Testing for the Mixture Hypothesis of Poisson Regression Models

JIN SEO CHO

School of Economics, Yonsei University

Seodaemun-gu, Seoul 03722, Korea

jinseocho@yonsei.ac.kr

August 2025

Abstract

This paper considers testing the mixture hypothesis of Poisson regression models using the likelihood ratio (LR) test. The main motivation for the mixture hypothesis is the unobserved heterogeneity. The null hypothesis of interest in this paper is that there is no unobserved heterogeneity in the data. Due to the nonstandard conditions described in the text, the LR test does not weakly converge to the standard chi-squared random variable under the null. We derive its limiting distribution by assuming popular Poisson regression models in terms of a Gaussian process. Furthermore, we discuss how to obtain the asymptotic critical values of the LR test consistently. Finally, we implement various Monte Carlo experiments and compare its power with the specification test proposed by Lee (1986). The simulation shows that the LR test is more powerful than the specification test when its small sample size distortion is adjusted.

Key Words: Mixture of Poisson Regression Models; Likelihood Ratio Test; Asymptotic Null Distribution; Gaussian Process.

JEL Classifications: C12, C22, C32, C52.

Acknowledgments: The co-editor, Cheolwoo Park, and two anonymous referees provided very helpful comments for which I am most grateful. Cho acknowledges the research grant provided by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF2020S1A5-A2A0104-0235).

1 Introduction

Poisson regression models are popularly applied for count data. For example, Hausman, Hall, and Griliches (1984) provide stylized econometric model specifications for count data using Poisson regression models.

Misspecified Poisson models have been investigated in the literature. For example, Lee (1986) provides a specification test statistic for Poisson regression models. In the econometrics literature, Gourieroux, Monfort, and Trognon (1984) examine misspecified Poisson regression models and relevant tests. These misspecified models are often believed to be due to the presence of unobserved heterogeneity. In the statistics literature, we find similar works, and a specific distribution is typically assumed for unobserved heterogeneity. The most popular distribution assumption is a Bernoulli or Binomial distribution. This assumption results in a finite mixture of Poisson regression models, overcoming model misspecification. For example, Karlis and Xekalaki (1999, 2001) and Schlattmann (2003) consider estimating the number of components in a finite mixture of Poisson regression models. They rely on a computationally intensive re-sampling procedure.

Nevertheless, testing the mixture hypothesis of the Poisson regression models has not been successfully resolved. As examined in the literature of mixture, testing the hypothesis using the standard likelihood ratio (LR) statistic has identification and boundary parameter problems. Thus, without resolving these issues, the asymptotic null distribution of the LR statistic cannot be determined effectively.

The main goal of this paper is, therefore, to demonstrate the use of the LR statistic designed to test for the mixture hypothesis of the Poisson regression models. For this, we rely on the methodology of Cho and White (2007). They provide a set of regularity conditions to test the mixture hypothesis for general mixture models and demonstrate the application of this methodology to testing the mixture of normals. We apply their methodology to the mixture of Poisson regression models by deriving the null limit distribution of the LR test. Furthermore, we provide a simulation method to deliver the asymptotic critical values consistently. In achieving this goal, we specifically assume the Poisson regression model specified by Hausman et al. (1984), although their exponential assumption is relaxed.

In the literature, testing the mixture hypothesis has been examined by a number of authors. Hartigan (1985) considers an example of a normal mixture to demonstrate that the null limit distribution of the LR test is dependent upon the parameter space unidentified under the null. Ghosh and Sen (1985) derive the null limit distribution of the LR test under the so–called strong identification assumption. Chernoff and Lander (1995) develop Ghosh and Sen's (1985) methodology to the case of mixture of binomial distributions. They also introduce a simulation method to deliver the asymptotic critical values consistently. Dacunha-Castelle and Gassiat (1999) examine general mixture models and apply their polar conic parametrization method to test the mixture hypothesis. Chen and Chen (2001) also examine the same problem using another methodology, affirming the results in Dacunha-Castelle and Gassiat (1999). In particular, Chen and Chen (2001) examine the simplest mixture of Poisson distributions whose model scope is extended in this paper. Cho and White (2007) note that many popular mixtures require much higher–order approximations than those examined in the prior literature. Due to this, they extend the mixture scope up to the case where models are differentiable eight–times continuously. Cho and White (2010) apply their methodology to the case of mixture of exponential or Weibull distributions. Cho and White (2007, 2010) also provide simulation methods delivering asymptotic critical values consistently in a similar manner to Chernoff and Lander (1995).

The plan of this paper is as follows. In Section 2, we consider a mixture of Poisson regression models and derive the null limit distribution of the LR test. We further introduce a simulation method to deliver the asymptotic critical values consistently. In Section 3, we conduct Monte Carlo simulations, and some concluding remarks are given in Section 4. Finally, we collect the regularity conditions for the Poission mixture model in the Appendix.

2 Mixture of Poisson Regression Models

Suppose that the Poisson regression mixture model is correctly specified. When a sequence of identically and independently distributed (IID) random variables $\{(X_t, \mathbf{Z}_t) \in \mathbb{N} \times \mathbb{R}^k\}$ can be correctly provided as follows:

$$m(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta}; X_t|\mathbf{Z}_t) = \pi f[g(\mathbf{Z}_t; \lambda_1, \boldsymbol{\beta}); X_t] + (1-\pi) f[g(\mathbf{Z}_t; \lambda_2, \boldsymbol{\beta}); X_t],$$

where for each λ and $\boldsymbol{\beta}$,

$$f[g(\mathbf{Z}_t; \lambda, \boldsymbol{\beta}); X_t = k] = \frac{\exp[-g(\mathbf{Z}_t; \lambda, \boldsymbol{\beta})] g(\mathbf{Z}_t; \lambda, \boldsymbol{\beta})^k}{k!},$$

and $\mathbf{Z}_t$ denotes a vector of covariates not including the constant. This is assumed for an identification purpose.

The motivation of this specification is due to the presence of unobserved heterogeneity. The prior literature points out that estimating the key parameter $\boldsymbol{\beta}_*$ is critically affected by unobserved heterogeneity (see (see Hausman et al., 1984; Heckman and Singer, 1984; Gourieroux et al., 1984; van den Berg and Ridder, 1998; Wooldridge, 1999; Cho and White, 2010, and references therein). Specifically, we suppose that there exists another variable independent of $(X_t, \mathbf{Z}_t)$, say W_t, affecting the conditional distribution of $X_t|\mathbf{Z}_t$

through λ, so that the general mixture model can be written as

$$f[g(\mathbf{Z}_t;\lambda,\boldsymbol{\beta}); X_t=k] = \int \frac{\exp[-g(\mathbf{Z}_t;\lambda(w),\boldsymbol{\beta})]g(\mathbf{Z}_t;\lambda(w),\boldsymbol{\beta})^k}{k!}h(w)dw,$$

where $h(\cdot)$ is the probability distribution function of W_t. The mixture model $m(\pi,\lambda_1,\lambda_2,\boldsymbol{\beta};X_t|\mathbf{Z}_t)$ is obtained by assuming that the heterogeneity variable W_t conforms to a Bernoulli distribution. We do not specify the functional form of $g(\cdot)$. It does not have to be the exponential function that is popular in the literature (see Hausman et al., 1984, for example). Without assuming a particular form of $g(\cdot)$, we proceed with our discussions. The regularity conditions for this are provided in the Appendix.

Given this, we can exploit the LR test principle to test for the mixture hypothesis. For this, we suppose that $(\pi_*,\lambda_{1*},\lambda_{2*},\boldsymbol{\beta}_*)$ maximizes $\mathbb{E}[\log\{m(\,\cdot\,;X_t|\mathbf{Z}_t)\}]$, and we test the following hypotheses: for some unknown and unique $\lambda_*\in[\underline{\lambda},\bar{\lambda}]$,

$$H_0:\pi_*=1,\lambda_{1*}=\lambda_*,;\ \pi_*=0,\lambda_{2*}=\lambda_*;\ \text{or}\ \lambda_{1*}=\lambda_{2*}=\lambda_* \quad \text{versus}$$

$$H_1:\pi_*\in(0,1)\ \text{and}\ \lambda_{1*}\neq\lambda_{2*}.$$

Here, $\underline{\lambda}$ and $\bar{\lambda}$ are lower and upper bounds of λ greater than zero, respectively. Note that the null model implies that the Poisson regression model is correctly specified for the distribution of X_t on $\mathbf{Z}_t$.

The null hypothesis is different from the standard null in the literature. It is a joint hypothesis, describing the Poisson regression model, and two non-standard problems are implied. First, there is an identification problem. If $\pi_*=1$ (resp. $\pi_*=0$), then λ_{2*} (resp. λ_{1*}) is not identified. Likewise, if $\lambda_{1*}=\lambda_{2*}$, then π_* is not identified. These are so-called Davies' (1977; 1987) identification problem. Second, if $\pi_*=1$ or 0, then π_* is on the boundary of parameter space, so that the interiority problem violates for the LR test to behave regularly under the null hypothesis.

In the prior literature, the null limit distribution of the LR test is obtained by overcoming the nonstandard problems. These nonstandard aspects are examined by numerous authors. For example, Ghosh and Sen (1985) examine the null limit distribution of the LR test under the strong identification assumption, and Chernoff and Lander (1995) apply Ghosh and Sen's (1985) methodology to the case of binomial mixtures. Dacunha-Castelle and Gassiat (1999) examine general mixture models and apply their polar conic parametrization method to test the mixture hypothesis. Chen and Chen (2001) also examine the same problem, including the simple Poisson mixture. In particular, Cho and White (2007) assume a general mixture model and derive the null limit distribution of the LR test generically. Following the methodology of Cho and White (2007), the null limit distributions of the LR tests are further investigated for specific mixture models. Cho and Han (2009) and Cho, Park, and Park (2018) focus on the geometric mixture, and Cho and White (2010) focus the exponential and Weibull mixtures. Furthermore, Cho (2025) focuses on the normal mixtures. All the null limit distributions of the LR test are different as it depends on the model properties.

By applying the general framework in Cho and White (2007) to the current Poisson mixture model, we here provide the null limit distribution of the LR test. Under the regularity conditions in the Appendix, the LR test has the following null weak limit:

$$LR_n \Rightarrow \sup_{\lambda \in [\underline{\lambda},\ \bar{\lambda}]} \min^2[0, \mathscr{Y}(\lambda)],$$

where

$$LR_n := 2n \left\{ \max_{\pi,\lambda_1,\lambda_2,\boldsymbol{\beta}} \sum_{t=1}^{n} \log(m(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta}; X_t|\mathbf{Z}_t)) - \max_{\lambda,\boldsymbol{\beta}} \sum_{t=1}^{n} \log(f(g(\mathbf{Z}_t; \lambda, \boldsymbol{\beta}); X_t)) \right\},$$

and $\mathscr{Y}$ is a Gaussian process such that for each λ, λ',

$$\mathbb{E}[\mathscr{Y}(\lambda)\mathscr{Y}(\lambda')] = \frac{r(\lambda, \lambda')}{\sqrt{r(\lambda, \lambda)}\sqrt{r(\lambda', \lambda')}}, \tag{1}$$

such that

$$\begin{aligned} r(\lambda, \lambda') :=& \mathbb{E}[\exp\{g(\mathbf{Z}_t; \boldsymbol{\beta}_*)(\sqrt{\lambda_*} - \lambda/\sqrt{\lambda_*})(\sqrt{\lambda_*} - \lambda'/\sqrt{\lambda_*})\}] \\ &- 1 - \mathbb{E}[g(\mathbf{Z}_t; \boldsymbol{\beta}_*)(\sqrt{\lambda_*} - \lambda/\sqrt{\lambda_*})(\sqrt{\lambda_*} - \lambda'/\sqrt{\lambda_*})]. \end{aligned}$$

The covariance kernel $r(\cdot, \cdot)$ implies that the null limit distribution is affected by the distribution of $g(\mathbf{Z}_t; \boldsymbol{\beta}_*)$. We consider two different cases. First, if $g(\cdot; \cdot) \equiv 1$, (1) reduces to the simplest mixture case of Poisson distributions without covariates. Chen and Chen (2001) examine the null limit distribution of the LR test using another approach and obtain the same covariance kernel. Second, if the distribution of $g(\mathbf{Z}_t; \boldsymbol{\beta}_*)$ is nontrivial, the covariance kernel can have different functional forms for different distributions. For such a case, it is useful to exploit a computationally intensive testing procedure for the LR test such as the parametric bootstrap (see Amengual, Bei, Carrasco, and Sentana, 2025, for example).

One of the notable things with this null limit distribution is in the fact that it depends on the size of parameter space $[\underline{\lambda},\ \bar{\lambda}]$ as pointed out by Hartigan (1985). Certainly, if a bigger parameter space is assumed, bigger critical values are obtained. Prior literature ignoring this aspect reports simulation results whose critical values do not appear to converge. When obtaining bootstrap–based critical values, different specifications for the parameter space are expected to produce different testing results (see Karlis and Xekalaki, 1999, 2001; Schlattmann, 2003, for example).

We now consider methodologies to obtain the asymptotic critical values consistently or their approximations. First, we suppose that $g(\cdot;\cdot) \equiv 1$. For this case, the asymptotic critical values of the LR test can be efficiently obtained by following the approximation method in Cho (2006). That is, we can provide an analytical Gaussian process having the same covariance structure as (1), so that we can obtain the asymptotic critical values by simulation. For each $\lambda \in [\underline{\lambda},\ \bar{\lambda}]$, let

$$\mathscr{G}(\lambda;\lambda_*) := \frac{1}{\sqrt{\rho(\lambda,\lambda)}}\sum_{j=2}^{\infty}\frac{1}{\sqrt{j!}}\left(\sqrt{\lambda_*}-\frac{\lambda}{\sqrt{\lambda_*}}\right)^j W_j, \tag{2}$$

where $W_j \sim$ IID $N(0,1)$. Then it is not hard to verify that $\mathbb{E}[\mathscr{Y}(\lambda)\mathscr{Y}(\lambda')] = \mathbb{E}[\mathscr{G}(\lambda;\lambda_*)\mathscr{G}(\lambda';\lambda_*)]$. Thus, we can simulate

$$\sup_{\lambda\in[\underline{\lambda},\ \bar{\lambda}]} \min{}^2[0, \mathscr{G}(\lambda;\lambda_*)]$$

many times to obtain the asymptotic critical values. The empirical distribution obtained in this way can consistently deliver the asymptotic null distribution of the LR test. While implementing this procedure, we note that one of the ingredients of $\mathscr{G}$ is λ_*, which is unknown. This unknown parameter can be estimated consistently. For example, we can estimate it using the null model. The Monte Carlo experiments given below verify that the parameter estimation error can be neglected if the sample size is moderately large.

Second, we again suppose that $g(\cdot;\cdot) \equiv 1$. There is another analytical Gaussian process whose covariance kernel is identical to (1). For this provision, for each ξ, we let

$$\mathscr{X}(\xi) := \frac{1}{\sqrt{s(\xi,\xi)}}\sum_{j=2}^{\infty}\frac{\xi^j}{\sqrt{j!}}W_j, \tag{3}$$

where for each ξ, ξ', $s(\xi,\xi') := \exp(\xi\xi') - 1 - \xi\xi'$. Note that (3) is obtained by letting $\xi := \sqrt{\lambda_*} - \lambda/\sqrt{\lambda_*}$. Thus, we can alternatively simulate

$$\sup_{\xi\in[\underline{\xi},\ \bar{\xi}]} \min{}^2[0, \mathscr{X}(\xi)] \tag{4}$$

to obtain the asymptotic critical values, where $\underline{\xi}$ and $\bar{\xi}$ are $\sqrt{\lambda_*}-\bar{\lambda}/\sqrt{\lambda_*}$ and $\sqrt{\lambda_*}-\underline{\lambda}/\sqrt{\lambda_*}$, respectively. We also note that $\mathscr{X}(\cdot)$ is the Hermite Gaussian process introduced by Cho and White (2007) and Cho (2025), which is the Gaussian process obtained while testing for the mixture normal. Although there is no direct relationship between the Poisson mixture and the normal mixture, the same Gaussian process is obtained to characterize the null limit distribution of the LR test. For other mixtures, the null limit distribution of the LR test is characterized by different Gaussian processes whose covariance kernels are different from $\mathscr{X}(\cdot)$. Cho and Han (2009), Cho and White (2010), and Cho (2025) provide analytical Gaussian processes different from (3) to obtain the null limit distribution of the LR test by simulation.

Third, we consider the null limit distribution of the LR test when $g(\mathbf{Z}_t; \boldsymbol{\beta}_*)$ has a nontrivial distribution. We first define a conditional Gaussian process given covariate $\mathbf{Z}_t = \boldsymbol{z}$ as

$$\tilde{\mathscr{X}}(\delta|\mathbf{Z}_t = \boldsymbol{z}) := \frac{1}{\sqrt{s(\delta(\boldsymbol{z}), \delta(\boldsymbol{z})')}} \sum_{j=2}^{\infty} \frac{\delta(\boldsymbol{z})^j}{\sqrt{j!}} W_j$$

for each $\delta(\boldsymbol{z})$, where for each λ, $\delta(\boldsymbol{z}) := g(\boldsymbol{z}; \beta_*)^{1/2}(\sqrt{\lambda_*} - \lambda/\sqrt{\lambda_*})$. This is a Gaussian process forming the null limit distribution when the covariate $\mathbf{Z}_t$ is fixed at $\boldsymbol{z}$. That is, it follows that for given $\mathbf{Z}_t = \boldsymbol{z}$,

$$LR_n(\boldsymbol{z}) \Rightarrow \sup_{\lambda \in [\underline{\lambda},\, \bar{\lambda}]} \min^2[0, \tilde{\mathscr{X}}(\delta|\boldsymbol{z})].$$

Note that if $g(\cdot;\cdot) \equiv 1$, $\tilde{\mathscr{X}}(\cdot|\mathbf{Z}_t = \boldsymbol{z}) \equiv \mathscr{X}(\cdot)$. We further note that this weak limit can be rewritten as a function of $\mathscr{X}(\cdot)$ by transforming the domain of λ. That is, if we let $\underline{\nu}(\boldsymbol{z}) := \underline{\xi} g(\boldsymbol{z}; \boldsymbol{\beta})^{1/2}$ and $\bar{\nu}(\boldsymbol{z}) := \bar{\xi} g(\boldsymbol{z}; \boldsymbol{\beta})^{1/2}$, respectively, it trivially follows that

$$\sup_{\lambda \in [\underline{\lambda},\, \bar{\lambda}]} \min^2[0, \tilde{\mathscr{X}}(\delta|\boldsymbol{z})] = \sup_{\nu \in [\underline{\nu}(\boldsymbol{z}),\, \bar{\nu}(\boldsymbol{z})]} \min^2[0, \mathscr{X}(\nu)] \tag{5}$$

The random feature of $\tilde{\mathscr{X}}(\cdot)$ driven by $\mathbf{Z}_t$ is now transferred to the random parameter space $[\underline{\nu}(\boldsymbol{z}), \bar{\nu}(\boldsymbol{z})]$ on the right-hand side (RHS) of (5).

Next, we apply Piterbarg (1996) to handle the random parameter space and approximate the tail distribution of (5). By theorem 7.1 of Piterbarg (1996), it follows that as u tends to infinity, the unconditional tail probability for an extremum is given as

$$P\left(\sup_{\nu \in [\underline{\nu}(\boldsymbol{z}),\, \bar{\nu}(\boldsymbol{z})]} \min^2[0, \mathscr{X}(\nu)] > u^2\right) = H_\alpha \mathbb{E}[\boldsymbol{\lambda}([\underline{\nu}(\boldsymbol{z}),\ \bar{\nu}(\boldsymbol{z})])] u^{2/\alpha}(1 - \Phi(u))(1 + o(1)), \tag{6}$$

where H_α is the asymptotic double-sum coefficient defined in (Piterbarg, 1996, p. 16); $\boldsymbol{\lambda}(\,\cdot\,)$ stands for Lebesgue measure; and $\Phi(\,\cdot\,)$ is the standard normal cumulative distribution function (CDF). Therefore, if we let the RHS of (6) be the level of significance, its corresponding u^2 becomes the asymptotic critical value. Here, we note that (6) is effective for a substantially large u^2. In case u^2 is not large enough or the level of significance is not sufficiently small, the equality in (6) does not hold. For such a case, the critical value obtained from the equality in (6) becomes conservative. Due to this, we should treat the critical value obtained from (6) as a conservative approximation.

The RHS of (6) can also be approximated by using the Hermitian Gaussian process. The random parameter space for ν on the left-hand side (LHS) of (6) can be replaced by $[L_*, U_*] := [\underline{\xi}\omega_*, \bar{\xi}\omega_*]$, where $\omega_* := \mathbb{E}[g(\mathbf{Z}_t; \beta_*)^{1/2}]$. Thus, simulating

$$\sup_{\nu \in [L_*,\, U_*]} \min^2[0, \mathscr{X}(\nu)] \tag{7}$$

can deliver the tail asymptotic null distribution consistently. In case $[L_*, U_*]$ is unknown, we can consistently estimate it using the null model as before. That is, we can replace $\underline{\xi}$, $\bar{\xi}$, and ω_* by their estimates:

$$\widehat{\underline{\xi}}_n := (\widehat{\lambda}_n^0)^{1/2} - \bar{\lambda}/(\widehat{\lambda}_n^0)^{1/2}, \quad \widehat{\bar{\xi}}_n := (\widehat{\lambda}_n^0)^{1/2} - \underline{\lambda}/(\widehat{\lambda}_n^0)^{1/2}, \quad \text{and} \quad \widehat{\omega}_n^0 := n^{-1} \sum g(\mathbf{Z}_t; \widehat{\boldsymbol{\beta}}_n^0)^{1/2}, \tag{8}$$

respectively, where $(\widehat{\lambda}_n^0, \widehat{\boldsymbol{\beta}}_n^0)$ is the maximum-likelihood estimator (MLE) obtained from the null model assumption. The tail asymptotic null distribution can be consistently delivered by simulating the following many times:

$$\sup_{\nu \in [\widehat{L}_n,\, \widehat{U}_n]} \min^2[0, \mathscr{X}(\nu)], \tag{9}$$

where $[\widehat{L}_n,\ \widehat{U}_n] := [\widehat{\underline{\xi}}_n \widehat{\omega}_n^0,\ \widehat{\bar{\xi}}_n \widehat{\omega}_n^0]$.

3 Monte Carlo Experiments

We suppose that $\{(X_t, Z_t) \in \mathbb{N} \times \mathbb{R}\}$ is generated according to $X_t | Z_t \sim$ IID Pois$(2\exp(Z_t))$ and $Z_t \sim$ IID $U(-1, 1)$ for simulations under the null. Therefore, $g(\mathbf{Z}_t; \lambda_*, \boldsymbol{\beta}_*) = \lambda_* \exp(\beta_* Z_t)$ in terms of the notation in Section 2, and $(\lambda_*, \beta_*) = (2, 1)$. A mixture model for this is specified as follows:

$$\pi \text{Pois}(\lambda_1 \exp(\beta Z_t)) + (1 - \pi)\text{Pois}(\lambda_2 \exp(\beta Z_t)),$$

where $\pi \in [0, 1]$, $\lambda_1, \lambda_2 \in [1, 3]$, and there is no restriction on β. Here, the space for λ_1 and λ_2 are selected in a manner the mixture model of doubt is correctly specified under both null and alternative hypotheses. As $\lambda_* \exp(\beta_* Z_t)$ has a nontrivial distribution, we test for the Poisson mixture by applying the asymptotically approximated critical values in (6).

First, we examine the asymptotic critical values. Table 1 shows the critical values obtained under various assumptions on the sample size. These critical values are obtained by simulating (9) 50,000 times and contained in Table 1. Simulating (9) is not affected by the estimation error, and their differences decrease as the sample size increases. The last column shows the asymptotic critical values by supposing that $(\lambda_*, \beta_*) = (2, 1)$ is known, so that we obtain the critical values by simulating (9) with $[L_*, U_*] = [\underline{\xi}\omega_*, \bar{\xi}\omega_*]$, where

$\omega_* := \mathbb{E}[\sqrt{2}\exp(\frac{1}{2}Z_t)]$, $\underline{\xi} := \sqrt{\lambda_*} - \bar{\lambda}/\sqrt{\lambda_*}$, and $\bar{\xi} := \sqrt{\lambda_*} - \underline{\lambda}/\sqrt{\lambda_*}$. If the sample size is finite, we replace (λ_*, β_*) with $(\widehat{\lambda}_n^0, \widehat{\beta}_n^0)$, viz., the MLE obtained from the null model, so that the lower and upper bounds for ν can be estimated as given in (8) to simulate (9).

Second, we examine the empirical rejection rates of the LR test under the null. We contain the simulation results in Table 2. As shown in Table 2, the small sample size distortion exists and the distortion does not disappear although the sample size is substantially large. Furthermore, the difference between the nominal level and the empirical rejection rate increases as the nominal level increases. This feature shows that the critical values obtained by simulating (6) are conservative approximates. When the critical values are approximated by the tail probability for an extremum, this distortion is expected. Despite the small sample size distortion, we also note that the distortion size can be controlled by reducing the level of significance.

In addition to the null simulation, we examine the power properties of the LR test. For the power comparison, we employ another test. Lee (1986) considers the following specification test:

$$\mathscr{S}\mathscr{R} = \frac{1}{\sqrt{2n}} \sum_{t=1}^{n} \frac{\{X_t(X_t - 1) - \widehat{\lambda}_n^0 \exp(\widehat{\beta}_n^0 Z_t)\}}{\widehat{\lambda}_n^0 \exp(\widehat{\beta}_n^0 Z_t)},$$

which weakly converges to the standard normal random variable under the null. We compare the power of the LR test with that of $\mathscr{S}\mathscr{R}$ when the sample size is 100, and the number of repetition is 3,000. If the critical values obtained by the limiting and empirical distributions, we denote them as $\mathscr{L}\mathscr{R}'$ and $\mathscr{L}\mathscr{R}''$, respectively. Specifically, the empirical rejection rates indexed by $\mathscr{L}\mathscr{R}'$ are obtained by comparing the LR tests with the critical value in Table 1, viz., 3.6930. Meanwhile the rejection rates indexed by $\mathscr{L}\mathscr{R}''$ are computed by comparing the LR tests with the critical value from its empirical distribution. Note that 10,000 LR test values are already available under the null while filling in Table 2. Top 5% value of them is used as the critical value for the LR test to examine its power properties when it does not suffer from size distortion under the null. Specific data generating processes for $\{X_t|Z_t\}$ is $\pi_*\text{Pois}(\lambda_{1*}\exp(Z_t)) + (1-\pi_*)\text{Pois}(\lambda_{2*}\exp(Z_t))$ and $Z_t \sim$ IID $U(-1,1)$. The values of π_* and $(\lambda_{1*}, \lambda_{2*})$ are given in Table 3, such that λ_{1*} and $\lambda_{2*} \in [1,3]$. By this, the mixture model is correctly specified under the alternative hypothesis. When $\mathscr{L}\mathscr{R}'$ is compared with $\mathscr{S}\mathscr{R}$, the results are nuanced. When π_* approaches zero, $\mathscr{S}\mathscr{R}$ is more powerful than $\mathscr{L}\mathscr{R}'$. Otherwise, $\mathscr{L}\mathscr{R}'$ is more powerful than $\mathscr{S}\mathscr{R}$. Also, $\mathscr{L}\mathscr{R}'$ is more powerful than $\mathscr{S}\mathscr{R}$ as π_* is away from zero or one. This aspect is expected because $\mathscr{L}\mathscr{R}'$ does not accommodate the small sample size distortion. Nevertheless, these nuances disappear when $\mathscr{S}\mathscr{R}$ is compared to $\mathscr{L}\mathscr{R}''$. In every case, $\mathscr{L}\mathscr{R}''$ is more powerful than $\mathscr{S}\mathscr{R}$. From this feature, we can say that the specification test has a relatively low power in testing for the mixture hypothesis and the LR test has a respectful power properties.

Before moving to the next section, we provide additional remark on the power of the LR test. In case the space for λ_{1*} and λ_{2*} is selected too small, so that λ_{1*} and/or $\lambda_{2*} \notin [1, 3]$, then the power of the LR test can be smaller than that under the current environment. If the space is too small, then the model is misspecified, so that the MLE for $(\lambda_{1*}, \lambda_{2*})$ does not converge to the unknown parameter under the alternative, and this decreases the power of the LR test. This aspect implies that the mixture model has to be carefully specified so that it is correctly specified under both null and alternative hypotheses.

4 Conclusion

In this paper, we consider testing the mixture hypothesis of Poisson regression models by focusing on the popularly applied Poisson regression models. In particular, we employ the LR test for the goal of this paper by deriving its limit distribution under the null that the conditional distribution is the Poisson regression model.

In achieving the goal of this study, we exploit the methodology developed by Cho and White (2007). The main result is that the LR test weakly converges to a functional of the Hermitian Gaussian process in case the regressor does not exist. When the regressor does exist, the asymptotic critical values are further obtained by providing a simulation method delivered by the tail probability for an extremum.

In addition, we conduct Monte Carlo simulations to examine the performances of the LR test. Specifically, we examine the empirical size of the LR test by comparing it with the asymptotic critical values obtained using the simulation methods. Furthermore, we compare the power properties of the LR test with the specification test developed by Lee (1986). When the LR test is applied to the critical values accommodating small sample size distortion, it is very useful for the inference of Poisson mixtures.

Despite its appealing features of the LR test, it has the following drawbacks. First, the Poisson mixture model has to be specified by letting the space of λ_* be sufficiently large without knowing the level of unobserved heterogeneity. If the space is selected too small, so that the mixture model is misspecified under the alternative, the power of the LR test can be lost. Second, when the asymptotic critical values are obtained by the tail probability for an extremum, the level of significance has to be sufficiently small. Otherwise, the asymptotic critical values are conservative, making it hard to control the type-I error. We leave tackling these limitations as future research topics.

A Assumptions

The following regularity conditions are adapted from Cho and White (2007) by accommodating the stylized aspects of popular Poisson regression models.

A1: (i) An observed data set $\{(X_t, \mathbf{Z}_t')' \in \mathbb{N} \times \mathbb{R}^k\}$ ($k \in \mathbb{N}$), is a set of strictly stationary and geometric β–mixing random variables; and $\{\mathbf{Z}_i\}$ is time–invariant and does not contain a constant term.
(ii) The conditional X_t given $\mathbf{Z}_t$ is identically and independently distributed, and for some element(s) $(\pi_*, \lambda_{1*}, \lambda_{2*}, \boldsymbol{\beta}_*) \in [0,1] \times [\underline{\lambda}, \bar{\lambda}] \times [\underline{\lambda}, \bar{\lambda}] \times B$, its conditional distribution is identical to

$$\pi_* f(\lambda_{1*} g(\mathbf{Z}_t; \boldsymbol{\beta}_*); X_t) + (1-\pi_*) f(\lambda_{2*} g(\mathbf{Z}_t; \boldsymbol{\beta}_*); X_t)$$

where for $i = 1, 2$,

$$f(\lambda_{i*} g(\mathbf{Z}_t; \boldsymbol{\beta}_*); X_t = k) = \frac{\exp\{-\lambda_{i*} g(\mathbf{Z}_t; \boldsymbol{\beta}_*)\}\{\lambda_{i*} g(\mathbf{Z}_t; \boldsymbol{\beta}_*)\}^k}{k!},$$

and $[\underline{\lambda}, \bar{\lambda}] \times B$ is a compact and convex set in $\mathbb{R}^+ \times \mathbb{R}^d$ ($d \in \mathbb{N}$). Further, for each $\boldsymbol{\beta} \in B$, $g(\,\cdot\,; \boldsymbol{\beta})$ is a positively valued measurable function.

A2: (i) A null model for the conditional distribution of X_t given $\mathbf{Z}_t$ is specified as

$$\{f(\lambda g(\mathbf{Z}_t; \boldsymbol{\beta}); X_t) : (\lambda, \boldsymbol{\beta}) \in [\underline{\lambda}, \bar{\lambda}] \times B\}$$

such that $f(\lambda g(\mathbf{Z}_t; \,\cdot\,))$ is four–times continuously differentiable almost surely.
(ii) An alternative model for the conditional distribution of X_t given $\mathbf{Z}_t$ is specified as

$$\{\pi f(\lambda_1 g(\mathbf{Z}_t; \boldsymbol{\beta}); X_t) + (1-\pi) f(\lambda_2 g(\mathbf{Z}_t; \boldsymbol{\beta}); X_t) : (\pi, \lambda_1, \lambda_2, \boldsymbol{\beta}) \in [0,1] \times [\underline{\lambda}, \bar{\lambda}] \times [\underline{\lambda}, \bar{\lambda}] \times B\},$$

and for each $(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta})$, $\mathbb{E}[\ell_t(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta})]$ exists and is finite, where

$$\ell_t(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta}) := \log[\pi f(\lambda_1 g(\mathbf{Z}_t; \boldsymbol{\beta}); X_t) + (1-\pi) f(\lambda_2 g(\mathbf{Z}_t; \boldsymbol{\beta}); X_t)].$$

A3: There exists a sequence of positive, strictly stationary, and ergodic random variables, $\{M_t\}$, such that for some $\delta > 0$,

1. $\mathbb{E}[M_t^{1+\delta}] < \Delta < \infty$;
2. $\sup_{(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta})} |\nabla_{j_1} \ell_t(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta}) \nabla_{j_2} \ell_t(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta})| \leq M_t$;
3. $\sup_{(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta})} |\nabla_{j_1, j_2} \ell_t(\pi, \lambda_1, \lambda_2, \boldsymbol{\beta})| \leq M_t$;
4. $\sup_{(\lambda, \boldsymbol{\beta}, \gamma)} |\nabla_{i_1} f(\lambda g(\mathbf{Z}_t; \boldsymbol{\beta}); X_t) / f(\lambda g(\mathbf{Z}_t; \boldsymbol{\beta}); X_t)|^4 \leq M_t$;

5. $\sup_{(\lambda,\boldsymbol{\beta},\gamma)} |\nabla_{i_1}\nabla_{i_2} f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}); X_t)/f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}); X_t)|^2 \leq M_t$;

6. $\sup_{(\lambda,\boldsymbol{\beta},\gamma)} |\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}); X_t)/f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}); X_t)|^2 \leq M_t$;

7. $\sup_{(\lambda,\boldsymbol{\beta},\gamma)} |\nabla_{i_1}\nabla_{i_2}\nabla_{i_3}\nabla_{i_4} f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}); X_t)/f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}); X_t)| \leq M_t$,

where $j_1, j_2 \in \{\pi, \lambda_1, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_d\}$, and $i_1, \cdots, i_4 \in \{\lambda, \boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_d\}$.

For each λ, λ', denote the matrices

$$B(\lambda,\lambda') := \begin{bmatrix} \mathbb{E}[r_t(\lambda)r_t(\lambda')] & \mathbb{E}[r_t(\lambda')s_t'] \\ \mathbb{E}[r_t(\lambda)s_t] & \mathbb{E}[s_t s_t'] \end{bmatrix}, \quad C := \begin{bmatrix} \mathbb{E}[t_t^2] & \mathbb{E}[t_t s_t'] \\ \mathbb{E}[t_t s_t] & \mathbb{E}[s_t s_t'] \end{bmatrix}$$

and let $\lambda_{\min}$ and $\lambda_{\max}$ be the minimum and the maximum eigenvalues of a given matrix, where for each λ,

$$\begin{aligned} r_t(\lambda) &:= 1 - f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}_*); X_t)/f(\lambda_* g(\mathbf{Z}_t;\boldsymbol{\beta}_*); X_t), \\ s_t &:= \nabla_{(\lambda,\boldsymbol{\beta})} f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}); X_t)/f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}); X_t)|_{(\lambda_*,\boldsymbol{\beta}_*)}, \\ t_t &:= \nabla_\lambda^2 f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}_*); X_t)/f(\lambda g(\mathbf{Z}_t;\boldsymbol{\beta}_*); X_t)|_{\lambda=\lambda_*}, \end{aligned}$$

and λ_* is an unique element in $(\underline{\lambda}, \bar{\lambda})$ given by the hypothesis.

A4: (*i*) For each $(\lambda,\lambda') \neq (\lambda_*,\lambda_*)$, $\lambda_{\min}\{B(\lambda,\lambda')\} > 0$ and $\lambda_{\max}\{B(\lambda,\lambda')\} < \infty$.
(*ii*) $\lambda_{\max}(C) < \infty$ and $\lambda_{\min}(C) > 0$.

References

AMENGUAL, D., X. BEI, M. CARRASCO, AND E. SENTANA (2025): "Score-Type Tests for Normal Mixtures," *Journal of Econometrics*, 248, 105717. 4

CHEN, H. AND J. CHEN (2001): "The Likelihood Ratio Test for Homogeneity in the Finite Mixture Models," *Canadian Journal of Statistics*, 29, 201–216. 2, 3, 4

CHERNOFF, H. AND E. LANDER (1995): "Asymptotic Distribution of the Likelihood Ratio Test that a Mixture of Two Binomials is a Single Binomial," *Journal of Statistical Planning and Inference*, 43, 19–40. 1, 2, 3

CHO, J. S. (2006): "Testing Hypotheses by Mixture of Null and Alternative Models," Working paper. 5

——— (2025): "Practical Testing for Normal Mixtures," Working Paper 2025rwp-248, Yonsei Economics Research Institute. 4, 5

CHO, J. S. AND C. HAN (2009): “Testing for the Mixture Hypothesis of Geometric Distributions,” *Journal of Economic Theory and Econometrics*, 20:3, 31–55. 4, 5

CHO, J. S., J. S. PARK, AND S. W. PARK (2018): “Testing for the Mixture Hypothesis of Conditional Geometric and Exponential Distributions,” *Journal of Economic Theory and Econometrics*, 29:2, 1–27. 4

CHO, J. S. AND H. L. WHITE (2007): “Testing for Regime–Switching,” *Econometrica*, 75, 1671–1720. 1, 2, 3, 4, 5, 9

——— (2010): “Testing for Unobserved Heterogeneity in Weibull and Exponential Duration Models,” *Journal of Econometrics*, 157, 458–480. 2, 4, 5

DACUNHA-CASTELLE, D. AND E. GASSIAT (1999): “Testing the Order of a Model Using Locally Conic Parametrization: Population Mixtures and Stationary ARMA Processes,” *Annals of Statistics*, 27, 1178–1209. 1, 2, 3

DAVIES, R. (1977): “Hypothesis Testing When a Nuisance Parameter is Present only under the Alternative,” *Biometrika*, 64, 247–254. 3

——— (1987): “Hypothesis Testing When a Nuisance Parameter is Present only under the Alternative,” *Biometrika*, 74, 33–43. 3

GHOSH, J. AND P. SEN (1985): “On the Asymptotic Performance of the Log Likelihood Ratio Statistic for the Mixture Model and Related Results,” in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, ed. by L. L. Cam and R. Olshen, Berkeley: University of California Press, vol. 2, 789–806. 1, 3

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): “Pseudo Maximum Likelihood Methods: Applications to Poisson Models,” *Econometrica*, 52, 701–720. 1, 2

HARTIGAN, J. (1985): “Failure of Log–Likelihood Ratio Test,” in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, ed. by L. M. L. Cam and R. A. Olshen, Berkeley: University of California Press, vol. 2, 807–810. 1, 4

HAUSMAN, J., B. HALL, AND Z. GRILICHES (1984): “Econometric Models for Count Data with an Application to the Patents–R&D Relationship,” *Econometrica*, 52, 909–938. 1, 2, 3

HECKMAN, J. AND B. SINGER (1984): “A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data,” *Econometrica*, 52, 271–320. 2

KARLIS, D. AND E. XEKALAKI (1999): "On Testing for the Number of Components in a Mixed Poisson Model," *Annals of the Institute of Statistical Mathematics*, 51, 149–162. 1, 4

——— (2001): "Robust Inference for Finite Poisson Mixtures," *Journal of Statistical Planning and Inference*, 93, 93–115. 1, 5

LEE, L.-F. (1986): "Specification Test for Poisson Regression Models," *International Economic Review*, 27, 689–706. , 1, 8, 9

PITERBARG, V. (1996): in *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, Providence: American Mathematical Society, vol. 148 of *Translations of Mathematical Monographs*. 6

SCHLATTMANN, P. (2003): "Estimating the Number of Components in a Finite Mixture Model: The Special Case of Homogeneity," *Computational Statistics and Data Analysis*, 41, 441–451. 1, 5

VAN DEN BERG, G. AND G. RIDDER (1998): "An Empirical Equilibrium Search Model of the Labor Market," *Econometrica*, 66, 1183–1221.

WOOLDRIDGE, J. (1999): "Distribution-Free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics*, 90, 77–97. 2

Nominal Level \ Sample Size	50	100	200	∞
1.00 %	6.7449	6.6231	6.5874	6.4012
	(0.4524)	(0.4878)	(0.3764)	(0.3374)
5.00 %	3.7040	3.6930	3.6804	3.6750
	(0.0784)	(0.0841)	(0.0866)	(0.0806)
10.0 %	2.4412	2.4567	2.4615	2.4581
	(0.0391)	(0.0400)	(0.0366)	(0.0397)
15.0 %	1.7597	1.7614	1.7834	1.7588
	(0.0248)	(0.0237)	(0.0260)	(0.0256)

Table 1: CRITICAL VALUES OF THE LR TEST. Figures show the critical values obtained by repeating (7) or (9) independently, and figures in parentheses denote the standard errors. Number of Replications: 50,000. Data Generating Process: $X_t|Z_t \sim$ IID Pois$(2\exp(Z_t))$ and $Z_t \sim$ IID $U(-1,1)$. Model: $X_t|Z_t \sim \pi$Pois$(\lambda_1 \exp(\beta Z_t)) + (1-\pi)$Pois$(\lambda_2 \exp(\beta Z_t))$ and $\lambda_1, \lambda_2 \in [1,3]$.

Nominal Level \ Sample Size	100	300	500	700	1,000
1.00	0.47	0.76	0.83	0.85	0.85
5.00	2.79	3.67	3.69	4.05	4.05
10.0	6.09	7.67	8.30	8.84	8.84
15.0	9.77	11.88	12.83	13.73	13.73

Table 2: EMPIRICAL REJECTION RATES OF THE LR TEST UNDER THE NULL (IN PERCENT). Figures show the empirical rejection rates under the null hypothesis, which are obtained by repeating independent experiments. Number of Replications: 10,000. Data Generating Process: $X_t|Z_t \sim$ IID Pois$(2\exp(Z_t))$ and $Z_t \sim$ IID $U(-1,1)$. Model: $X_t|Z_t \sim \pi$Pois$(\lambda_1 \exp(\beta Z_t)) + (1-\pi)$Pois$(\lambda_2 \exp(\beta Z_t))$ and $\lambda_1, \lambda_2 \in [1,3]$.

	λ_{1*}	1.80	1.60	1.40	1.20
	λ_{2*}	2.20	2.40	2.60	2.80
	$\mathscr{LR}'$	3.56	4.56	7.53	16.63
$\pi_* = 0.1$	$\mathscr{LR}''$	6.43	7.73	13.23	24.66
	$\mathscr{SR}$	4.40	4.70	7.16	11.11
	$\mathscr{LR}'$	4.26	9.33	25.03	55.63
$\pi_* = 0.3$	$\mathscr{LR}''$	7.43	14.60	34.36	65.36
	$\mathscr{SR}$	4.83	8.63	19.73	40.70
	$\mathscr{LR}'$	3.73	11.63	35.60	72.70
$\pi_* = 0.5$	$\mathscr{LR}''$	6.56	16.80	45.43	80.13
	$\mathscr{SR}$	5.43	9.93	27.40	61.11
	$\mathscr{LR}'$	4.13	10.90	32.46	67.70
$\pi_* = 0.7$	$\mathscr{LR}''$	7.53	16.46	41.83	75.06
	$\mathscr{SR}$	5.46	8.03	27.43	61.36
	$\mathscr{LR}'$	3.80	6.23	14.16	30.20
$\pi_* = 0.9$	$\mathscr{LR}''$	7.40	10.43	20.70	37.80
	$\mathscr{SR}$	4.86	7.00	12.70	26.93

Table 3: POWER OF THE TESTS (IN PERCENT, 5% NOMINAL LEVEL). Figures show the empirical rejection rates of $\mathscr{LR}'$, $\mathscr{LR}''$, and $\mathscr{SR}$ under the alternative hypothesis, which are obtained by repeating independent experiments. Number of Replications: 3,000. Data Generating Process: $Z_t \sim$ IID $U(-1,1)$ and $X_t|Z_t \sim$ IID π_*Pois$(\lambda_{1*}\exp(Z_t)) + (1-\pi_*)$Pois$(\lambda_{2*}\exp(Z_t))$. Model: $X_t|Z_t \sim \pi$Pois$(\lambda_1 \exp(\beta Z_t)) + (1-\pi)$Pois$(\lambda_2 \exp(\beta Z_t))$ and $\lambda_1, \lambda_2 \in [1,3]$.