A Monte Carlo Simulation Study on the Power of Autocorrelation Tests for ARMA Models

Zachary Wenning*, Emily Valenci

Department of Statistics, Pennsylvania State University, University Park, PA

https://doi.org/10.33697/ajur.2019.030

Students: zrw5043@psu.edu*, edv5042@psu.edu
Mentor: xxf13@psu.edu*

ABSTRACT

It is often the case when assessing the goodness of fit for an ARMA time series model that a portmanteau test of the residuals is conducted to assess residual serial correlation of the fitted ARMA model. Of the many portmanteau tests available for this purpose, one of the most famous and widely used is a variant of the original Box-Pierce test, the Ljung-Box test. Despite the popularity of this test, however, there are several other more modern portmanteau tests available to assess residual serial autocorrelation of the fitted ARMA model. These include two portmanteau tests proposed by Monti and Peña and Rodríguez. This paper focuses on the results of a power analysis comparing these three different portmanteau tests against different fits of ARMA - derived time series, as well as the behavior of the three different test-statistics examined when applied to a real-world data set. We confirm that for situations in which the moving average component of a fitted ARMA model is underestimated or when the sample size is small, the portmanteau test proposed by Monti is a viable alternative to the Ljung-Box test. We show new evidence that the Peña and Rodríguez test may also be a viable option for testing for residual autocorrelation in cases where the sample size is small.

KEYWORDS

Time Series; Monte Carlo; ARMA Models; Power; Simulation; Autocorrelation Tests; Portmanteau Tests; Monti; Ljung-Box; Peña and Rodríguez

INTRODUCTION

In many situations in statistics, data used for statistical modeling is required to obey the assumption of independence, as is often the case in variants of regression analysis and analysis of variance. However, there are endless examples of real-world scenarios where such an assumption is invalid and the data display autocorrelation with respect to time, such as in financial and agricultural data. This is often the case in time series, which are defined by data being collected periodically over time, and we may denote one as Y_t for $t = 1, ..., T$ where T is the total number of observations of the time series.[1] In the case where the data is quite simple in structure and seasonal variations are not present, it suffices to model autocorrelated data with an autoregressive-moving average model (denoted as ARMA(p, q) where p and q are the orders of the autoregressive and moving average components of the fitted ARMA model respectively) which can be written as:[2]

$$Y_t = \sum_{j=1}^{p} \phi_j Y_{t-j} + \sum_{k=1}^{q} \theta_k e_{t-k} + e_t \qquad \text{Equation 1.}$$

Where e_t denotes the errors of the fitted ARMA model and $\phi_1, ..., \phi_p$ and $\theta_1, ..., \theta_q$ are the corresponding parameters which are estimated using Conditional Least Squares, Method of Moments, or Maximum Likelihood methods.[1,2]

When fitting the above data, we require that 1) the error terms e_t be independent and identically distributed Gaussian white noise terms (e.g., normally distributed with mean 0 and variance σ^2), and 2) weak stationarity of the time series Y_t.[2] It is this first assumption which we will be assessing in our power analysis.

Note that, in theory, the goodness of fit for a particular ARMA model fit to a time series Y_t can be assessed by the validity of the first assumption. If the ARMA model is a sufficient fit for the time series, then it will take into account all serial autocorrelation present in the data, including autocorrelation of the data to its own past values and any current and past values of the stochastic term e_t.[1]

Much of this paper aims to confirm the findings of Safi and Al-Reqep's research on the power of portmanteau tests.[3] In this paper, we additionally seek to provide evidence that the Peña and Rodríguez test for residual autocorrelation is suitable for small sample sizes. We will now present three autocorrelation tests which assess this lack of fit.

METHODS AND PROCEDURES

Autocorrelation Tests for ARMA Models.

Autocorrelation is tested for using portmanteau tests as they are powerful for detecting deviations from independence. Each of the following portmanteau tests described in this paper test the lack of fit of an ARMA model by measuring the residual autocorrelation present after fitting the model to the time series data. Furthermore, the null and alternative hypothesis for each of the following tests can be written as follows:[4, 6, 7]

H_0 : The errors of the fitted ARMA model are independently distributed

H_A : The errors of the fitted ARMA model are not independent; they display serial autocorrelation

Arguably the most popular of the three different portmanteau tests examined for this power analysis is the Ljung-Box test. The test statistic for the Ljung-Box test is:[4]

$$Q = n(n+2) \sum_{k=1}^{h} \frac{r_k^2}{n-k}$$

Equation 2.

Here, n denotes the sample size, h is the number of lags being tested, and r_k is the residual autocorrelation at lag k. Ljung and Box showed that for residuals derived from a fitted ARMA(p, q) model, Q is asymptotically distributed as χ^2 with h - p - q degrees of freedom.[4]

A similar test was proposed by Monti, and is quite similar to the Ljung-Box test statistic, although the residual autocorrelation at lag k is replaced by the residual partial autocorrelation at lag k. The test statistic for the portmanteau test proposed by Monti can be denoted as:[6]

$$M = n(n+2) \sum_{k=1}^{h} \frac{\pi_k^2}{n-k}$$

Equation 3.

Here, n denotes the sample size, h is the number of lags being tested, and π_k is the residual partial autocorrelation at lag k. Monti showed that for residuals derived from a fitted ARMA(p, q) model, M is asymptotically distributed as χ^2 with h - p - q degrees of freedom.[6]

The last test being examined in this paper is one which was proposed in Peña and Rodríguez. The test statistic for this portmanteau test, which follows the same null and alternative hypotheses specified above, can be written as:[7]

$$D_m = n[1 - |\mathbf{R}_m|^{1/m}]$$ Equation 4.

Where n is the sample size. Here, $\mathbf{R}_m$ refers to the residual correlation matrix of dimension m, which can be written as:

$$\mathbf{R}_m = \begin{vmatrix} 1 & r_1 & \cdots & r_m \\ r_1 & 1 & \cdots & r_{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_m & r_{m-1} & \cdots & 1 \end{vmatrix}$$ Equation 5.

Where r_m is the residual autocorrelation at lag m. Peña and Rodríguez showed that for relatively large sample sizes, D_m is approximately distributed as Gamma with mean $(m+1)/2-(p+q)$ and variance $(2m+1)(m+1)/3m-2(p+q)$.[7]

Keeping these three different tests in mind, we will now briefly define power for a binary hypothesis test, and how an estimate for the power of each of the above tests can be derived via Monte Carlo simulation. Then, we will move on to the execution and results of the power analysis.

Power for a Binary Hypothesis Test.
Recall the hypotheses being tested by the above portmanteau tests (see *Introduction*), and let λ denote the power of the above test against the alternative hypothesis H_A. Then we can define the power of the above hypothesis test as:

$$\lambda = P(Reject\ H_0 | H_A\ is\ true)$$ Equation 6.

One problem that this paper seeks to address is the derivation of an estimate for λ for each of the above portmanteau tests. A reasonable method of achieving such an estimate is to model the probability of the portmanteau test rejecting the null hypothesis H_0 given the alternative hypothesis H_A is true by a Bernoulli distribution (and thus, the simulation of such trials as a Bernoulli process).

Hence, if one simulates n time series derived from an ARMA(p, q) process and models the simulated time series using an ARMA(p*, q*) model (such that the ARMA(p*, q*) model is an *underfit* of the ARMA(p, q) process), then the power of each of the above portmanteau tests can be estimated by calculating the proportion of times that the test correctly rejects the null hypothesis H_0 for the underfitted model.[1] Let X_i equal 1 if the ith application of the portmanteau test correctly classifies the ARMA(p*, q*) model as an underfit, and 0 otherwise. Then a reasonable estimate for λ would be:

$$\hat{\lambda} = \sum_{i=1}^{n} \frac{X_i}{n}$$ Equation 7.

Hence, the sampled X_i are independently and identically distributed random variables such that:

$$X_i = \begin{cases} 1 & w/\ probability\ \lambda \\ 0 & w/\ probability\ 1-\lambda \end{cases}$$ Equation 8.

Note that the above estimation is valid only for modeling the derived ARMA(p, q) process with an underfitting model. Therefore, we must know that the alternative hypothesis is true, else we would not obtain a valid estimate of the portmanteau test's power. This is because a model which is a proper fit (or even a model which suffers from overfitting) for the derived ARMA(p, q) process would theoretically be able to sufficiently model all serial autocorrelation present in the data, causing the p-value of the corresponding portmanteau test to inflate and thereby making it less likely that we would reject H_0.[1] Therefore, this condition is required to be satisfied if we are to obtain an accurate estimate of λ.

Execution of Power Analysis in R.

The execution of the power analysis in R consisted of using a custom built R function which allowed for the simulation of an ARMA(p, q) process with specified length and model fit.

The methods employed in the power analysis were as follows. The Ljung-Box, Monti, and Peña and Rodríguez portmanteau tests were conducted on either White Noise, AR(1), or MA(1) fits. These models were fitted against an ARMA(p, q) derived process with specified length and parameters which were generated randomly according to a uniform distribution $U(-1, 1)$ for each combination of N and k. These are shown in the corresponding tables below in *Results*. For this simulation study, only AR(2), MA(2) and ARMA(1, 1) derived processes were considered. Sample sizes N = 30, 100, 300 were used in this simulation as the goal was to investigate small to moderate sample sizes. Number of lags k = 5, 10, 15 were investigated as Safi and Al-Reqep showed that high number of lags resulted in a decrease in power.[3] Furthermore, 10,000 replications were used for each simulated process in order to achieve an estimate for the power of the corresponding portmanteau test accurate to four significant digits and also for the purpose of improving upon the accuracy from Safi and Al-Reqep's study.[3] Such power estimates were calculated using the theoretical approach outlined above in *Methods and Procedures* and generally align with the methods from Safi and Al-Reqep's simulation study.[3] All models were fitted to the corresponding derived ARMA(p, q) process using a Conditional Sum of Squares - Maximum Likelihood approach. This method is very similar to the power analysis conducted by Fisher, Monti, and Peña and Rodríguez.[1, 6, 7]

RESULTS

Power Analysis.

Below are the results of the power analysis for the derived AR(2), MA(2), and ARMA(1, 1) processes. Note that in Tables 1, 2, and 3 (displayed below), N denotes the length of the derived process, k denotes the number of lags used to calculate the test statistics of the three portmanteau tests, ϕ_1, ϕ_2, θ_1, and θ_2 denote the parameters of the derived ARMA(p, q) process, and $\hat{\lambda}_{D_m}, \hat{\lambda}_M$, and $\hat{\lambda}_Q$ denote the corresponding estimates for the power of the Peña and Rodríguez, Monti, and Ljung-Box tests.

Model Fit	N	k	ϕ_1	ϕ_2	θ_1	θ_2	$\hat{\lambda}_{D_m}$	$\hat{\lambda}_M$	$\hat{\lambda}_Q$
White Noise	30	5	.1860	N/A	.8632	N/A	**.9471**	.9446	.7096
AR(1)	30	5	.1860	N/A	.8632	N/A	**.6437**	.6140	.3955
MA(1)	30	5	.1860	N/A	.8632	N/A	**.1116**	.1032	.1079
White Noise	100	10	-.6639	N/A	.1716	N/A	**.9936**	.9811	.9873
AR(1)	100	10	-.6639	N/A	.1716	N/A	.0613	.0686	**.0688**
MA(1)	100	10	-.6639	N/A	.1716	N/A	**.7051**	.5755	.6488
White Noise	300	15	-.1472	N/A	.3672	N/A	**.8286**	.6452	.6015
AR(1)	300	15	-.1472	N/A	.3672	N/A	**.1499**	.1119	.1183
MA(1)	300	15	-.1472	N/A	.3672	N/A	.0644	.0662	**.0689**

Table 1. Power Analysis Results for ARMA(1,1) Derived Process

Model Fit	N	k	ϕ_1	ϕ_2	θ_1	θ_2	$\hat{\lambda}_{D_m}$	$\hat{\lambda}_M$	$\hat{\lambda}_Q$
White Noise	30	5	N/A	N/A	.8632	-.4690	.3046	**.4092**	.2278
AR(1)	30	5	N/A	N/A	.8632	-.4690	.3826	**.3897**	.2776
MA(1)	30	5	N/A	N/A	.8632	-.4690	**.1994**	.1907	.1217
White Noise	100	10	N/A	N/A	.1716	.2125	**.4154**	.3355	.3811
AR(1)	100	10	N/A	N/A	.1716	.2125	**.2000**	.1677	.1740
MA(1)	100	10	N/A	N/A	.1716	.2125	**.2746**	.2125	.2070
White Noise	300	15	N/A	N/A	.3672	-.1208	**1.0000**	.9997	.9899
AR(1)	300	15	N/A	N/A	.3672	-.1208	**.9225**	.7865	.7157
MA(1)	300	15	N/A	N/A	.3672	-.1208	**.3115**	.2101	.2100

Table 2. Power Analysis Results for MA(2) Derived Process

Model Fit	N	k	ϕ_1	ϕ_2	θ_1	θ_2	$\hat{\lambda}_{D_m}$	$\hat{\lambda}_M$	$\hat{\lambda}_Q$
White Noise	30	5	.1860	-.6302	N/A	N/A	.7572	.7782	**.7890**
AR(1)	30	5	.1860	-.6302	N/A	N/A	**.8718**	.8334	.8344
MA(1)	30	5	.1860	-.6302	N/A	N/A	**.7713**	.7326	.6032
White Noise	100	10	-.6639	-.5996	N/A	N/A	1.0000	**1.0000**	.9999
AR(1)	100	10	-.6639	-.5996	N/A	N/A	**.9998**	.9976	.9979
MA(1)	100	10	-.6639	-.5996	N/A	N/A	**.9319**	.8811	.8949
White Noise	300	15	-.1472	-.3845	N/A	N/A	**1.0000**	.9992	.9991
AR(1)	300	15	-.1472	-.3845	N/A	N/A	**.9999**	.9985	.9984
MA(1)	300	15	-.1472	-.3845	N/A	N/A	**.9994**	.9950	.9934

Table 3. Power Analysis Results for AR(2) Derived Process

We can see from the above that the portmanteau test proposed by Monti had an estimated power better than or similar to that of the Ljung-Box test on several occasions, the most notable of these being when either the moving average component of the derived ARMA(p, q) process was underestimated or when the sample size and lag used for testing was sufficiently small (e.g., N = 30 and k = 5, respectively). These results agree with those obtained by Monti.[6]

Furthermore, in 7 out of the 9 cases in which the derived ARMA(p, q) process had a large sample size and lag used for testing was high (e.g., N = 300 and k = 15), the results of the power analysis for the portmanteau test suggested by Monti continued to outperform that of the Ljung-Box test.

The results of the power analysis for the Ljung-Box test meanwhile, were quite mixed. Although the Ljung-Box test tended to have an estimated power greater than than that of Monti's portmanteau test for occasions when the number of lags used for testing and the sample size of the derived ARMA(p, q) process was moderately large (e.g., N = 100 and k = 10), the Ljung-Box test was almost always outperformed by Peña and Rodríguez's portmanteau test, regardless of sample size, number of lags used for testing, or type of derived process. Indeed, in 24 of the 27 simulations ran for the power analysis, the portmanteau test suggested by Peña and Rodríguez had an estimated power greater than that of the Ljung-Box test.

However, the results seen in the performance of the portmanteau test suggested by Peña and Rodríguez compared to that of the Ljung-Box test are not surprising for the sample sizes which were greater than or equal to 100, and further support the results originally obtained by Peña and Rodríguez.[7] What should be noted though, is that in 6 of the 9 cases tested in which the derived ARMA(p, q) process had a small sample size and the number of lags used for testing was also small (e.g., N = 30 and k = 5), the portmanteau test proposed by Peña and Rodríguez outperformed the other two portmanteau tests (e.g., had greater estimated power).

Applied example: Sheep.

In order to show the results of the power analysis in an applied context, we performed a time series analysis of a real data set. The data set, *Sheep*, is of the total annual sheep population (1000s) in England and Wales from 1867 to 1939 ($N = 73$).[8] Firstly, an initial visual inspection of the time series was performed in order to address the requirement of weak stationarity prior to fitting any ARMA models to the data.
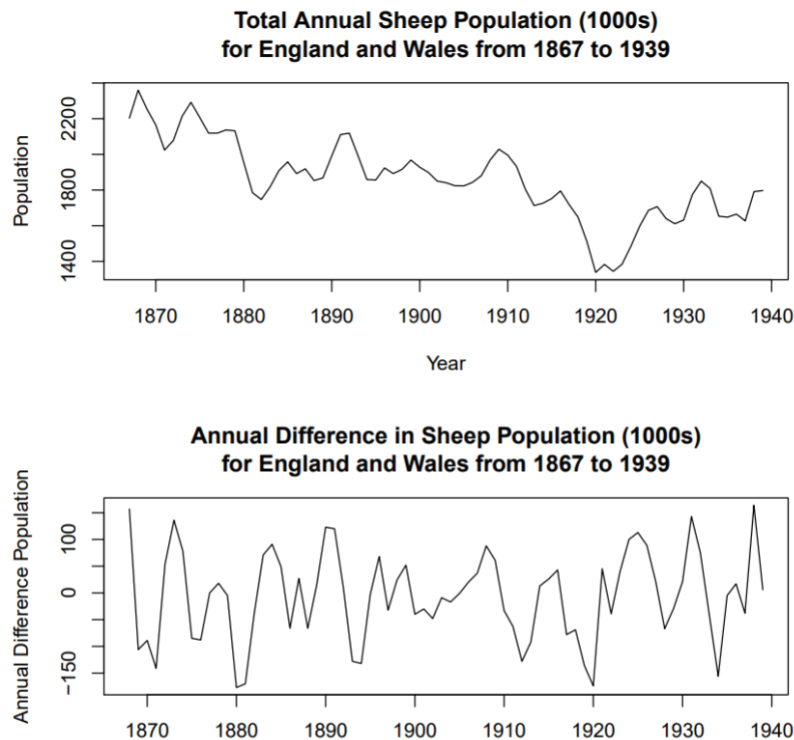

Total Annual Sheep Population (1000s) for England and Wales from 1867 to 1939

Annual Difference in Sheep Population (1000s) for England and Wales from 1867 to 1939

Figure 1. *Sheep* data visualization. The top graph shows a time series plot of the data which exhibits a negative linear trend and the bottom graph shows the differenced time series which appears to be stationary.

The time series plot of the data shows a general decreasing linear mean trend. Hence, the data was stationarized in order to satisfy the assumptions of fitting an ARMA(p, q) model to the time series. It is clear from the second plot that we successfully stationarized the sheep data by taking the first difference of the time series. We can now safely examine the autocorrelation function (ACF) and partial autocorrelation function (PACF) in order to conduct ARMA(p, q) modeling of the differenced data.

Judging from the plots of the ACF and PACF in Figure 2, an AR(2) model was found to be the most likely candidate for modeling the differenced time series as a higher order AR model would likely be an overfit of the data. From the power analysis, it was found that for moderately large sample sizes (e.g., $N \geq 100$), the Peña and Rodríguez test often has greater statistical power compared to the Monti and Ljung-Box tests. Therefore, we expected that the portmanteau test proposed by Peña and Rodríguez would be the most conservative in testing for residual autocorrelation of fitted ARMA models to the differenced *Sheep* data (likely followed by the Ljung-Box test and Monti's portmanteau test).

Table 4 shows the estimated coefficients for each of the ARMA models fit to the differenced *Sheep* time series data. The Akaike Information Criterion (AIC) is included to give an idea of the goodness of fit for each model relative to that of the others.[5] A smaller AIC value indicates a better fit. It can be noted that the ARMA(2,1) model has the lowest AIC out of all of the models considered and therefore is the model (out of the subset of ARMA fits considered) which best fits the differenced *Sheep* time series. Note that the white noise model has the highest AIC of all the ARMA models considered and therefore is the model which has the worst relative fit to the data.
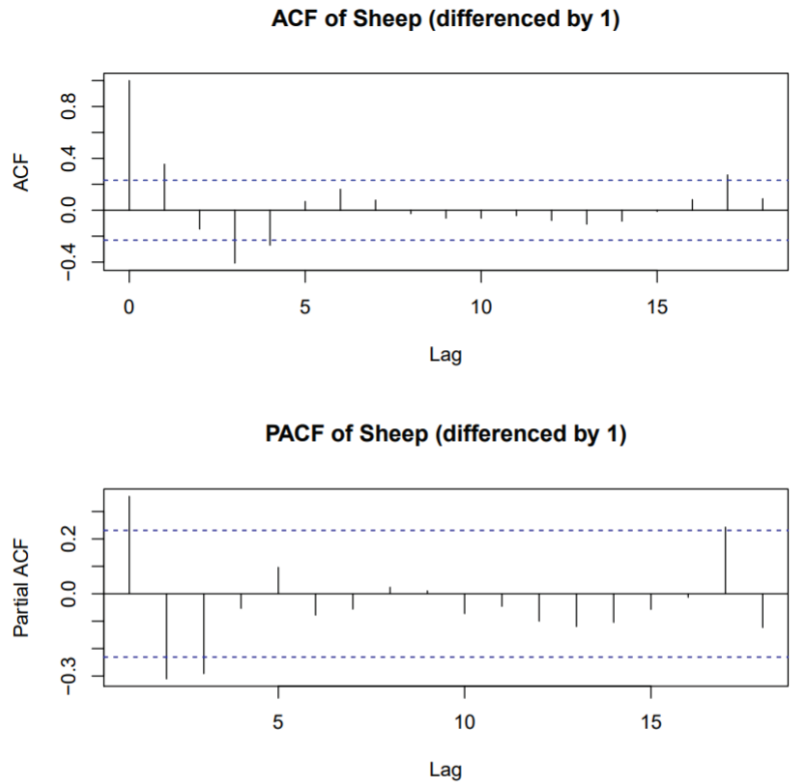
ACF of Sheep (differenced by 1)

PACF of Sheep (differenced by 1)

Figure 2. Residual Autocorrelation Analysis of Differenced *Sheep* Time Series Data.

Model Fit	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	AIC
AR(1)	.3734	N/A	N/A	832.71
MA(1)	N/A	N/A	.4798	828.83
AR(2)	.5348	-.3873	N/A	824.82
ARMA(1,1)	.0671	N/A	.4267	830.75
ARMA(2,1)	.9150	-.5454	-.4553	822.70
White Noise	N/A	N/A	N/A	841.12

Table 4. Results of *Sheep* data model fits.

For this example, we considered three types of models: "good" models (models of adequate fit, which we will define as those below the 50th percentile of AIC scores i.e. ARMA(2,1), AR(2), and MA(1)), "not so good" models (models of obvious underfit, which we will define as those above the 50th percentile of AIC scores i.e. ARMA(1,1) and AR(1)), and White Noise models (e.g., models of a Gaussian stochastic process).

Model Fit	$p^*_{D_m}$	p^*_M	p^*_Q	Overall Conclusion
AR(1)	.0123	.0637	.0168	Inconclusive
MA(1)	.0630	.1769	.1776	No Residual Autocorrelation Detected
AR(2)	.1548	.4670	.5120	No Residual Autocorrelation Detected
ARMA(1,1)	.0193	.1244	.1313	Inconclusive
ARMA(2,1)	.4560	.6759	.7042	No Residual Autocorrelation Detected
White Noise	.0004	.0040	.0001	Residual Autocorrelation Detected

Table 5. Results of Portmanteau Tests of Sheep Data Models.

Table 5 shows the p-values corresponding to each of the portmanteau tests considered against each of the ARMA models applied to the differenced *Sheep* time series data. It was found that for the AR(1) model, at a significance level of 5% and specified number of lags set to 10, Peña and Rodríguez's portmanteau test and the Ljung-Box test both detected sig-

nificant residual autocorrelation in the fitted model while the portmanteau test suggested by Monti failed to detect any significant residual autocorrelation.

Furthermore, it was found that for the MA(1), AR(2), and ARMA(2,1) models, at a significance level of 5% and specified number of lags set to 10, all of the portmanteau tests failed to detect any significant residual autocorrelation for these fitted models. The ARMA(2,1) model was found to be one of the "good" models for the fitted data, so as expected, none of the portmanteau tests found any significant residual autocorrelation for this model.

Additionally, for the ARMA(1,1) model, at a significance level of 5% and specified number of lags set to 10, the portmanteau test suggested by Peña and Rodríguez resulted in the rejection of the null hypothesis and therefore detected significant residual autocorrelation in the fitted ARMA model, while the rest of the portmanteau tests failed to detect any significant residual autocorrelation.

Lastly, the white noise model, at a significance level of 5% and specified number of lags set to 10, was found to have significant residual autocorrelation present by all three portmanteau tests. This was expected as the results from Table 3 showed all of the portmanteau tests have high estimated power for testing underfitting white noise models.

DISCUSSION

From the power analysis, it was found that Monti's portmanteau test had an estimated power better than or similar to that of the Ljung-Box test when either the moving average component of the derived ARMA(p, q) process was underestimated or when the sample size (and lag used for testing) were sufficiently small. These results support those originally obtained by Monti and those by Safi and Al-Reqep.[3, 6] Even for higher sample sizes and increased lag, Monti's portmanteau test continued to fare better than that of the Ljung-Box test.

Meanwhile, the estimated power for the Ljung-Box test compared to that of Peña and Rodríguez's portmanteau test was quite poor, and had an estimated power greater than Peña and Rodríguez's portmanteau test in only 3 of the 27 simulations conducted. For large sample sizes, Peña and Rodríguez found similar results,[7] but our power analysis demonstrated that Peña and Rodríguez's portmanteau test still had sufficiently high estimated power for a majority of the cases where the sample size (and number of lags used for testing) was small. This expands upon the findings of Safi and Al-Reqep as their paper does not discuss Peña and Rodríguez's original test statistic as a viable option in this situation.[3]

These results are quite surprising, as Peña and Rodríguez showed that the approximate convergence of their D_m test statistic to the Gamma distribution with mean $(m + 1)/2 - (p + q)$ and variance $(2m + 1)(m + 1)/3m - 2(p + q)$ is quite slow and often lacking in power for small sample sizes.[7] These results challenge that idea, as although Peña and Rodríguez's portmanteau test was shown to suffer from reduced power on a few occasions in which the sample size was small, the estimated power of Peña and Rodríguez's portmanteau test was still close to that of the portmanteau test suggested by Monti (which in turned fared better than the Ljung-Box test for 6 of the 9 simulations in which the sample size and lag used for testing was small).

CONCLUSIONS

In summary, our results confirm that the portmanteau test proposed by Monti is a viable alternative to the Ljung-Box test for when either the moving average component of the fitted ARMA model is underestimated or when sample size is small. However, for time series with small to moderate sample sizes, the portmanteau test proposed by Peña and Rodríguez may be considered, as this portmanteau test performed well for all sample sizes examined.

Furthermore, the time series analysis of the *Sheep* data supports the results found in our initial power analysis of the portmanteau tests considered in this paper, and this is evident in the results of the goodness of fit tests performed on the ARMA(p, q) models fit to the data. We can see that for when a possible moving average component was underestimated, Monti's portmanteau test was more conservative, and overall, gave results similar to that of the Ljung-Box

test, while Peña and Rodríguez's portmanteau test was the most conservative with regards to all of the non white-noise model fits (see Table 5 above). Moreover, we provide evidence that the original test statistic proposed by Peña and Rodríguez may be suitable for sample sizes as small as 30, contradicting the initial findings of Peña and Rodríguez as well as Safi and Al-Reqep.[3, 7]

Future work regarding this subject matter should investigate further applications of the portmanteau tests considered in this study. Furthermore, the theory of these portmanteau tests extends to testing for conditional heteroscedasticity in raw time series data (e.g., checking for the presence of GARCH effects in an otherwise stochastic time series process). There seems to be very little statistical literature regarding applying Monti's and Peña and Rodríguez's portmanteau tests to this task.

REFERENCES
1. Fisher, T. J. (2011) Testing Adequacy of ARMA Models Using a Weighted Portmanteau Test on the Residual Autocorrelations, *SAS Global Forum 2011 - Statistics and Data Analysis. support.sas.com/resources/papers/proceedings11/327-2011.pdf*
2. Moran, P. A., and Whittle, P. (1951) Hypothesis Testing in Time Series Analysis, *Journal of the Royal Statistical Society*, Series A (General), Vol. 114, No. 4, pp. 579.
3. Safi, S. K., and Al-Reqep, A. A. (2014) Comparative Study of Portmanteau Tests for the Residuals Autocorrelation in ARMA Models. *Science Journal of Applied Mathematics and Statistics* Vol. 2, No. 1, pp. 1–13. *doi.org/10.11648/j.sjams.20140201.11*
4. Ljung, G. M., and Box, G. E. P. (1978) On a measure of lack of fit in time series models, *Biometrika*, Vol. 65, No. 2, pp. 297–303. *https://doi.org/10.1093/biomet/65.2.297*
5. Akaike, H. (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, pp. 716–723. *doi.org/10.1109/TAC.1974.1100705*
6. Monti, A. C. (1994) A Proposal for a Residual Autocorrelation Test in Linear Models, *Biometrika*, Vol. 81, No. 4, pp. 776–790. *https://doi.org/10.1093/biomet/81.4.776*
7. Peña, D., and Rodríguez, J. (2002) A Powerful Portmanteau Test of Lack of Fit for Time Series, *JASA*, Vol. 97, No. 458, pp. 601–610. *https://doi.org/10.1198/016214502760047122*
8. Kendall, M. G., and Ord, J. K. (1990) Annual Sheep Population (1000s) in England & Wales 1867–1939, DataMarket. *datamarket.com/data/set/22px/annual-sheep-population-1000s-in-england-wales-1867-1939#!ds=22px&display=line* (accessed May 2019)

ABOUT THE STUDENT AUTHORS
Zachary Wenning and Emily Valenci are recent undergrad graduates from the Pennsylvania State University. Both graduated with Bachelor of Science degrees in Statistics under the Biostatistics option. Zachary Wenning is currently a Junior Analytics Consultant, and is a graduate student of the Pennsylvania State University's World Campus pursuing a Master of Applied Statistics degree. Emily Valenci is a Data Analyst and Statistical Researcher.

PRESS SUMMARY
When modeling autocorrelated data such as time series, it is often required to check how well the model describes the data. One such way to do this is to apply statistical tests called portmanteau tests. Our paper compares the statistical power, or the proportion of times a statistical test correctly rejects a null assumption, for three different portmanteau tests. We show that there is noticeable improvement in statistical power for two more modern portmanteau tests compared to the commonly used Ljung-Box test.