



# A Novel Correction for the Adjusted Box-Pierce Test

Sidy Danioko, Jianwei Zheng\*, Kyle Anderson, Alexander Barrett and Cyril S. Rakovski

Schmid College of Science and Technology, Chapman University, Orange, CA, United States

The classical Box-Pierce and Ljung-Box tests for auto-correlation of residuals possess severe deviations from nominal type I error rates. Previous studies have attempted to address this issue by either revising existing tests or designing new techniques. The Adjusted Box-Pierce achieves the best results with respect to attaining type I error rates closer to nominal values. This research paper proposes a further correction to the adjusted Box-Pierce test that possesses near perfect type I error rates. The approach is based on an inflation of the rejection region for all sample sizes and lags calculated via a linear model applied to simulated data that encompasses a large range of data scenarios. Our results show that the new approach possesses the best type I error rates of all goodness-of-fit time series statistics.

## OPEN ACCESS

### Edited by:

Avner Bar-Hen,  
Conservatoire National des Arts et  
Métiers (CNAM), France

### Reviewed by:

Hossein Hassani,  
University of Tehran, Iran  
Christian Derquenne,  
Electricité de France, France

### \*Correspondence:

Jianwei Zheng  
zheng120@mail.chapman.edu

### Specialty section:

This article was submitted to  
Statistics and Probability,  
a section of the journal  
Frontiers in Applied Mathematics and  
Statistics

Received: 11 February 2022

Accepted: 12 April 2022

Published: 19 May 2022

### Citation:

Danioko S, Zheng J, Anderson K,  
Barrett A and Rakovski CS (2022) A  
Novel Correction for the Adjusted  
Box-Pierce Test.  
Front. Appl. Math. Stat. 8:873746.  
doi: 10.3389/fams.2022.873746

**Keywords:** model selection, residuals, auto-correlation, type I error, diagnostic test, portmanteau Q statistic

## 1. INTRODUCTION

The Box-Jenkins algorithm is a general systematic approach for model checking of a time series model. Examples of the approach can be found in [1–3]. A well-fitting model produces residuals that are free of correlation. Thus, standard goodness-of-fit approaches are in essence global tests for absence of correlation among estimated residuals. Accordingly, many statistical techniques have been designed to assess the absence of correlation among the time series model residuals.

Following classical notation, let  $\{X_t\}$  be an observed time series generated by a stationary and invertible ARMA( $p, q$ ) process  $\phi(B)X_t = \theta(B)\epsilon_t$ , where  $\phi(B)$  and  $\theta(B)$  are the autoregressive and moving average characteristic polynomial and  $B^k X_t = X_{t-k}$  is the backshift operator. The desired parameters,  $\phi_i$  and  $\theta_i$  are estimated using maximum likelihood or least squares methods to obtain  $\hat{\phi}_i$  and  $\hat{\theta}_i$ , the residuals are calculated via  $\hat{\epsilon}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)X_t$  and the sample auto-correlation coefficients are in turn obtained from  $\hat{r}_k = \sum_{t=k+1}^n \hat{\epsilon}_t \hat{\epsilon}_{t-k} / \sum_{t=1}^n \hat{\epsilon}_t^2$ .

In recent years, many techniques have been employed to test the global hypothesis of all autocorrelations up to a certain lag,  $H_0 : r_1 = r_2 = \dots = r_m = 0$ . In general, these techniques are designed as weighted sums of squares of the estimated autocorrelations and they can produce misleading conclusions due to deviations from the asymptotic limiting distribution in moderate size samples [4–6]. Thus, a new and more robust test is proposed in this research that attains precise type I error rates for all sample sizes.

The history of portmanteau tests traces its roots back to the Box-Pierce diagnostic test defined as [6, 7] :

$$Q_{BP} = n \sum_{k=1}^m \hat{r}_k^2, \quad (1)$$

where  $n$ ,  $m$ , and  $\hat{r}_k$  represent the sample size, number of lags being tested and the sample auto-correlation of order  $k$  of the residuals, respectively. The authors showed that the asymptotic distribution of  $Q_{BP}$  is approximately  $\chi^2(m-p-q)$  but considerable deviations for moderate sample sizes have been observed [7–9]. That deficiency entails imperfections of type I error rates and prompted the design of a weighted and improved versions of the test. In their stimulation studies, Ray and Xiaolou [4] focused on investigating the type I errors in the  $\chi_m^2$  setting. They remarked that the Box-Pierce test has imperfect type I error rates for most sample size and lag values.

Ljung and Box [7] were the first ones to propose a design that assigns larger weights to residuals estimated with more data:

$$Q_{LB} = n(n + 2) \sum_{k=1}^m \frac{\hat{r}_k^2}{n - k} = n \sum_{k=1}^m \frac{n + 2}{n - k} \hat{r}_k^2. \tag{2}$$

The Box-Pierce and Ljung-Box tests are asymptotically equivalent. The Ljung-Box test has been shown to overcorrect in moderate samples [4]. They also realized that Ljung-Box inflates the test statistic using a variance estimate of the residuals. They further showed that on moderate sized data,  $Q_{LB}$  rejects too often because the test statistic is too small.

Li and McLeod [9] refined the  $Q_{BP}$  test by proposing the following statistic,

$$Q_{LB} = Q_{BP} + \frac{m(m + 1)}{2n} = \frac{m(m + 1)}{2n} + n \sum_{k=1}^m \hat{r}_k, \tag{3}$$

This approach only corrects the mean of the Box-Pierce statistic and consequently fails to properly adjust the type I error rates.

Monti [10] proposed a portmanteau test based on the residual partial autocorrelations. The test is defined as,

$$Q_M = n(n + 2) \sum_{k=1}^m \frac{\hat{\pi}_k^2}{n - k}, \tag{4}$$

Monti [10] showed *via* simulations that the performance of  $Q_M$  is comparable to that  $Q_{LB}$ . In addition, he concluded that in certain scenarios,  $Q_{LB}$  outperforms  $Q_M$ .

Peña and Rodríguez [11] proposed a test based on a different measure of dependence of the residual autocorrelations,

$$D = n(1 - |\hat{R}_m|^{1/m}), \tag{5}$$

where

$$\tilde{R}_m = \begin{pmatrix} 1 & \hat{r}_1 & \dots & \hat{r}_m \\ \hat{r}_1 & 1 & \dots & \hat{r}_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_m & \hat{r}_{m-1} & \dots & 1 \end{pmatrix} \tag{6}$$

In their work, the authors showed that under particular conditions, their test greatly outperformed  $Q_{LB}$  test. Furthermore, they demonstrated that the test had an advantage

over the McLeod and Li’s test regardless of sample size. However, the convergence of the asymptotic distribution of the test developed by Peña and Rodríguez is very slow [12].

Fisher proposed new weighted versions of the Box-Pierce and Monti’s tests, the Q statistic [5]:

$$\tilde{Q}_{WL} = n(n + 2) \sum_{k=1}^m \frac{m - k + 1}{m(n - k)} \hat{r}_k^2, \tag{7}$$

and

$$\tilde{Q}_{WM} = n(n + 2) \sum_{k=1}^m \frac{m - k + 1}{m(n - k)} \hat{\pi}_k^2, \tag{8}$$

A comparison simulation study by Safi and Al-Reqep [13] showed that for small sample size and  $m$  values  $Q_{WL}$  performs better than  $Q_{LB}$ . For moderate sample sized data, they also found that  $Q_{WL}$  does better than  $Q_{LB}$  and  $Q_{WM}$  outperforms  $Q_M$ .

To remedy some of the shortcomings of all previously existing tests, Kan and Wang [4] proposed a new modification of the portmanteau test, widely called the adjusted Box-Pierce test. They defined their statistic as,

$$Q_{BP}^a = m + \sqrt{\frac{2m}{Var[Q_{BP}]}} (Q_{BP} - E[Q_{BP}]), \tag{9}$$

The authors conducted an evaluation of various tests including Box-Pierce and Ljung-Box. The design of the adjusted Box-Pierce statistic (9) explicitly recenters and rescales  $Q_{BP}$  to attain the mean and variance of a  $\chi^2(m)$  variable. The authors showed through simulations that the test possesses very good adherence to nominal type I error rates. In their comparison study, they found that both the distributions of  $Q_{BP}$  and  $Q_{LB}$  deviate from the expected variance of  $\chi^2(m)$  distribution for small and moderate sample sizes and almost all choices for the value of  $m$ .

All of the above-mentioned tests exhibit deviations from the nominal type I error rates that compromise their performance. Hassani and Yeganegi [14, 15] conducted simulation studies to evaluate the optimal lag value for the Ljung-Box test. They found that the optimal number of lags not only depends on the length of the time series, but also on the significance level of the test. Thus, a new approach is proposed which aims at correcting the rejection region instead of redesigning the test statistic itself. This technique was introduced by Bernard in his effort to construct a more powerful alternative to Fisher’s exact test [16, 17] and later by Boschloo [18]. The same idea of rejection region correction has been recently employed by Ehwerhemuepha et al. [19] to produce the best performing test for homogeneity for multinational distributions.

## 2. METHODS

A model based correction of the rejection region of the adjusted Box-Pierce test was designed. A large scale simulation study was then conducted to not only estimate the correction, but to also assess the performance advantages (defined as adherence to the

nominal type I error rates for all scenarios) of the proposed corrected method.

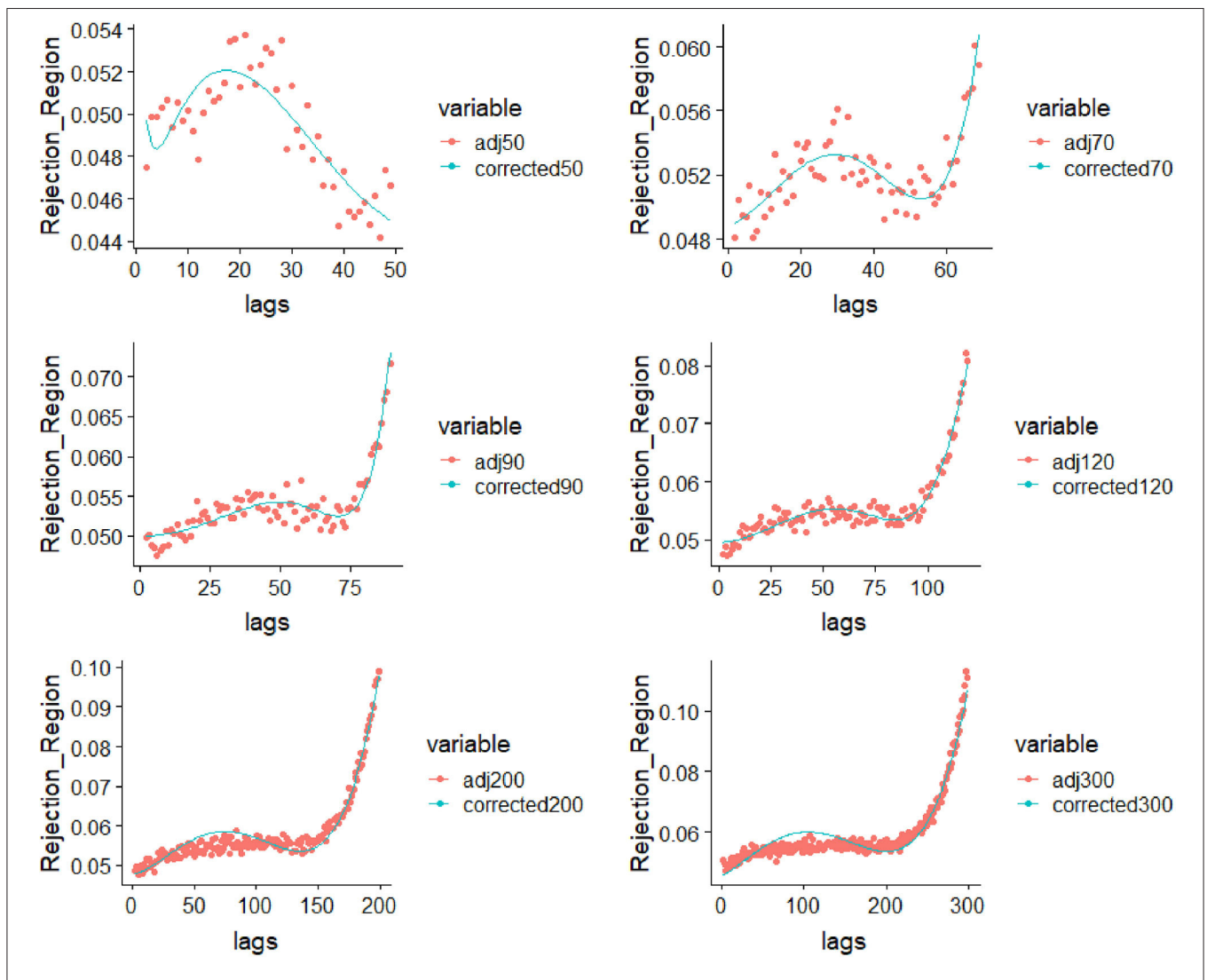
$n$  were stacked to get an aggregated dataset with number of rows  $\sum_{n=4}^{30} 10n(10n - 2) = 934,920$ .

### 2.1. Simulation Study

For sample size values of  $n = 40, 50, \dots, 300$ , we simulated  $10^6$  white noise samples,  $s_{n1}, s_{n2}, \dots, s_{n10^6} \sim N_n(0, I)$ . These mimic the behavior of residuals of a well-fitting time series model (under the null). Next, the adjusted Box-Pierce test was applied to every sample and for all possible lags,  $m$  ( $2 \leq m \leq n - 1$ ) and the corresponding  $p$ -values,  $p_{nm1}, p_{nm2}, \dots, p_{nm10^6}$  were obtained. For each pair  $(n, m)$ , the estimated type I error rate of the adjusted Box-Pierce test at alpha level of 0.05 was empirically estimated by  $P_{\alpha=0.05}^{n,m} = \sum_{i=1}^{10^6} I\{p_{nmi} < 0.05\} / 10^6$ . Thus, for each sample size  $n$ ,  $n - 2$  empirically estimated type I error rates yielding a dataset with three columns,  $n$ ,  $m$ , and  $P_{\alpha=0.05}^{n,m}$ . Further, these datasets obtained from all individual sample sizes

### 2.2. Linear Model

The primary idea of this study was to provide a model-based correction to the rejection region of the adjusted Box-Pierce test in order to attain improved type I error rates for all sample sizes and lags. We created six linear regression models trained on the simulated data described in the section above. These six models were trained on different subsets of the data split into sample size intervals  $[0, 50]$ ,  $[51, 70]$ ,  $[71, 90]$ ,  $[91, 120]$ ,  $[121, 200]$ , and  $[201, 300]$ . The difference in the type I error rate patterns for distinct sample sizes (shown in **Figure 1**) necessitated the use of separate models to achieve the desired level of fit. These linear models are complex as they encompass different powers of  $n$ ,  $m$ , and



**FIGURE 1** | Parametric correction to the rejection region for sample sizes 50, 70, 90, 130, 200, and 300.

their 2-way interactions. The general formula adopted for the models was,

$$Y - 0.05 = \alpha_1 n^s + \alpha_2 m^p + \alpha_3(n^s * m^p) + \alpha_4(n^{2s} * m^{2p}) + \alpha_5 n^{2s} + \alpha_6(n^{3s} * m^{2p}) + \alpha_7(n^{3s} * m^{3p}) + \alpha_8 m^{4p} + \alpha_9 m^{5p}. \tag{10}$$

Further, within the general form (10) an extensive grid search to find the best values of the power transformation parameters  $s$  and  $p$  was performed. The type I error rates from the selected best models are presented in **Table 1**. The rates were calculated using validation data with sample sizes of  $n_{val} = 45, 65, 85, 100, 250$ .

**TABLE 1** | Performance summary of the correction to the Adjusted Box-Pierce.

Sample size	s	p	AdjBoxPierce	Corrected version
$n = 45$	0.2	0.3	0.04868907	0.05001953
$n = 65$	10.0	1.0	0.05163921	0.05002905
$n = 85$	7.0	2.0	0.05305157	0.05045904
$n = 100$	1.3	1.7	0.05447408	0.05020469
$n = 160$	0.8	0.9	0.05629981	0.04987525
$n = 250$	1.9	0.8	0.05813593	0.05037286

**TABLE 2** | Summary statistics for selected variables in interval sample size <50.

Variable	Estimate	Std.Error	t-value	p-value
$n^s$	0.425295	0.251604	1.690	0.095008
$m^p$	-1.353900	0.793110	-1.707	0.091837
$n^s * m^p$	0.593460	0.396921	1.495	0.138960
$n^{2s} * m^{2p}$	0.149028	0.056476	2.639	0.010065*
$n^{2s}$	-0.183531	0.122355	-1.500	0.137706
$n^{3s} * m^{2p}$	-0.070355	0.030893	-2.277	0.025539*
$n^{3s} * m^{3p}$	0.004419	0.002064	2.141	0.035436*
$m^{4p}$	-0.017762	0.004355	-4.079	0.000109***
$m^{5p}$	0.002106	0.000461	4.570	1.83e-05***

The symbols \*, \*\* designate the statistical significance level of the variables in a given model.

**TABLE 3** | Summary statistics for selected variables in finite sample size between 51 and 70.

Variable	Estimate	Std.Error	t-value	p-value
$n^s$	-2.652e-06	8.296e-07	-3.196	0.00179 **
$m^p$	1.209e-03	2.984e-04	4.053	9.12e-05 ***
$n^s * m^p$	-2.283e-07	7.347e-08	-3.108	0.00237 **
$n^{2s} * m^{2p}$	-2.068e-12	3.852e-13	-5.369	4.07e-07 ***
$n^{2s}$	4.910e-10	1.869e-10	2.627	0.00977 **
$n^{3s} * m^{2p}$	4.637e-16	8.877e-17	5.223	7.75e-07 ***
$n^{3s} * m^{3p}$	-1.167e-18	2.439e-19	-4.784	5.05e-06 ***
$m^{4p}$	6.138e-10	2.856e-10	2.150	0.03364 *
$m^{5p}$	2.552e-12	1.811e-12	1.409	0.16150

The symbols \*, \*\*, \*\*\* designate the statistical significance level of the variables in a given model.

### 3. RESULTS

Noticeable differences between the patterns of type I error rates across the analyzed sample sizes (40–300) were discovered. Therefore, sample-size specific models (0–50, 51–70, 71–90, 91–120, 120–200, 201–300) were constructed to capture the exact pattern for that particular scenario. **Table 1** displays a condensed form of the comparative study between revised

**TABLE 4** | Summary statistics for selected variables in finite sample size between 71 and 90.

Variable	Estimate	Std.Error	t-value	p-value
$n^s$	3.214e-17	2.901e-17	1.108	0.269585
$m^p$	3.833e-06	1.130e-06	3.392	0.000877 ***
$n^s * m^p$	-1.392e-20	3.309e-20	-0.421	0.674609
$n^{2s} * m^{2p}$	-4.627e-36	6.406e-37	-7.224	2.02e-11 ***
$n^{2s}$	-6.756e-31	6.616e-31	-1.021	0.308740
$n^{3s} * m^{2p}$	9.423e-50	1.523e-50	6.189	5.00e-09 ***
$n^{3s} * m^{3p}$	-1.759e-54	4.077e-55	-4.315	2.80e-05 ***
$m^{4p}$	2.816e-17	2.774e-18	10.153	< 2e-16 ***

The symbol \*\*\* designates the statistical significance level of the variables in a given model.

**TABLE 5** | Summary statistics for selected variables in finite sample size between 91 and 120.

Variable	Estimate	Std.Error	t-value	p-value
$n^s$	5.169e-06	3.434e-06	1.505	0.133211
$m^p$	1.266e-05	3.809e-06	3.323	0.000994***
$n^s * m^p$	-1.569e-09	9.362e-09	-0.168	0.867045
$n^{2s} * m^{2p}$	-2.021e-13	1.482e-14	-13.641	< 2e-16***
$n^{2s}$	-1.216e-08	7.488e-09	-1.624	0.105408
$n^{3s} * m^{2p}$	3.782e-16	3.539e-17	10.687	< 2e-16***
$n^{3s} * m^{3p}$	-4.778e-20	4.874e-21	-9.804	< 2e-16***
$m^{4p}$	3.367e-15	1.792e-16	18.793	< 2e-16***
$m^{5p}$	-4.058e-19	3.561e-20	-11.397	< 2e-16***

The symbol \*\*\* designates the statistical significance level of the variables in a given model.

**TABLE 6** | Summary statistics for selected variables in finite sample size between 121 and 200.

Variable	Estimate	Std.Error	t-value	p-value
$n^s$	5.966e-05	2.343e-05	2.546	0.01102*
$m^p$	8.195e-04	5.830e-05	14.056	< 2e-16***
$n^s * m^p$	-1.227e-05	1.336e-06	-9.181	< 2e-16***
$n^{2s} * m^{2p}$	-8.989e-09	3.701e-10	-24.290	< 2e-16***
$n^{2s}$	-1.271e-06	3.925e-07	-3.237	0.00124**
$n^{3s} * m^{2p}$	1.864e-10	5.775e-12	32.280	< 2e-16***
$n^{3s} * m^{3p}$	-1.079e-12	2.925e-14	-36.873	< 2e-16***
$m^{4p}$	1.233e-09	8.712e-11	14.147	< 2e-16***
$m^{5p}$	6.308e-12	6.042e-13	10.440	< 2e-16***

The symbols \*, \*\*, \*\*\* designate the statistical significance level of the variables in a given model.

version of Box-Pierce, which to the best of our knowledge is the last version, and the correction that we have brought into

the study. For different time series lengths, the corresponding *s*- and *p*-values along with the type I error rates for the adjusted Box-Pierce and those of the corrected version that we designed. It is important to realize that the results from the implementation of these models show that in all settings, the proposed regression-based correction provided almost perfect type I error rates. In particular, the adjusted type I error rates after the correction to the rejection regions were exactly 0.05 with detailed results.

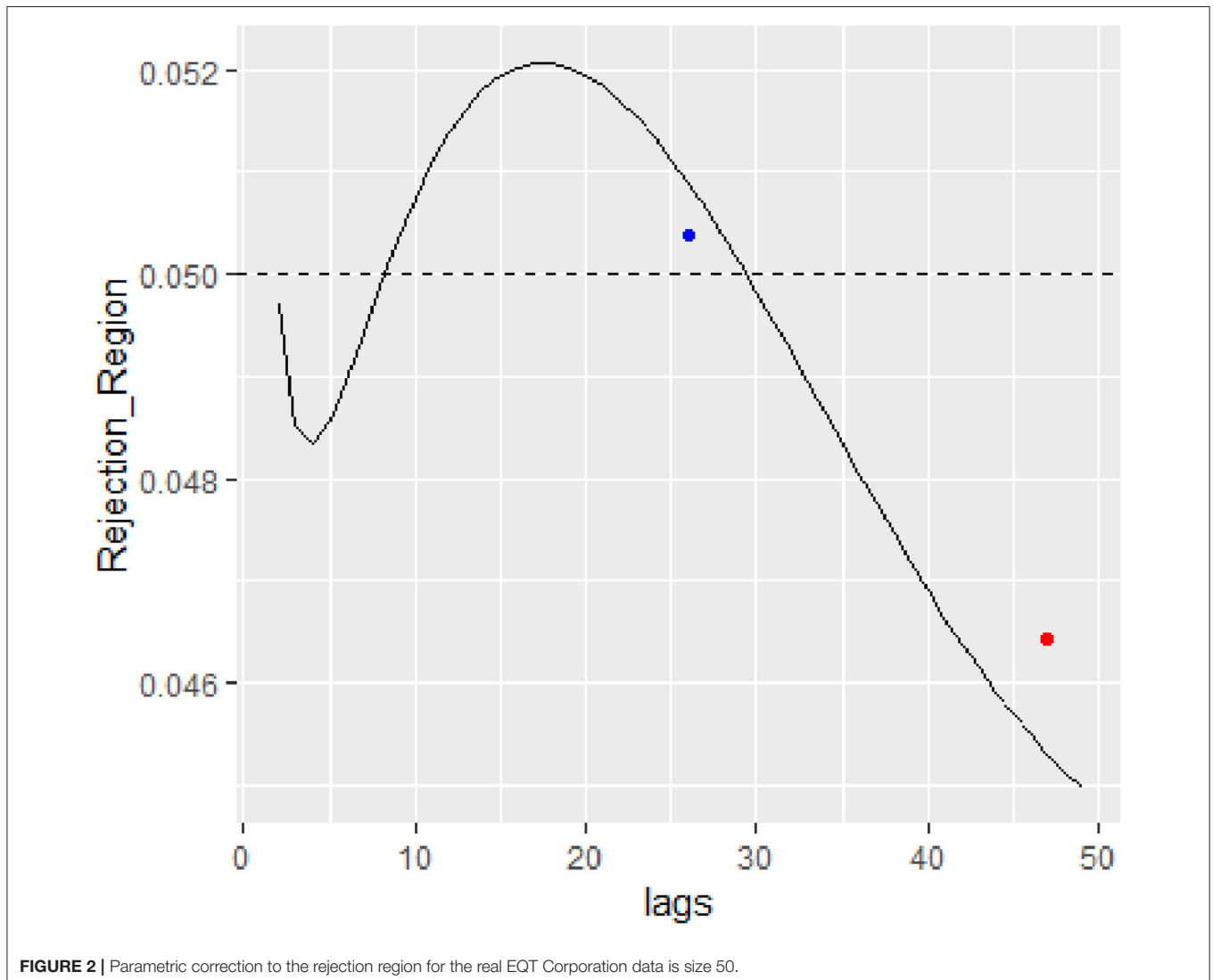
**TABLE 7** | Summary statistics for selected variables in finite sample size between 201 and 300.

Variable	Estimate	Std.Error	t-value	p-value
$n^s$	1.740e-07	5.213e-08	3.338	0.000868***
$m^p$	2.056e-04	5.313e-05	3.870	0.000114***
$n^s * m^p$	1.206e-08	5.327e-09	2.263	0.023777*
$n^{2s} * m^{2p}$	-9.680e-14	6.970e-15	-13.889	< 2e-16***
$n^{2s}$	-1.845e-11	2.884e-12	-6.396	2.22e-10***
$n^{3s} * m^{2p}$	5.841e-18	1.928e-19	30.295	< 2e-16***
$n^{3s} * m^{3p}$	-5.966e-20	2.469e-21	-24.161	< 2e-16***
$m^{4p}$	-4.111e-09	5.612e-10	-7.326	4.14e-13***
$m^{5p}$	1.660e-10	7.322e-12	22.678	< 2e-16***

The symbols \*, \*\*\*, designate the statistical significance level of the variables in a given model.

Tables 2–7 show detailed summary from the sample-size specific model fits. These models provide a parametric correction of the type I error rates. Graphical representation of results from the implementation of these models for several scenarios are shown in Figure 1.

Form left-to-right-up-to-down the fitting curves with appropriately found models in cases where ( $n = 50, 70, 90, 120, 300$ ) can be viewed. Empirically, it can be seen that the models that best fit the specific curve in a given data were found.



**FIGURE 2** | Parametric correction to the rejection region for the real EQT Corporation data is size 50.



## 4. DATA EXAMPLE

An application of our corrected version of the adjusted Box-Pierce test was performed using S&P 500 stock data. We provide instances of both false positive and false negative results obtained by the standard adjusted Box-Pierce test using EQT Corporation stock. This corporation created in 1884 and headquartered in Pittsburg is one of the leading companies extensively devoted to the exploration and transportation of hydrocarbon (Petroleum, natural gas, natural gas liquid). The average daily price of the EQT Corporation was calculated by collecting its opening and closing prices over a period over 8 years (2010–2018). For a window size of 50, numerous false negative and false positive points were found at different lags. In this case, instead of a critical value we have a critical boundary or curve exists. In this setting, the same rejection conditions are the same as in the normal case.

In **Figure 2**, instances of a false positive rejection at lag 26 are shown where the adjusted Box-Pierce test obtains a  $p$ -value of 0.0504 but the proposed model correction inflates the rejection region to start at 0.058. The graph also shows a false negative results with  $p$ -value of 0.046 at lag 47. However, the proposed correction shrinks the rejection region to start at 0.045.

## 5. DISCUSSION

In this work a new approach for correction of adjusted Box-Pierce test recently developed by Kan and Wang [4]. Conceptually, the rejection region correction idea is similar to the ones successfully employed in the work of [18, 19] to counterbalance the conservativeness of exact homogeneity tests. The provided method combines large scale simulations with subsequent scenario-specific regression modeling that includes complex interaction terms to achieve exceptionally good fit that entails nominal type I error rates for all sample sizes and lags used in the test statistic. The regression models that were constructed depend on the length of the series ( $n$ ) and the lag order ( $m$ ). The

exponents ( $s$ ) and ( $p$ ) of different variables present in the models are treated as hyperparameters in order to control the learning process. To obtain optimal values for those hyperparameters an extensive search through chosen subset values for ( $s$ ) and ( $p$ ) was conducted. The simulation study showed that the test outperforms all existing competing goodness-of-fit approaches for sample sizes up to 300.

It shall be noted that, in this study, we are not developing any new statistic but improving the best test among the current goodness-of-fit methods for time series. Our contribution is the introduction of a completely new idea to time series diagnostics, a rejection region correction *via* a range of parametric regressions fitted to large sample simulation data. Our study is an extension of the Adjusted Box-Pierce, as presented earlier.

The merit to the novel correction to the adjusted Box-Pierce proposed in this study is that it allows to find a test with vastly improved type I error rates for all sample size and lag values. This proposed technique of rejection region correction has direct implication on precise decision making by investors and financial institutions. The same technique can be easily extended to larger sample sizes.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

## AUTHOR CONTRIBUTIONS

SD conducted the simulation study, drafted, and revised the manuscript. JZ participated in interpreting the simulation results. KA reviewed and revised the manuscript, and prepared the final draft. AB contributed to the model building and the study design. CR conceived and designed the study, contributed to drafting, reviewing, and revising the manuscript. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Cryer JD, Chan K-S. *Time Series Analysis: With Applications in R, Vol. 2*. Springer (2008).
- Rachev ST, Mittnik S, Fabozzi FJ, Focardi SM, Jašić T. *Financial Econometrics: From Basics to Advanced Modeling Techniques*. Vol. 150. John Wiley & Sons (2007).
- Lindberg J. *Autoregressive Conditional Density*. (2016). doi: 10.2139/ssrn.2785499
- Kan R, Wang X. On the distribution of the sample autocorrelation coefficients. *J Econ*. (2010) 154:101–21. doi: 10.1016/j.jeconom.2009.06.010
- Fisher TJ. Testing adequacy of arma models using a weighted portmanteau test on the residual autocorrelations. In: *Contributed Paper to 2011 SAS Global Forum*. (2011) 327.
- Box GE, Pierce DA. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J Am Stat Assoc*. (1970) 65:1509–26. doi: 10.1080/01621459.1970.10481180
- Ljung GM, Box GE. On a measure of lack of fit in time series models. *Biometrika*. (1978) 65:297–303. doi: 10.1093/biomet/65.2.297
- Arranz MA. Tol-Project Portmanteau test statistics in time series. *Time Orientated Lang*. (2005) 1–8.
- Chen Y. On the robustness of Ljung-Box and McLeod-Li Q tests: a simulation study. *Econ Bull*. (2002) 3:1–10.
- Monti AC. A proposal for a residual autocorrelation test in linear models. *Biometrika*. (1994) 81:776–80. doi: 10.1093/biomet/81.4.776
- Peña D, Rodríguez J. A powerful portmanteau test of lack of fit for time series. *J Am Stat Assoc*. (2002) 97:601–10. doi: 10.1198/016214502760047122
- Lin JW, McLeod AI. Improved Peña-Rodríguez portmanteau test. *Comput Stat Data Anal*. (2006) 51:1731–8. doi: 10.1016/j.csda.2006.06.010
- Safi SK, Al-Reqep AA. Comparative study of portmanteau tests for the residuals autocorrelation in ARMA models. *Sci J Appl Math Stat*. (2014) 2:1–13. doi: 10.11648/j.sjams.20140201.11
- Hassani H, Yeganegi MR. Selecting optimal lag order in Ljung-Box test. *Phys A*. (2020) 541:123700. doi: 10.1016/j.physa.2019.123700
- Hassani H, Yeganegi MR. Sum of squared ACF and the Ljung-Box statistics. *Phys A*. (2019) 520:81–6. doi: 10.1016/j.physa.2018.12.028
- Barnard G. A new test for  $2 \times 2$  tables. *Nature*. (1945) 156:177. doi: 10.1038/156177a0
- Fisher RA. On the interpretation of  $\chi^2$  from contingency tables, and the calculation of P. *J R Stat Soc*. (1922) 85:87–94. doi: 10.2307/2340521
- Boschloo R. Raised conditional level of significance for the  $2 \times 2$ -table when testing the equality of two probabilities. *Stat Neerland*. (1970) 24:1–9. doi: 10.1111/j.1467-9574.1970.tb00104.x

19. Ehwerhemuepha L, Sok H, Rakovski C. A more powerful unconditional exact test of homogeneity for  $2 \times c$  contingency table analysis. *J Appl Stat.* (2019) 46:2572–82. doi: 10.1080/02664763.2019.1601689

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Danioko, Zheng, Anderson, Barrett and Rakovski. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*