	0.135	0.936	0.983
0.800	0.705	1.000	0.500	0.965	1.000	0.300	0.990	1.000	0.150	0.969	0.996

Table 3: Simulated size and power, Equal mixture of Poisson(1) and Binomial(10,0.75) arrivals

$T=100$			$T=200$			$T=400$			$T=800$		
β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$
0.000	0.046	0.043	0.000	0.047	0.036	0.000	0.050	0.041	0.000	0.046	0.043
0.050	0.251	0.110	0.030	0.270	0.086	0.020	0.309	0.095	0.010	0.232	0.079
0.100	0.507	0.221	0.060	0.598	0.166	0.040	0.677	0.185	0.020	0.543	0.134
0.150	0.670	0.387	0.090	0.814	0.309	0.060	0.903	0.308	0.030	0.798	0.201
0.200	0.756	0.580	0.120	0.922	0.476	0.080	0.976	0.465	0.040	0.937	0.290
0.250	0.762	0.757	0.150	0.967	0.640	0.100	0.996	0.616	0.050	0.983	0.393
0.300	0.736	0.880	0.180	0.983	0.783	0.120	0.999	0.759	0.060	0.998	0.497
0.350	0.678	0.952	0.210	0.990	0.884	0.140	1.000	0.862	0.070	1.000	0.607
0.400	0.605	0.985	0.240	0.992	0.949	0.160	1.000	0.928	0.080	1.000	0.711
0.450	0.532	0.997	0.270	0.992	0.980	0.180	1.000	0.965	0.090	1.000	0.806
0.500	0.466	1.000	0.300	0.992	0.994	0.200	1.000	0.989	0.100	1.000	0.873

Table 4: Simulated size and power, Uniform{0,1,2,3} arrivals

$T = 100$			$T = 200$			$T = 400$			$T = 800$		
β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$
0.000	0.044	0.043	0.000	0.050	0.048	0.000	0.045	0.045	0.000	0.046	0.044
0.050	0.167	0.100	0.030	0.214	0.104	0.015	0.217	0.083	0.012	0.371	0.090
0.100	0.338	0.221	0.060	0.431	0.195	0.030	0.437	0.131	0.024	0.659	0.160
0.150	0.521	0.394	0.090	0.618	0.320	0.045	0.616	0.204	0.036	0.839	0.257
0.200	0.656	0.591	0.120	0.763	0.486	0.060	0.755	0.297	0.048	0.924	0.377
0.250	0.755	0.760	0.150	0.850	0.642	0.075	0.850	0.405	0.060	0.974	0.509
0.300	0.837	0.883	0.180	0.912	0.787	0.090	0.920	0.524	0.072	0.990	0.632
0.350	0.896	0.954	0.210	0.946	0.883	0.105	0.954	0.646	0.084	0.996	0.746
0.400	0.930	0.983	0.240	0.964	0.944	0.120	0.974	0.752	0.096	0.999	0.842
0.450	0.958	0.995	0.270	0.982	0.978	0.135	0.986	0.831	0.108	1.000	0.911
0.500	0.974	0.999	0.300	0.990	0.993	0.150	0.992	0.894	0.120	1.000	0.954

Table 5: Simulated size and power, Binomial(6,0.5) arrivals

$T = 100$			$T = 200$			$T = 400$			$T = 800$		
β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$
0.000	0.042	0.039	0.000	0.041	0.045	0.000	0.046	0.048	0.000	0.047	0.047
0.080	0.138	0.168	0.040	0.118	0.120	0.020	0.103	0.099	0.015	0.120	0.105
0.160	0.322	0.415	0.080	0.246	0.271	0.040	0.192	0.180	0.030	0.236	0.209
0.240	0.534	0.722	0.120	0.435	0.488	0.060	0.308	0.297	0.045	0.375	0.337
0.320	0.730	0.916	0.160	0.607	0.693	0.080	0.448	0.448	0.060	0.537	0.498
0.400	0.853	0.986	0.200	0.766	0.861	0.100	0.597	0.613	0.075	0.696	0.667
0.480	0.915	0.999	0.240	0.877	0.949	0.120	0.725	0.755	0.090	0.824	0.804
0.560	0.948	1.000	0.280	0.935	0.983	0.140	0.836	0.862	0.105	0.910	0.902
0.640	0.958	1.000	0.320	0.970	0.995	0.160	0.909	0.931	0.120	0.957	0.950
0.720	0.961	1.000	0.360	0.988	0.999	0.180	0.952	0.970	0.135	0.982	0.982
0.800	0.936	1.000	0.400	0.993	1.000	0.200	0.974	0.988	0.150	0.993	0.993

Table 6: Simulated size and power, Binomial(6,0.5), gap at {4,5} arrivals

$T = 100$			$T = 200$			$T = 400$			$T = 800$		
β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$	β	$\hat{\xi}_T$	$\hat{\rho}_T$
0.000	0.056	0.040	0.000	0.046	0.045	0.000	0.052	0.046	0.000	0.050	0.046
0.060	0.453	0.123	0.030	0.453	0.098	0.020	0.553	0.101	0.015	0.707	0.106
0.120	0.687	0.285	0.060	0.742	0.191	0.040	0.850	0.181	0.030	0.943	0.207
0.180	0.766	0.499	0.090	0.880	0.325	0.060	0.954	0.303	0.045	0.992	0.337
0.240	0.739	0.723	0.120	0.940	0.480	0.080	0.986	0.448	0.060	0.999	0.501
0.300	0.708	0.879	0.150	0.969	0.647	0.100	0.995	0.602	0.075	1.000	0.656
0.360	0.714	0.960	0.180	0.976	0.787	0.120	0.998	0.750	0.090	1.000	0.799
0.420	0.721	0.988	0.210	0.982	0.892	0.140	1.000	0.862	0.105	1.000	0.901
0.480	0.735	0.998	0.240	0.979	0.949	0.160	1.000	0.926	0.120	1.000	0.955
0.540	0.761	1.000	0.270	0.975	0.978	0.180	1.000	0.965	0.135	1.000	0.980
0.600	0.779	1.000	0.300	0.978	0.990	0.200	1.000	0.987	0.150	1.000	0.994

Table 7: Simulated size and power, Equal mixture of Poisson(1) and Negative Binomial (10,0.75) arrivals

$T = 100$				$T = 200$				$T = 800$			
β	$\hat{\xi}_T$	$\hat{\rho}_T$	Hybrid	β	$\hat{\xi}_T$	$\hat{\rho}_T$	Hybrid	β	$\hat{\xi}_T$	$\hat{\rho}_T$	Hybrid
0.000	0.059	0.047	0.060	0.000	0.048	0.048	0.053	0.000	0.051	0.059	0.059
0.080	0.230	0.160	0.236	0.050	0.248	0.160	0.245	0.015	0.185	0.118	0.169
0.160	0.445	0.406	0.531	0.100	0.525	0.359	0.555	0.030	0.410	0.208	0.383
0.240	0.575	0.704	0.792	0.150	0.739	0.638	0.806	0.045	0.664	0.340	0.626
0.320	0.652	0.918	0.936	0.200	0.859	0.862	0.946	0.060	0.846	0.499	0.820
0.400	0.692	0.986	0.989	0.250	0.912	0.967	0.988	0.075	0.947	0.670	0.931
0.480	0.680	0.998	0.999	0.300	0.940	0.994	0.998	0.090	0.983	0.806	0.978
0.560	0.668	1.000	1.000	0.350	0.954	0.999	1.000	0.105	0.996	0.898	0.996
0.640	0.643	1.000	1.000	0.400	0.963	1.000	1.000	0.120	0.999	0.954	1.000
0.720	0.559	1.000	1.000	0.450	0.966	1.000	1.000	0.135	1.000	0.985	1.000
0.800	0.464	1.000	1.000	0.500	0.967	1.000	1.000	0.150	1.000	0.996	1.000

Table 8: Simulated size and power, Uniform $\{0, 1, 2, 3\}$ arrivals

		$T = 100$						$T = 200$			
β	Hybrid	Γ_T	$\widehat{\Gamma}_T$	Γ_T^*	$\widehat{\Gamma}_T^*$	β	Hybrid	Γ_T	$\widehat{\Gamma}_T$	Γ_T^*	$\widehat{\Gamma}_T^*$
0.000	0.035	0.000	0.001	0.051	0.048	0.000	0.040	0.000	0.000	0.053	0.050
0.050	0.127	0.848	0.848	0.854	0.855	0.030	0.152	0.894	0.894	0.899	0.897
0.100	0.299	0.983	0.983	0.983	0.984	0.060	0.352	0.991	0.991	0.991	0.991
0.150	0.479	0.999	0.999	0.999	0.999	0.090	0.544	0.999	0.999	0.999	0.999
0.200	0.653	1.000	1.000	1.000	1.000	0.120	0.695	1.000	1.000	1.000	1.000
0.250	0.789	1.000	1.000	1.000	1.000	0.150	0.808	1.000	1.000	1.000	1.000
0.300	0.896	1.000	1.000	1.000	1.000	0.180	0.892	1.000	1.000	1.000	1.000

Table 9: Simulated properties of $\widehat{M}$, $\widetilde{M}$, $\widehat{\Gamma}_T$ under H_0 with Binomial$(6, 0.5)$ arrivals

T	$\widehat{M} = M$	$\widetilde{M} = M$	$\widehat{\Gamma}_T$
100	0.791	0.024	0.943
200	0.967	0.054	0.934
400	0.995	0.092	0.908
800	1.000	0.171	0.829
1600	1.000	0.324	0.676
3200	1.000	0.542	0.458
6400	1.000	0.773	0.227
12800	1.000	0.963	0.037
25600	1.000	0.997	0.003

$\widehat{M} = M$: proportion of replications with $\widehat{M} = M$
$\widetilde{M} = M$: proportion of replications with $\widetilde{M} = M$
$\widehat{\Gamma}_T$: proportion of replications where $\widehat{\Gamma}_T$ rejects H_0

A Proofs of Main Results

A.1 Proof of Theorem 1

Both Theorems 1 and 2 hold under Assumption 2, in which the arrivals support is specified to be $\mathcal{U} = \mathbb{N}$. This implies that π_j is *strictly* positive for every $j \in \mathbb{N}$. This fact will be used throughout the proofs of these two theorems.

A.1.1 Proof of Theorem 1: The LAN Property

The log-likelihood of the model (1) is given by (2). Theorem A.1 below shows that the log-likelihood ratio for this model has the LAN property under H_0. Statement of the result requires the specification of a local sequence at $\beta = 0$

$$\beta_T(h_\beta) = T^{-1/2} h_\beta, \quad h_\beta > 0 \tag{A.1}$$

and a local sequence for Π given by a linear map $\Pi_T : \mathcal{H}_\pi \mapsto \mathcal{P}_{\mathbb{N}}$ with typical element

$$\pi_{T,k}(h_\pi) = \pi_k \left(1 + T^{-1/2}\left(h_{\pi,k} - \sum_{j=0}^{\infty} \pi_j h_{\pi,j}\right)\right), \tag{A.2}$$

from $\mathcal{H}_\pi = \{h_\pi \in \ell^\infty(\mathbb{N}) : h_{\pi,0} = 0\}$ to $\mathcal{P}_{\mathbb{N}}$, the set of probability distributions on $\mathbb{N}$. So Π_T acts linearly on members of $\mathcal{H}_\pi$ i.e. $h_\pi = \{h_{\pi,j}\}$, which are convergent sequences (in $\ell^\infty(\mathbb{N})$) and transforms them into probability distributions in $\mathcal{P}_{\mathbb{N}}$. The condition $h_{\pi,0} = 0$ makes explicit the adding-up restriction for any $\Pi = \{\pi_k\}$ expressed as $\pi_0 = 1 - \sum_{k=1}^{\infty} \pi_k$. The combined directions for β and Π are denoted $h = (h_\beta, h_\pi)$. Define notation $\theta = (\beta, \Pi)$, $\theta_0 = (0, \Pi)$ and $\theta_T = (\beta_T, \Pi_T)$.

Theorem A.1 *Under Assumption 2*

$$\log \frac{L_T(\theta_T h)}{L_T(\theta_0)} = S_T(h) - \frac{1}{2}\langle h, Vh\rangle + r_T(h) \tag{A.3}$$

where

(i) $S_T = (S_{T,\beta}, S_{T,\pi})$ *is a random linear functional with elements defined by*

$$S_{T,\beta} h_\beta = T^{-1/2} \sum_{t=1}^{T} y_{t-1}\left(\frac{\pi_{y_t-1}}{\pi_{y_t}} - 1\right) h_\beta, \tag{A.4}$$

$$S_{T,\pi} h_\pi = T^{-1/2} \sum_{t=1}^{T} \left(h_{\pi,y_t} - \sum_{j=1}^{\infty} \pi_j h_{\pi,j}\right), \tag{A.5}$$

that satisfies $S_T \rightsquigarrow Z$*, where* Z *is a tight Gaussian process with mean zero and variance* V *under* θ_0*,* *(ii) V is a positive definite self-adjoint bounded linear operator with elements specified by*

$$V_{\beta\beta} = E\left(u_t^2\right)\left(\sum_{k=1}^{\infty}\frac{\pi_{k-1}^2}{\pi_k} - 1\right) \quad \text{(A.6)}$$

$$V_{\beta\pi}h_\pi = \mu_u\sum_{k=1}^{\infty}\left(\pi_{k-1} - \pi_k\right)h_{\pi,k} \quad \text{(A.7)}$$

$$\langle h_\pi, V_{\pi\pi}h_\pi\rangle = \sum_{k=1}^{\infty}\pi_k h_{\pi,k}^2 - \left(\sum_{k=1}^{\infty}\pi_k h_{\pi,k}\right)^2 \quad \text{(A.8)}$$

where $\langle ., V.\rangle$ denotes the inner product induced by V,
(iii) $r_T(h) \xrightarrow{p} 0$ for every h under θ_0.

The proof of this Theorem is provided in the online supplementary appendix section B.1.

A.1.2 Proof of Theorem 1: The infeasible effective score test

Having proved the LAN property the next step in constructing an optimal test is to assume the nuisance parameters $\{\pi_k\}$ are known. Based on these known $\{\pi_k\}$, define the statistics

$$S_{T,\beta}^* = T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1} - \mu_u\right)\left(g_t - 1\right) \quad \text{(A.9)}$$

and

$$\omega^2 = \sigma_u^2\left(\sum_{k=1}^{\infty}\frac{\pi_{k-1}^2}{\pi_k} - 1\right),$$

where $g_t = \pi_{y_t-1}/\pi_{y_t}$. Note that Assumption 2 ensures that $\pi_{y_t} > 0$ for every y_t, while $\pi_{-1} = 0$ when $y_t = 0$.

Lemma A.1 *Suppose Assumption 2 holds. The standardised effective score test statistic is*

$$\xi_T = \frac{S_{T,\beta}^*}{\omega}.$$

Under H_0 this statistic satisfies $\xi_T \rightsquigarrow Z^* \sim N(0,1)$, so the effective score test that rejects H_0 for $\xi_T > z_\alpha$ therefore has asymptotic size of α. This effective score test is asymptotically uniformly most powerful at a given Π. In addition, under local alternatives θ_T, $\Pr(\xi_T > z_\alpha) \to 1 - \Phi(z_\alpha - \omega h_\beta)$ which gives the local power.

The proof of this Lemma is provided in the online supplementary appendix section B.2.

A.1.3 Proof of Theorem 1: The feasible effective score test

The final step to prove Theorem 1 is accomplished by showing that a feasible test and the infeasible one are equivalent asymptotically. The feasible test is given by $\hat{\xi}_T$ in (4) and we now show that the distribution of ξ_T is unchanged when the nuisance parameters are replaced by $\sqrt{T}$-consistent estimators, hence concluding that the test based on $\hat{\xi}_T$ is asymptotically uniformly most powerful.

The following detailed derivations show that $\hat{\xi}_T - \xi_T \xrightarrow{p} 0$ under H_0, i.e. under the parameter vector θ_0. The same result under θ_T then follows immediately from Le Cam's third lemma and the LAN property proved in Theorem A.1 above. Therefore we proceed with $\beta = 0$.

Recalling the notation $\hat{g}_t = \hat{\pi}_{y_t-1}/\hat{\pi}_{y_t}$ and $g_t = \pi_{y_t-1}/\pi_{y_t}$

$$\begin{aligned}
\hat{S}^*_{T,\beta} - S^*_{T,\beta} &= T^{-1/2}\sum_{t=1}^{T}(y_{t-1}-\hat{\mu}_u)(\hat{g}_t - 1) - T^{-1/2}\sum_{t=1}^{T}(y_{t-1}-\mu_u)(g_t-1) \\
&= T^{-1/2}\sum_{t=1}^{T}(y_{t-1}-\mu_u)(\hat{g}_t - g_t) && \text{(A.10)} \\
&\quad - (\hat{\mu}_u - \mu_u)\, T^{-1/2}\sum_{t=1}^{T}(\hat{g}_t - g_t) && \text{(A.11)} \\
&\quad - (\hat{\mu}_u - \mu_u)\, T^{-1/2}\sum_{t=1}^{T}(g_t - 1). && \text{(A.12)}
\end{aligned}$$

The sample mean $\hat{\mu}_u$ is consistent for μ_u and $g_t - 1$ satisfies a CLT under Assumption 2, so (A.12) is $o_p(1)$. Both (A.10) and (A.11) require analysis of $\hat{g}_t - g_t$.

The restricted estimator (3) can be re-expressed as

$$\hat{\Pi} = \Pi_T\left(\hat{h}_\pi\right),$$

where Π_T is defined in (A.2) and

$$\hat{h}_\pi = \arg\max_{h_\pi} \log \frac{L_T(0, \Pi_T(h_\pi))}{L_T(0,\Pi)}.$$

Regardless of whether Assumption 2 or 3 applies, these log-likelihood ratios (with $\beta = 0$ imposed) have the LAN representation under the null

$$\log \frac{L_T(0, \Pi_T(h_\pi))}{L_T(0,\Pi)} = S_{T,\pi} h_\pi - \frac{1}{2}\langle h_\pi, V_{\pi\pi} h_\pi\rangle + o_p(1)$$

so that $\hat{h}_\pi$ can be represented.

$$\hat{h}_\pi = V_{\pi\pi}^{-1} S_{T,\pi} + o_p(1). \tag{A.13}$$

Now, using (A.2),

$$\begin{aligned}
\hat{g}_t - g_t &= \frac{\hat{\pi}_{y_t-1}}{\hat{\pi}_{y_t}} - \frac{\pi_{y_t-1}}{\pi_{y_t}} \\
&= T^{-1/2}\sum_{k\in\mathcal{U}} 1_k(y_t)\,\frac{\pi_{k-1}}{\pi_k}\left(\frac{\hat{h}_{\pi,k-1} - \hat{h}_{\pi,k}}{1 + T^{-1/2}\left(\hat{h}_{\pi,k} - \sum_{j=1}^{\infty}\pi_j \hat{h}_{\pi,j}\right)}\right).
\end{aligned}$$

Thus

$$\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\left(\hat{g}_t-g_t\right) &= \sum_{k\in\mathcal{U}}\left(T^{-1}\sum_{t=1}^{T}1_k\left(y_t\right)\right)\frac{\pi_{k-1}}{\pi_k}\left(\frac{\hat{h}_{\pi,k-1}-\hat{h}_{\pi,k}}{1+T^{-1/2}\left(\hat{h}_{\pi,k}-\sum_{j=1}^{\infty}\pi_j\hat{h}_{\pi,j}\right)}\right)\\
&= \sum_{k\in\mathcal{U}}\hat{\pi}_k\frac{\pi_{k-1}}{\pi_k}\left(\frac{\hat{h}_{\pi,k-1}-\hat{h}_{\pi,k}}{1+T^{-1/2}\left(\hat{h}_{\pi,k}-\sum_{j=1}^{\infty}\pi_j\hat{h}_{\pi,j}\right)}\right)\\
&= \sum_{k\in\mathcal{U}}\pi_{k-1}\left(\hat{h}_{\pi,k-1}-\hat{h}_{\pi,k}\right)\\
&= -\sum_{k\in\mathcal{U}}\left(\pi_{k-1}-\pi_k\right)\hat{h}_{\pi,k}.
\end{aligned}$$

Combining this with (A.7) and (A.13) gives

$$T^{-1/2}\sum_{t=1}^{T}\left(\hat{g}_t-g_t\right)=-\frac{1}{\mu_1}S_{T,\pi}V_{\pi\pi}^{-1}V_{\pi\beta}=O_p\left(1\right),$$

so that (A.11) is $o_p\left(1\right)$ by the consistency of $\hat{\mu}_u$. Similarly

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1}-\mu_u\right)\left(\hat{g}_t-g_t\right)\\
&= \sum_{k\in\mathcal{U}}\left(T^{-1}\sum_{t=1}^{T}\left(y_{t-1}-\mu_u\right)1_k\left(y_t\right)\right)\frac{\pi_{k-1}}{\pi_k}\left(\frac{\hat{h}_{\pi,k-1}-\hat{h}_{\pi,k}}{1+T^{-1/2}\left(\hat{h}_{\pi,k}-\sum_{j=1}^{\infty}\pi_j\hat{h}_{\pi,j}\right)}\right)\\
&= \left(\hat{\mu}_u-\mu_u\right)\sum_{k\in\mathcal{U}}\pi_{k-1}\left(\hat{h}_{\pi,k-1}-\hat{h}_{\pi,k}\right)+o_p\left(1\right),
\end{aligned}$$

so that (A.10) is also $o_p\left(1\right)$.

Finally the variance $\hat{\omega}^2$ is consistent for ω^2 by the consistency of $\hat{\Pi}$ and the sample variances, so $\hat{\xi}_T=\hat{S}^*_{T,\beta}/\hat{\omega}=S^*_{T,\beta}/\omega+o_p(1)=\xi_T+o_p(1)$ under H_0, which is the required result. ■

A.2 Proof of Theorem 2

(i) The LAN representation (A.3) implies that

$$T^{1/2}\tilde{\beta}=\left(S^*_{T,\beta}/\omega^2+o_p\left(1\right)\right)\vee 0. \tag{A.14}$$

It follows from (**??**) that the asymptotic joint null distribution of $S^*_{T,\beta}/\omega^2$ and the quadratic approximation to the likelihood ratios

$$\lambda_T h=S_T h-\frac{1}{2}\left\langle h,Vh\right\rangle$$

is, for any $h=(h_\beta,h_\pi)$, $h_\beta>0$,

$$\begin{pmatrix}\lambda_T h\\ S^*_{T,\beta}/\omega^2\end{pmatrix}\rightsquigarrow N\left(\begin{pmatrix}-\frac{1}{2}\left\langle h,Vh\right\rangle\\ 0\end{pmatrix},\begin{pmatrix}\left\langle h,Vh\right\rangle & h_\beta\\ h_\beta & 1/\omega^2\end{pmatrix}\right).$$

Le Cam's third lemma therefore implies that

$$S_{T,\beta}^{*}/\omega^{2} \rightsquigarrow N\left(h_{\beta}, 1/\omega^{2}\right) = Z_{\beta}$$

under $\theta_T h$, which combines with (A.14) to give the conclusion.

(ii) Given a consistent estimator $\hat{\omega}^2$ (which can be demonstrated as in Theorem 1), it follows from part (i) that the Wald statistic satisfies

$$W_T = \left(\frac{T^{1/2}\tilde{\beta}}{\hat{\omega}}\right)^2 \rightsquigarrow \left(\omega^{-1} Z_{\beta} \vee 0\right)^2 = \left(Z_{\Lambda} \vee 0\right)^2.$$

The LR statistic satisfies

$$\begin{aligned}
\Lambda_T &= 0, \text{ if } \tilde{\beta} = 0, \text{ (and hence } \tilde{\Pi} = \hat{\Pi}) \\
\Lambda_T &= 2\left(\left\langle S_T, V^{-1} S_T\right\rangle - \left\langle S_{T,\pi}, V_{\pi\pi}^{-1} S_{T,\pi}\right\rangle + o_p(1)\right) = \left(\frac{S_{T,\beta}^{*}}{\omega}\right)^2 + o_p(1) \text{ if } \tilde{\beta} > 0.
\end{aligned}$$

Combining these with (A.14) implies

$$\Lambda_T = \left(\frac{T^{1/2}\tilde{\beta}}{\omega}\right)^2 + o_p(1) = W_T + o_p(1).$$

(iii) The score statistic satisfies

$$\Psi_T = \left(\frac{S_{T,\beta}^{*}}{\omega}\right)^2 + o_p(1) \rightsquigarrow Z_{\Lambda}^2,$$

while its one-sided version satisfies

$$\Psi_T^{+} = \frac{S_{T,\beta}^{*}}{\omega} + o_p(1) = \xi_T + o_p(1)$$

■

A.3 Proof of Theorem 3

The following proofs of Theorems 3 and 4 are carried out under Assumption 3, in which the arrivals support is specified to be $\mathcal{U} \neq \mathbb{N}$. This implies that $\pi_j = 0$ for some $j \in \mathbb{N}$, which changes many of the results proved for Theorems 1 and 2.

Theorem 3 is stated and proved assuming H_0 is true, implying $y_t \in \mathcal{U}$ for every t, and hence $\pi_{y_t} > 0$ for every t. The presence of gaps may imply $\pi_{y_t - 1} = 0$ for some t, so that $\pi_{y_t-1}/\pi_{y_t} = 0$, but this causes no problems in the derivations to follow.

(i) Under Assumption 3, $E\left(\pi_{y_t-1}/\pi_{y_t}\right)=\sum_{k\in\mathcal{U}^{(0)}}\pi_{k-1}=\pi^{(0)}<1$. Since $\hat{\mu}_u=T^{-1}\sum_{t=1}^{T}y_{t-1}+O_p\left(T^{-1}\right)$, the sample effective score can be represented

$$\begin{aligned}
\hat{S}^{*}_{T,\beta} &= T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1}-\hat{\mu}_u\right)\left(\hat{g}_t-\pi^{(0)}\right)+o_p(1) && \text{(A.15)}\\
&= T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1}-\mu_u\right)\left(g_t-\pi^{(0)}\right) && \text{(A.16)}\\
&\quad+\left(\hat{\mu}_u-\mu_u\right)T^{-1/2}\sum_{t=1}^{T}\left(g_t-\pi^{(0)}\right) && \text{(A.17)}\\
&\quad+T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1}-\hat{\mu}_u\right)\left(\hat{g}_t-g_t\right) && \text{(A.18)}\\
&= T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1}-\mu_u\right)\left(g_t-\pi^{(0)}\right)+o_p(1), && \text{(A.19)}
\end{aligned}$$

in which (A.15) follows by the summation properties of the sample mean, (A.17) is $o_p(1)$ because $\hat{\mu}_u$ is consistent for μ_u and a CLT applies for zero mean i.i.d. process $\left(\pi_{y_t-1}/\pi_{y_t}-\pi^{(0)}\right)$, and (A.18) is $o_p(1)$ since the treatment of (A.10) and (A.11) in the proof of Theorem 1 also applies under Assumption 3. The summand in (A.19) is a stationary ergodic martingale difference sequence, which therefore satisfies a central limit theorem with limiting variance $\omega^2=\sigma_u^2\sigma_g^2$. This is consistently estimated by $\hat{\omega}^2$ in (5).

(ii) Since

$$E\left(y_{t-1}\left(g_t-1\right)\right)=\mu_u\left(\pi^{(0)}-1\right)=-\mu_u\pi^{(1)}<0$$

under Assumption 3, the score with respect to β satisfies

$$S_{T,\beta}=T^{-1/2}\sum_{t=1}^{T}y_{t-1}\left(g_t-1\right)\xrightarrow{p}-\infty.$$

Also

$$\hat{S}_{T,\beta}-S_{T,\beta}=T^{-1/2}\sum_{t=1}^{T}y_{t-1}\left(\hat{g}_t-g_t\right)\xrightarrow{p}0,$$

again following the treatment of (A.10) and (A.11). Thus $\hat{S}_{T,\beta}\xrightarrow{p}-\infty$, giving $\Psi_T\xrightarrow{p}+\infty$ and $\Psi_T^{+}\xrightarrow{p}-\infty$ as required, since the variance estimator $\hat{\omega}^2$ remains bounded in probability.

(iii) First observe that $\hat{S}_{T,\beta}<0$ implies that $\tilde{\beta}=0$ since the log-likelihood is decreasing as β increases through zero. Since $\hat{S}_{T,\beta}\xrightarrow{p}-\infty$ implies $\Pr\left(\hat{S}_{T,\beta}<0\right)\to1$, we conclude $\Pr\left(\tilde{\beta}=0\right)\to1$, which in turn implies that $\Pr\left(W_T=0\right),\Pr\left(\Lambda_T=0\right)\to1$. ■

A.4 Proof of Lemma 1

We use $a_T\approx b_T$ to represent $a_T/b_T\to1$ as $T\to\infty$.

Lemma A.2 *Let $\mathcal{A}$ be any subset of $\mathbb{N} \setminus \mathcal{U}^{(0)}$. For any $\mathcal{T} = \{t_1, \ldots, t_k\}$ with $1 \leq t_1$, $t_k \leq T$, $t_j - t_{j-1} > 1$, and $\mathcal{A}_k = \{i_1, \ldots, i_k\} \subseteq \mathcal{A}$ (i_j not necessarily distinct)*

$$\begin{aligned} &\Pr\left(y_{t_1} = i_1, \ldots, y_{t_k} = i_k \text{ and } y_s \notin \mathcal{A} \text{ for all } s \neq t_1, \ldots, t_k\right) \\ \approx & \left(\prod_{i \in \mathcal{A}_k} \beta_T \mu_u \pi_{i-1}\right) \exp\left(-h_\beta \mu_u \sum_{i \in \mathcal{A}} \pi_{i-1}\right). \end{aligned}$$

Also

$$\Pr\left(y_t \notin \mathcal{A} \text{ for all } t = 1, \ldots, T\right) \approx \exp\left(-h_\beta \mu_u \sum_{i \in \mathcal{A}} \pi_{i-1}\right). \tag{A.20}$$

The proof of this Lemma is given in section B.3 of the online Appendix B.

Define $\mathcal{U}^{(k)} = \left\{i \notin \cup_{j<k} \mathcal{U}^{(j)} : i-1 \in \mathcal{U}^{(k-1)}\right\}$. This definition collects in $\mathcal{U}^{(k)}$ those integers i that are not included in the arrivals support $\mathcal{U}$ and that differ by k from the nearest element of $\mathcal{U}$ that is less than i. Setting $\mathcal{A} = \mathbb{N} \setminus \mathcal{U}^{(0)}$ in (A.20) gives

$$\Pr\left(y_t \in \mathcal{U}^{(0)} \text{ for all } t = 1, \ldots, T\right) \approx \exp\left(-h_\beta \mu_u \sum_{i \in \mathcal{U}^{(1)}} \pi_{i-1}\right),$$

since $\pi_{i-1} = 0$ for $i \in \mathcal{U}^{(k)}$, $k > 1$, and $\pi^{(1)} = \sum_{i \in \mathcal{U}^{(1)}} \pi_{i-1}$, which shows (7).

Lemma A.2 allows the derivation of the asymptotic distribution of $N_{T,i}$ for any $i \in \mathcal{U}^{(1)}$. For a fixed non-negative integer k, set $\mathcal{A} = \{i, \ldots, i\}$ in Lemma A.2 to find

$$\begin{aligned} \Pr\left(N_{T,i} = k\right) &= \sum_{t_k=k+1}^{T} \sum_{t_{k-1}=k}^{t_k-1} \cdots \sum_{t_2=2}^{t_3-1} \sum_{t_1=1}^{t_2-1} \Pr\left(y_{t_1} = i, \ldots, y_{t_k} = i \text{ and } y_s \neq i \text{ for all } s \neq t_1, \ldots, t_k\right) \\ &\approx T^{-k} \sum_{t_k=k+1}^{T} \sum_{t_{k-1}=k}^{t_k-1} \cdots \sum_{t_2=2}^{t_3-1} \sum_{t_1=1}^{t_2-1} \left(h_\beta \mu_u \pi_{i-1}\right)^k \exp\left(-h_\beta \mu_u \pi_{i-1}\right) \\ &\approx T^{-k} \binom{T}{k} \left(h_\beta \mu_u \pi_{i-1}\right)^k \exp\left(-h_\beta \mu_u \pi_{i-1}\right) \\ &\approx \frac{1}{k!} \left(h_\beta \mu_u \pi_{i-1}\right)^k \exp\left(h_\beta \mu_u \pi_{i-1}\right), \end{aligned} \tag{A.21}$$

which is the Poisson probability mass function with parameter $h_\beta \mu_u \pi_{i-1}$.

For the joint convergence of finite collections of $N_{T,i}$ over i, consider any set $\mathcal{A} = \{i_1, i_2, \ldots i_k\} \subseteq \mathcal{U}^{(1)}$. Lemma A.2 immediately gives

$$\begin{aligned} \Pr\left(N_{T,i} = 0 \text{ for all } i \in \mathcal{A}\right) &= \Pr\left(y_t \notin \mathcal{A} \text{ for all } t\right) \\ &\approx \exp\left(-h_\beta \mu_u \sum_{i \in \mathcal{A}} \pi_{i-1}\right) \\ &= \prod_{i \in \mathcal{A}} \exp\left(-h_\beta \mu_u \pi_{i-1}\right), \end{aligned} \tag{A.22}$$

the product of marginal Poisson probabilities. Similarly

$$\begin{aligned}
& \Pr\left(N_{T,i}=1 \text{ for all } i\in\mathcal{A}\right) \\
= & \sum_{\substack{t_1=1 \\ t_1\neq\ldots\neq t_k}}^{T}\cdots\sum_{t_k=1}^{T}\Pr\left(y_{t_1}=i_1,\ldots,y_{t_k}=i_k \text{ and } y_t\notin\mathcal{A} \text{ for all } s\neq t_1,\ldots,t_k\right) \\
\approx & \sum_{\substack{t_1=1 \\ |t_i-t_j|\geq 2\ \forall i\neq j}}^{T}\cdots\sum_{t_k=1}^{T}\Pr\left(y_{t_1}=i_1,\ldots,y_{t_k}=i_k \text{ and } y_t\notin\mathcal{A} \text{ for all } s\neq t_1,\ldots,t_k\right) \\
= & k!\sum_{t_k=2(k-1)+1}^{T}\ \sum_{t_{k-1}=2(k-2)+1}^{t_k-2}\cdots\sum_{t_2=3}^{t_3-2}\sum_{t_1=1}^{t_2-2}\Pr\left(y_{t_1}=i_1,\ldots,y_{t_k}=i_k \text{ and } y_t\notin\mathcal{A} \text{ for all } s\neq t_1,\ldots,t_k\right),
\end{aligned}$$

the approximation in the second line following because there are $k!\binom{T}{k}$ choices of $\{t_1,\ldots,t_k\}$ from $\{1,\ldots,T\}^k$ such that $t_i\neq t_j$ for all $i\neq j$ in the first sum, and there are $k!\binom{T-k+1}{k}$ choices of $\{t_1,\ldots,t_k\}$ such that $|t_i-t_j|\geq 2$ for all $i\neq j$ in the second sum, and $\binom{T-k+1}{k}\Big/\binom{T}{k}\to 1$ implying the omitted terms in the second sum are negligible. Applying Lemma A.2 then gives

$$\begin{aligned}
\Pr\left(N_{T,i}=1 \text{ for all } i\in\mathcal{A}\right) &\approx k!T^{-k}\binom{T-k+1}{k}\left(\prod_{i\in\mathcal{A}}h_\beta\mu_u\pi_{i-1}\right)\exp\left(-h_\beta\mu_u\sum_{i\in\mathcal{A}}\pi_{i-1}\right) \\
&\approx \prod_{i\in\mathcal{A}}\left(h_\beta\mu_u\pi_{i-1}\exp\left(-h_\beta\mu_u\pi_{i-1}\right)\right), \qquad \text{(A.23)}
\end{aligned}$$

again the product of marginal Poisson probabilities. Clearly the steps leading to (A.22) and (A.23) can be combined to give

$$\begin{aligned}
& \Pr\left(N_{T,i}=0 \text{ for all } i\in\mathcal{A}_0 \text{ and } N_{T,i}=1 \text{ for all } i\in\mathcal{A}_1\right) \\
\approx & \prod_{i\in\mathcal{A}_0}\exp\left(-h_\beta\mu_u\pi_{i-1}\right)\cdot\prod_{i\in\mathcal{A}_1}\left(h_\beta\mu_u\pi_{i-1}\exp\left(-h_\beta\mu_u\pi_{i-1}\right)\right),
\end{aligned}$$

for disjoint finite subsets $\mathcal{A}_0$ and $\mathcal{A}_1$ of $\mathcal{U}^{(1)}$. Generic values $N_{T,i}=k$ can also be included using the steps leading to (A.21). This proves the joint convergence of finite collections of $N_{T,i}$ to the stated independent Poisson distributions. If $\mathcal{U}^{(1)}$ is infinite then the convergence of infinite collections follows from that of the finite collections by application of Example 2.4 of Billingsley (1999).

A.5 Proof of Theorem 4

The marginal probabilities of y_t for any i can be found from

$$\begin{aligned}
\Pr\left(y_t=i\right) &= \Pr\left(y_t=i \text{ and } y_s\in\mathbb{N} \text{ for all } s<t\right) \\
&= \sum_{j_{t-1}\in\mathbb{N}}\cdots\sum_{j_1\in\mathbb{N}}\sum_{j_0\in\mathcal{U}^{(0)}}p_{i|j_{T-1}}\cdots p_{j_{t+1}|i}p_{i|j_{t-1}}\cdots p_{j_1|j_0}\pi_{j_0} \\
&\approx \sum_{j_{t-1}\in\mathbb{N}}\cdots\sum_{j_1\in\mathbb{N}}\sum_{j_0\in\mathcal{U}^{(0)}}q_{i|j_{T-1}}\cdots q_{j_{t+1}|i}q_{i|j_{t-1}}\cdots q_{j_1|j_0}\pi_{j_0},
\end{aligned}$$

and the remainders from the approximation in the second step handled in the same way as in Lemma A.2. Then using

$$\sum_{j\in\mathbb{N}}\begin{pmatrix}1\\ \beta_T j\end{pmatrix}\begin{pmatrix}\pi_j\\ \pi_{j-1}-\pi_j\end{pmatrix}'=\begin{pmatrix}1 & 0\\ \beta_T\mu_u & \beta_T\end{pmatrix}=C,$$

gives

$$\begin{aligned}\Pr\left(y_t=i\right) &\approx \sum_{j_{t-1}\in\mathbb{N}}\cdots\sum_{j_1\in\mathbb{N}}\sum_{j_0\in\mathcal{U}^{(0)}} q_{i|j_{T-1}}\cdots q_{j_{t+1}|i}q_{i|j_{t-1}}\cdots q_{j_1|j_0}\pi_{j_0}\\ &= \begin{pmatrix}\pi_i\\ \pi_{i-1}-\pi_i\end{pmatrix}' C^t\begin{pmatrix}1\\0\end{pmatrix}\\ &= \begin{pmatrix}\pi_i\\ \pi_{i-1}-\pi_i\end{pmatrix}'\begin{pmatrix}1 & 0\\ \mu_u\sum_{j=1}^t\beta_T^j & \beta_T^t\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}\\ &= \pi_i+\beta_T\mu_u\left(\pi_{i-1}-\pi_i\right)\frac{1-\beta_T^t}{1-\beta_T}. \qquad (A.24)\end{aligned}$$

It follows that for any $i\in\mathcal{Y}_T$,

$$\hat{\pi}_i=\pi_i+O_p\left(T^{-1/2}\right).$$

Thus if $i\in\mathcal{U}^{(k)}$, $k\geq 1$, then $\pi_i=0$ and $\hat{\pi}_i=O_p\left(T^{-1/2}\right)$. We use this consistency of $\hat{\pi}_i$ throughout the following proof.

We begin with the denominator of $\hat{\xi}_T$, which is relatively simpler than the numerator but illustrates the non-standard features of the limit theory under local alternatives. First we have $\bar{y}=\mu_u+O_p\left(T^{-1/2}\right)$ and $\hat{\sigma}_y^2=\sigma_u^2+O_p\left(T^{-1/2}\right)$ as usual, but $\hat{\sigma}_g^2$ requires more careful analysis. From (A.24)

$$\Pr\left(N_{T,i}>0\right)\approx 1-\left(1-\pi_i\right)^T\to 1 \text{ for } i\in\mathcal{U}^{(0)}, \qquad (A.25)$$

while Lemma 1 implies that

$$\Pr\left(N_{T,i}>0\right)\to 1-\exp\left(-h_\beta\mu_u\pi_{i-1}\right)\ \text{ for } i\in\mathcal{U}^{(1)}. \qquad (A.26)$$

Lemma 1 further implies that

$$\Pr\left(N_{T,i}>0\right)\to 0 \text{ for } i\in\mathcal{U}^{(k)},\ k>1. \qquad (A.27)$$

We can represent $\hat{g}_t=\hat{\pi}_{y_t-1}/\hat{\pi}_{y_t}$ as

$$\hat{g}_t=\sum_{i\in\mathcal{Y}_T}1_i\left(y_t\right)\frac{\hat{\pi}_{i-1}}{\hat{\pi}_i}=\sum_{k=0}^{\infty}\sum_{i\in\mathcal{U}^{(k)}:N_{T,i}>0}1_i\left(y_t\right)\frac{\hat{\pi}_{i-1}}{\hat{\pi}_i}.$$

Using this representation and applying (A.25), (A.26) and (A.27) we have

$$T^{-1}\sum_{t=1}^T\hat{g}_t=\sum_{k=0}^{\infty}\sum_{i\in\mathcal{U}^{(k)}:N_{T,i}>0}\hat{\pi}_{i-1}\rightsquigarrow\sum_{i\in\mathcal{U}^{(0)}}\pi_{i-1}+\sum_{i\in\mathcal{U}^{(1)}:N_i>0}\pi_{i-1},$$

with N_i the weak limit of $N_{T,i}$ defined in Lemma 1. This is a random limit because the second sum is over a random set $i \in \mathcal{U}^{(1)} : N_i > 0$. Similarly

$$\hat{g}_t^2 = \sum_{k=0}^{\infty} \sum_{i \in \mathcal{U}^{(k)} : N_{T,i} > 0} 1_i(y_t) \frac{\hat{\pi}_{i-1}^2}{\hat{\pi}_i^2},$$

so that

$$\begin{aligned} T^{-1} \sum_{t=1}^{T} \hat{g}_t^2 &= \sum_{k=0}^{\infty} \sum_{i \in \mathcal{U}^{(k)} : N_{T,i} > 0} \frac{\hat{\pi}_{i-1}^2}{\hat{\pi}_i} \\ &\approx \sum_{i \in \mathcal{U}^{(0)}} \frac{\hat{\pi}_{i-1}^2}{\hat{\pi}_i} + T \sum_{i \in \mathcal{U}^{(1)} : N_{T,i} > 0} \frac{\hat{\pi}_{i-1}^2}{N_{T,i}} \end{aligned} \tag{A.28}$$

where, using (A.27), we have omitted the asymptotically disappearing terms for $i \in \mathcal{U}^{(k)}$, $k > 1$ in the approximation. Since

$$\sum_{i \in \mathcal{U}^{(0)}} \frac{\hat{\pi}_{i-1}^2}{\hat{\pi}_i} \xrightarrow{p} \sum_{i \in \mathcal{U}^{(0)}} \frac{\pi_{i-1}^2}{\pi_i}$$

and

$$\sum_{i \in \mathcal{U}^{(1)} : N_{T,i} > 0} \frac{\hat{\pi}_{i-1}^2}{N_{T,i}} \rightsquigarrow \sum_{i \in \mathcal{U}^{(1)} : N_i > 0} \frac{\pi_{i-1}^2}{N_i} \tag{A.29}$$

we see that the correct standardisation of $\sum_{t=1}^{T} \hat{g}_t^2$ (i.e. T^{-2} or T^{-1}) depends on whether observations remain in $\mathcal{U}^{(1)}$ or not as T increases. The limit also varies, being either the fixed quantity $\sum_{i \in \mathcal{U}^{(0)}} \pi_{i-1}^2 / \pi_i$ (if $N_i = 0$ for all $i \in \mathcal{U}^{(1)}$) or the random variable $\sum_{i \in \mathcal{U}^{(1)} : N_i > 0} \pi_{t-1}^2 / N_i$. That is, we obtain the standard limit for $N_i = 0$ for all $i \in \mathcal{U}^{(1)}$ but a non-standard one for $N_i > 0$ for any $i \in \mathcal{U}^{(1)}$.

This same issue with standardisation and differing limits arises in the numerator of $\hat{\xi}_T$. Similarly to (A.28), we write the numerator in terms of its $\mathcal{U}^{(0)}$ and $\mathcal{U}^{(1)}$ components as:

$$\begin{aligned} T^{-1/2} \sum_{t=1}^{T} (y_{t-1} - \bar{y})(\hat{g}_t - 1) &\approx T^{-1/2} \sum_{t=1}^{T} (y_{t-1} - \mu_u) \sum_{i \in \mathcal{U}^{(0)}} 1_i(y_t) \left(\frac{\hat{\pi}_{i-1}}{\hat{\pi}_i} - 1 \right) \\ &+ T^{1/2} \sum_{i \in \mathcal{U}^{(1)} : N_{T,i} > 0} \sum_{t=1}^{T} (y_{t-1} - \mu_u) 1_i(y_t) \frac{\pi_{i-1}}{N_{T,i}}. \end{aligned} \tag{A.30}$$

For the limit of the first sum of $\mathcal{U}^{(0)}$ components, under local alternatives of the fast rate $\beta_T = T^{-1} h_\beta$, the proof of Theorem 3 given under the null carries through with only the addition of extra $O_p(T^{-1})$ remainders, giving the same result

$$T^{-1/2} \sum_{t=1}^{T} (y_{t-1} - \mu_u) \sum_{i \in \mathcal{U}^{(0)}} 1_i(y_t) \left(\frac{\hat{\pi}_{i-1}}{\hat{\pi}_i} - 1 \right) \rightsquigarrow N\left(0, \sigma_u^2 \sigma_g^2\right),$$

This would be the expected distribution for a statistic under alternatives that approach the null at $O\left(T^{-1}\right)$ instead of the usual $O\left(T^{-1/2}\right)$. However, if $N_{T,i}>0$ for any $i\in\mathcal{U}^{(1)}$ then the second term in (A.30) becomes dominant. We may write

$$\begin{aligned}\sum_{i\in\mathcal{U}^{(1)}:N_{T,i}>0}\sum_{t=1}^{T}\left(y_{t-1}-\mu_u\right)1_i\left(y_t\right)\frac{\pi_{i-1}}{N_{T,i}} &= \sum_{i\in\mathcal{U}^{(1)}:N_{T,i}>0}\pi_{i-1}\frac{\sum_{t=1}^{T}y_{t-1}1_i\left(y_t\right)-\mu_u N_{T,i}}{N_{T,i}}\\ &= \sum_{i\in\mathcal{U}^{(1)}:N_{T,i}>0}\pi_{i-1}\left(\frac{S_{T,i}}{N_{T,i}}-\mu_u\right),\end{aligned}$$

where

$$S_{T,i}=\sum_{t=1}^{T}y_{t-1}1_i\left(y_t\right).$$

and consider the joint convergence of $S_{T,i},N_{T,i}$ for $i\in\mathcal{U}^{(1)}$. Both are non-decreasing in T and therefore may be sub-martingales. To check this, consider $E\left|N_{T,i}\right|$ and $E\left|S_{T,i}\right|$. For $i\in\mathcal{U}^{(1)}$ (A.24) reduces to

$$\Pr\left(y_t=i\right)\approx\beta_T\mu_u\pi_{i-1},$$

so that

$$E\left|N_{T,i}\right|=\sum_{t=1}^{T}\Pr\left(y_t=i\right)\approx h_\beta\mu_u\pi_{i-1}<\infty$$

and similarly

$$E\left|S_{T,i}\right|=\sum_{t=1}^{T}E\left(y_{t-1}1_i\left(y_t\right)\right)\leq\sum_{t=1}^{T}E\left(y_{t-1}^2\right)^{1/2}\Pr\left(y_t=i\right)<\infty,$$

since $E\left(y_{t-1}^2\right)<\infty$ by Lemma 1(a) of Drost *et al* (2009) . Both $N_{T,i}$ and $S_{T,i}$ are therefore L_1 bounded sub-martingales and, from Theorem 35.5 of Billingsley (1995), have almost sure limits N_i and S_i respectively, where $E\left|N_i\right|<\infty$ and $E\left|S_i\right|<\infty$, which implies weak convergence as well. This convergence is joint for $\{N_{T,i},S_{T,i}\}$ and is automatically joint across all i if $\mathcal{U}^{(1)}$ is a finite set. If $\mathcal{U}^{(1)}$ is not a finite set then we again invoke Example 2.4 of Billingsley (1999). Therefore we can conclude in (A.30) that

$$\sum_{i\in\mathcal{U}^{(1)}:N_{T,i}>0}\pi_{i-1}\left(\frac{S_{T,i}}{N_{T,i}}-\mu_u\right)\rightsquigarrow\sum_{i\in\mathcal{U}^{(1)}:N_i>0}\pi_{i-1}\left(\frac{S_i}{N_i}-\mu_u\right).\tag{A.31}$$

While the distribution of N_i is known (i.e. Poisson), the distribution of S_i depends on Π and h_β in a more complicated way and a known form has not been found. It is evident, however, that S_i and N_i can be expected to be positively dependent since, for given Π, larger values of $N_{T,i}$ (i.e. larger numbers of t for which $1_i(y_t)=1$) imply additional non-negative terms in the sum $S_{T,i}$. Unreported simulations for several choices of Π are consistent with this dependence. Therefore (A.31) can be taken as a well-defined representation of a non-degenerate limiting distribution, but one which remains dependent on nuisance parameters and reliant on simulation for further exploration of its properties.

Define the indicator random variable $Q_T = 1\left(N_{T,i} > 0 \text{ for any } i \in \mathcal{U}^{(1)}\right)$ and its limit random variable $Q = 1\left(N_i > 0 \text{ for any } i \in \mathcal{U}^{(1)}\right)$. The probability $\Pr(Q=1) = 1 - \exp\left(-h_\beta \mu_u \pi^{(1)}\right)$ follows directly from (A.26).

Then we can write

$$\begin{aligned}
\hat{\xi}_T &= \frac{T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1}-\bar{y}\right)\left(\hat{g}_t - 1\right)}{\hat{\sigma}_y \hat{\sigma}_g} \\
&\approx \frac{T^{-1/2}\sum_{t=1}^{T}\left(y_{t-1}-\mu_u\right)\sum_{i\in\mathcal{U}^{(0)}} 1_i\left(y_t\right)\left(\frac{\hat{\pi}_{i-1}}{\hat{\pi}_i}-1\right)}{\sigma_u\left(\sum_{i\in\mathcal{U}^{(0)}}\frac{\hat{\pi}_{i-1}^2}{\hat{\pi}_i}-\left(\sum_{i\in\mathcal{U}^{(0)}}\hat{\pi}_{i-1}\right)^2\right)^{1/2}} \\
&\quad + \frac{\sum_{i\in\mathcal{U}^{(1)}:N_{T,i}>0}\sum_{t=1}^{T}\left(y_{t-1}-\mu_u\right)1_i\left(y_t\right)\frac{\hat{\pi}_{i-1}}{N_{T,i}}}{\sigma_u\left(\sum_{i\in\mathcal{U}^{(1)}:N_{T,i}>0}\frac{\hat{\pi}_{i-1}^2}{N_{T,i}}\right)} \\
&= \hat{\xi}_{T,0} + \hat{\xi}_{T,1}
\end{aligned}$$

Considering $\hat{\xi}_{T,1}$, if $Q=0$ there are no elements in $\mathcal{U}^{(1)}$ asymptotically and hence $\hat{\xi}_{T,1}$ disappears. On the other hand if $Q=1$ and there are elements in $\mathcal{U}^{(1)}$, $\hat{\xi}_{T,1}$ automatically stabilises itself, with matching behaviour in the numerator and denominator using (A.31) and (A.29). Thus $\hat{\xi}_{T,1}$ converges to

$$X = \frac{1}{\sigma_u}\left(\sum_{i\in\mathcal{U}^{(1)}:N_i>0}\frac{\pi_{i-1}^2}{N_i}\right)^{-1/2}\sum_{i\in\mathcal{U}^{(1)}:N_i>0}\pi_{i-1}\left(\frac{S_i}{N_i}-\mu_u\right)$$

under $Q=1$. Thus $\hat{\xi}_{T,1} \rightsquigarrow X_Q$ where $X_Q = 0$ if $Q=0$ and $X_Q = X$ if $Q=1$.

Considering $\hat{\xi}_{T,0}$, if $Q=0$ then $\hat{\xi}_{T,0}$ converges to $Z \sim N(0,1)$ (as in the null case) under the fast rate of the alternatives. If $Q=1$, $\hat{\xi}_{T,0}$ converges to zero in probability due to the explosive nature of $\hat{\sigma}_g$ (again using (A.29)). Thus $\hat{\xi}_{T,0} \rightsquigarrow Z_Q$ where $Z_Q = Z$ if $Q=0$ and $Z_Q = 0$ if $Q=1$.

Putting together these limits for $\hat{\xi}_{T,0}$ and $\hat{\xi}_{T,1}$ gives the conclusion of the theorem

$$\hat{\xi}_T \rightsquigarrow Z_Q + X_Q$$

as stated. This representation is not canonical in the sense that the random variables involved are not all independent or standard. Nevertheless, jointly $\left\{Q, \left(N_i, S_i, i \in \mathcal{U}^{(1)}\right)\right\}$ is a well defined asymptotic distribution and hence so is $Z_Q + X_Q$. ■