Gaussianity versus Correlation in Time Series Residual Diagnostics

James Livsey * Tucker McElroy * Anindya Roy *†

Abstract

The time series model selection problem is strongly rooted in residual analysis. Due to increased data collection, computing power and public interest, more-and-more important time series models are fit using iterative computer aided methods. These iterative methods usually fit competing models and compare residuals or likelihood based criteria to select then best models. The basis of this work is to improve the iterative methodology.

This paper will discuss current methods used during iterative model fitting; then, introduce a measure for model residuals that weighs whiteness and marginal distribution concurrently. For example, if residuals are assumed normal then perfect residuals are Gaussian noise. We seek a single score that is minimized at uncorrelated Gaussian and grows as residuals deviate from either property. Distributional theory of such a measure will be described. Finally, empirical studies will be presented supporting ideas and illustrating the practical uses of such a measure.

Key Words: density expansion, Hermite polynomials, normality, stationarity

1. Introduction

Many classical methods for fitting stationary time series models perform best when residuals follow a Gaussian distribution. Moreover, when utilizing the Box-Jenkins time series model fitting procedure, the typical progression (assuming data is already stationary with no heteroscedasticity) is to fit a model and check if residuals have any correlation present [1]. Then, one may further look to see if a suitable set of residuals follow the assumed Gaussian structure. This procedure can lead to a large number of models that all "pass" the correlation and normality tests. Final decisions are then usually made by experience, knowledge of the data generating process, parsimony, or a multitude of other facets.

A second way of fitting is to automate the process. One decides on a class of models and then lets a self-regulating algorithm choose the best model in the class. This is usually done via AIC (or one of the many AIC variants)[3, 8]. For example, many automated routines exist to fit ARMA models that, in theory, will have residuals that follow any prescribed assumptions. These routines conduct a search over possible models within order constraints given by the user and select a model based on likelihood driven criteria. Many automated routines are specific to a model class.

Throughout this paper, the premise of the proposed approach to time series modeling is that better models produce residuals that are more like Gaussian white noise. Questions we wish to address with this research are very fundamental: given two sets of residuals, which is closer to Gaussian white noise. This question cannot be answered by performing separate goodness-of-fit tests on both sample correlation and marginal structure. This is because such tests rely on finding evidence to refute the proposed characteristics. Hence, comparing multiple models which return satisfactory test values is not informative. We strive to produce a test with "black box" capabilities that does not depend on which model is chosen, i.e. given two sets of residuals, no knowledge of the model is needed to decide which set of residuals is closer to the idealized set of uncorrelated Gaussian residuals.

*U.S. Census Bureau, 4600 Silver Hill Rd, Washington, DC 20233

†Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21250

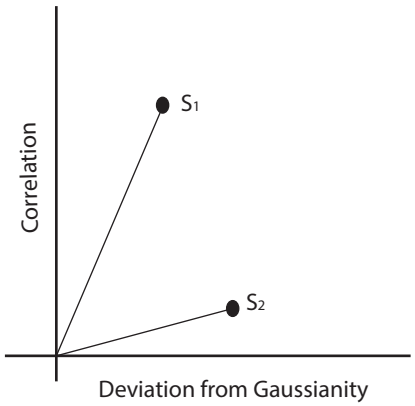

Figure 1: Example plot showing two hypothetical scores from two separate sets of residual values.

The proposed approach differs from current methods in that it will be able to compare two vastly different models. Much work has already been done in residual analysis. However, much of it focuses on only testing for correlation, only testing for marginal structure, or is model based, i.e. requires knowledge of the model that produced the residuals - possibly requiring them to produce a likelihood, be nested, etc. We create a single powerful test that can determine which set of residuals (and hence which model) more closely follows the assumption of uncorrelated Gaussian.

1.1 Simple motivating example

To further outline the goals of our work, consider two sets of hypothetical residuals $\{\epsilon_t\}$ and $\{\delta_t\}$, each of which pass a test for serial correlation (ACF plot, Ljung-Box test, Durbin-Watson test, etc) [5, 2] and a test for Gaussianity (QQ-norm plot, Shapiro-Wilks test, Kolmogorov-Smirnov test) [7]. We now wish to decide, in an automated manor, whether $\{\epsilon_t\}$ or $\{\delta_t\}$ is the better set of residuals. The goal of this research is to produce a test statistic, S, that will concurrently measure the amount of correlation present and the amount each set of residuals deviates from the normal distribution. Let S_1 denote the value of S obtained based on $\{\epsilon_t\}$ and S_2 be that obtained based on $\{\delta_t\}$. This is shown graphically in Figure 1. In this hypothetical situation, we would conclude that $\{\epsilon_t\}$ deviates less from the Gaussian distribution than $\{\delta_t\}$ and $\{\delta_t\}$ has less correlation than $\{\epsilon_t\}$. Most importantly, we see that $\{\delta_t\}$ is closer to Gaussian white noise than $\{\epsilon_t\}$. This same idea has been investigated in independent component analysis (ICA); the decomposition of a random vector [4].

2. Measuring Gaussianity

There are a multitude of well established methods for measuring how close a set of data, or in our case residuals, is to a set of normally distributed variables. Our initial empirical work has explored many methods and focused on two: one based on the empirical distribution function and one based on an expansion using Hermite polynomials.While we are not ready to state conclusively that Hermite polynomials are most fruitful, in our investigation Hermite expansion based tests performed well. The remainder of this proceedings will focus on measuring Gaussianity with a Hermite expansion.

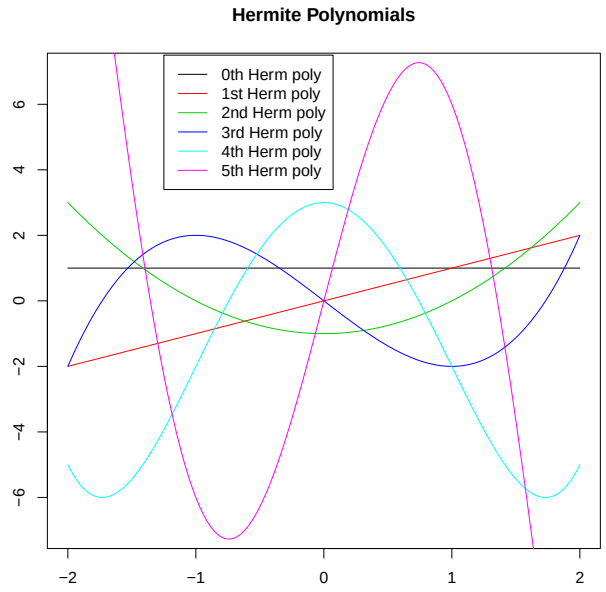

Figure 2: Plot of the first six Hermite Polynomials.

2.1 Hermite Polynomials

There are many variations of Hermite polynomial definitions. We define the Hermite polynomials as

$$H_n(x) = (-1)^n e^{\frac{1}{2}x^2} \frac{d^n}{dx^n} e^{-\frac{1}{2}x^2}, \tag{1}$$

for $-\infty < x < \infty$, $n = 0, 1, 2, \ldots$ and $\frac{d}{dx}$ is the derivative operator. The first few Hermite polynomials under definition (1), shown graphically in Figure 2, are:

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ H_4(x) &= x^4 - 6x^2 + 3, \\ H_5(x) &= x^5 - 10x^3 + 15x. \end{aligned}$$

We will use these polynomials to express the density of a given set of residuals. This particular definition is useful when measuring Gaussianity since under (1) the polynomials $\{H_n\}$ are orthogonal in terms of a inner product defined with respect to the standard normal density function, $\phi(x)$. Specifically,

$$< H_n,\ H_m >= n!\, \mathbf{1}_{(n=m)}$$

where $< f,\ g >= \int_{-\infty}^{\infty} f(x)g(x)\phi(x)dx$ is an inner product and $\mathbf{1}_{(A)}$ is the zero-one indicator function of the event A.

We consider two different density representations. First,

$$f(x) = \phi(x) \sum_{n=0}^{\infty} a_n H_n(x). \tag{2}$$

The convention of including $\phi(x)$ in the representation (2) assures that f will integrate to unity. Representation (2) does not guarantee $f(x) \geq 0$. However, this is not an issue; the end goal is not density estimation but a measure of how far f deviates from ϕ. The coefficients $\{a_n\}$ have a nice form when $f = \phi$. Namely, the sequence is extremely sparse:

$$\mathbf{a} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix}.$$

A second choice for density representation will be,

$$f(x) = e^{-\alpha x^2} \sum_{n=0}^{\infty} b_n H_n(x). \tag{3}$$

where $0 < \alpha \leq \frac{1}{2}$. This representation is proportional to (2) when $\alpha = 1/2$. The reason for considering two separate representations is as follows. Assume we are given a set of residuals $x_1, x_2, \ldots, x_T$ that has been standardized, i.e. centered about its mean and made to have standard deviation one. The density of the population that $\{x_t\}$ was drawn from $f(x)$ can be approximated by calculating estimates of the coefficients in (2) and (3) by treating the integration as an expectation with respect to the density f. Under the first representation (2)

$$\begin{aligned} a_n &= < \phi^{-1} f, H_n > \\ &= \int_{-\infty}^{\infty} \phi^{-1} f(x) H_n(x) \phi(x) dx \\ &= E_f[H_n(X)]. \end{aligned}$$

Which immediately leads to an estimator of

$$\hat{a_n} = T^{-1} \sum_{t=1}^{T} H_n(x_t). \tag{4}$$

Equation (4) provides a particularly simple and interesting representation for the coefficients. It can be viewed as a method of moments estimator. $\hat{a}_0 = 1$, $\hat{a}_1 = \bar{x}$, $\hat{a}_2 = T^{-1} \sum_{t=1}^{T} x_t^2 - 1, \ldots$. One issue with this representation (2) is the mean of each estimate $\hat{a}_n$ is not zero.

Alternatively, under the representation (3), we get

$$\begin{aligned} b_n &= < e^{\alpha x^2} f(x), H_n(x) > \\ &= \int_{-\infty}^{\infty} e^{\alpha x^2} f(x) H_n(x) \phi(x) dx \\ &= \frac{1}{\sqrt{2\pi}\, n!} E_f \left[e^{(\alpha - \frac{1}{2}) X^2} H_n(X) \right] \end{aligned}$$

which again provides a method of moments estimator,

$$\hat{b_n} = \frac{1}{\sqrt{2\pi}\, n!} \frac{1}{T} \sum_{t=1}^{T} e^{(\alpha - \frac{1}{2}) x_t^2} H_n(x_t). \tag{5}$$

To have a similar interpretable coefficients $\{b_n\}$, it is advantageous to no longer preprocess the residuals $\{x_t\}$ to have variance one. Instead, consider the normal density with variance $1/(2\alpha)$,

$$\phi_{\sigma^2=1/(2\alpha)}(x) = \sqrt{\frac{\alpha}{\pi}} e^{-\alpha x^2}.$$

Comparing this to (3), we see that for $f(x) = \phi_{\sigma^2=1/(2\alpha)}(x)$ the coefficients should be $b_0 = \sqrt{\frac{\alpha}{\pi}}$ and $b_n = 0$ for $n \geq 1$. It would be more convenient if $b_0 = 1$ and the variance of the data could be standardized to 1. This is achieved by using the coefficients $\sqrt{\alpha/\pi} b_n$ and setting $Y = \sqrt{2\alpha} X$. For notational convenience, we denote the modified coefficients as b_n. A natural estimator of b_n following (5) is

$$\hat{b}_n = \frac{1}{\sqrt{2\alpha}\, n!} \frac{1}{T} \sum_{t=1}^{T} e^{\frac{1}{2}(1-\frac{1}{2\alpha})y_t^2} H_n\left(\frac{y_t}{\sqrt{2\alpha}}\right) \tag{6}$$

Properties of this estimator are explored in [6]. The term $e^{\frac{1}{2}(1-\frac{1}{2\alpha})y_t^2}$ in (6) acts as an exponential dampening for outliers.

2.2 Covariance of estimated Coefficients

Our exposition will require the covariance of any two coefficients, $\hat{b}_n$ and $\hat{b}_m$. Let $\hat{\mathbf{b}} = (\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_J)'$ and denote its associated covariance matrix Σ_b. Our measure of Gaussianity does not assume uncorrelated residuals. As seen, calculation of covariance is done assuming $x_1, x_2, \ldots, x_T$, a covariance stationary sequence with mean 0 and autocovariance function $\gamma(\cdot)$. To devise a test for Gaussianity, we are primarily interested in the null distribution of the test statistic. Hence, we compute the covariance under the null hypothesis of Gaussian residuals and assume that the residuals follow a Gaussian process where for any s, t

$$\begin{pmatrix} x_s \\ x_t \end{pmatrix} = N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \gamma(0) & \gamma(|s-t|) \\ \gamma(|s-t|) & \gamma(0) \end{bmatrix}\right).$$

Each Hermite polynomial is given the representation

$$H_n(x) = \sum_{j=0}^{J} \xi_{n,j} x^j$$

where J is the maximum power of any Hermite polynomial used and $\xi_{n,j}$ are Hermite polynomial coefficients. The convention of a fixed J is used as opposed to letting J depend on n, leading to many $\xi_{n,j}$ being zero. For example, if $J = 10$ we would express $H_4(x) = x^4 - 6x^2 + 3 = \sum_{j=0}^{10} \xi_{4,j} x^j$ where $\xi_{4,4} = 1$, $\xi_{4,2} = -6$, $\xi_{4,0} = 3$ and all other $\xi_{4,j}$ equal zero.

The estimated coefficients following (6) can be written as

$$\hat{b}_n = T^{-1} \sum_{t=1}^{T} g_n(x_t)$$

where $g_n(x) = \frac{1}{\sqrt{2\alpha n!}} e^{\alpha x^2} H_n(x/\sqrt{2\alpha})$. We are interested in calculating

$$Cov(\hat{b}_n, \hat{b}_m) = Cov\left(T^{-1}\sum_{t=1}^{T} g_n(x_t),\ T^{-1}\sum_{s=1}^{T} g_m(x_s)\right)$$
$$= T^{-2}\sum_{t=1}^{T}\sum_{s=1}^{T} Cov(g_n(x_t), g_m(x_s)).$$

Expanding the Hermite polynomials we get

$$Cov(g_n(x_t), g_m(x_s)) = \frac{1}{2\alpha n! m!}\sum_{j=0}^{J}\sum_{k=0}^{J}(2\alpha)^{-\frac{j+k}{2}}\xi_{n,j}\xi_{m,k}Cov(e^{\alpha x_t^2}x_t^j, e^{\alpha x_s^2}x_s^k)$$
$$= \frac{1}{2\alpha n! m!}\sum_{j=0}^{J}\sum_{k=0}^{J}(2\alpha)^{-\frac{j+k}{2}}\xi_{n,j}\xi_{m,k}\underbrace{E[e^{\alpha x_t^2}x_t^j, e^{\alpha x_s^2}x_s^k]}_{\theta_{j,k,t,s}}.$$

To get an explicit expression for $\theta_{j,k,t,s}$ the residuals are scaled to unit variance.

$$\theta_{j,k,s,t} = E[e^{\alpha x_t^2}x_t^j, e^{\alpha x_s^2}x_s^k]$$
$$= \gamma(0)^{\frac{j+k}{2}}E\left[e^{\beta x^2}e^{\beta y^2}x^j y^k\right]$$

where $\beta = \gamma(0)\ \alpha$ and $\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\mathbf{0},\ \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$, for $\rho = \rho_{t,s} = \gamma(|t-s|)/\gamma(0)$. A formal derivation of $E\left[e^{\beta x^2}e^{\beta y^2}x^j y^k\right]$ is given in the Appendix and expressing the sums in quadratic form we have,

$$\Sigma_b = T^{-2}\sum_{h=0}^{T-1}\Xi Q_h \Xi'$$

where $[\Xi]_{nj} = \xi_{n,j}/n!$ and $[Q_h]_{jk} = (2\alpha)^{-\frac{j+k+2}{2}}E\left[e^{\alpha(x_t^2+x_{t+h}^2)}x_t^j x_{t+h}^k\right]$. The notation $[A]_{ij}$ gives the $(i,j)^{th}$ element of the matrix A.

2.3 A measure of Gaussianity

Under Gaussianity, the value of the true coefficient vector $\mathbf{b}$ is $\mathbf{e}_1 = (1, 0, \ldots, 0)'$. Thus , a natural measure of Gaussianity could be based on the distance between $\hat{\mathbf{b}}$ and $\mathbf{e}_1$. Since the different entries of $\hat{\mathbf{b}}$ have different variance, we consider a weighted distance as proposed measure of Gaussianity,

$$G_0 = (\hat{\mathbf{b}} - \mathbf{e}_1)'\Sigma_{\mathbf{b}}^{-1}(\hat{\mathbf{b}} - \mathbf{e}_1) \tag{7}$$

where $\Sigma_{\mathbf{b}} = Var(\hat{\mathbf{b}})$. We are now ready to state the main theorems that yield relevant distribution theory and can be used to set up a test of Gaussianity based on J.

Theorem 2.1 *Assuming $\{X_t\}$ is a mean zero stationary Gaussian process,*

$$\sqrt{T}(\hat{\mathbf{b}} - \mathbf{e_1}) \longrightarrow N(\mathbf{0}, \Sigma_b).$$

where $\mathbf{e}_1 = (1, 0, 0, \ldots, 0)$.

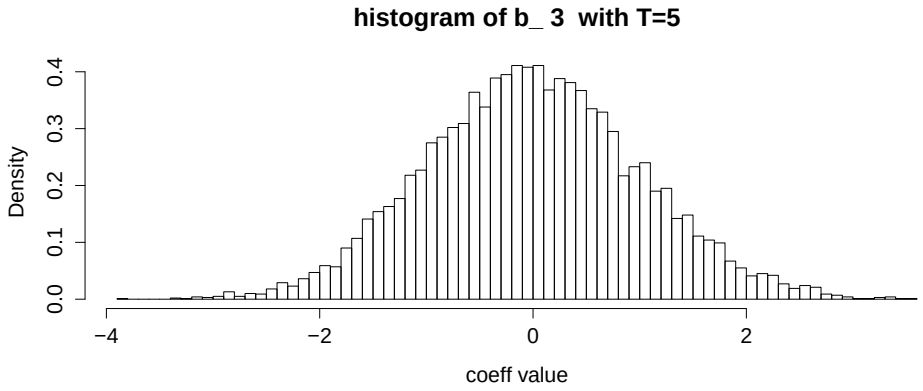

Figure 3: Histogram of 10000 simulations of standardized $\hat{b}_3$ coefficient estimates from $T = 5$ residual values.

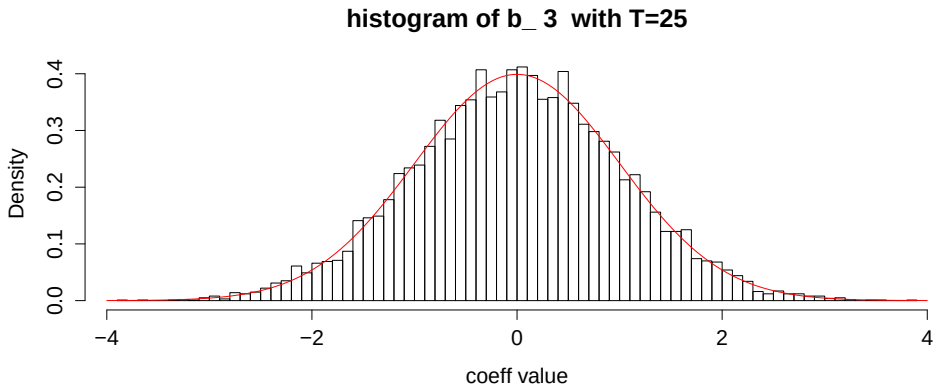

Figure 4: Histogram of 10000 simulations of standardized $\hat{b}_3$ coefficient estimates from $T = 25$ residual values. Over-plotted with the standard normal density

Theorem 2.2 *Assuming $\{X_t\}$ is a mean zero stationary Gaussian process,*

$$\sqrt{T}(\hat{\mathbf{b}} - \mathbf{e_1})'\Sigma_b^{-1}(\hat{\mathbf{b}} - \mathbf{e}_1) \longrightarrow \chi^2(J+1).$$

where $\chi^2(\nu)$ represents that chi-squared distribution with ν degrees of freedom.

Under the above assumptions that residuals follow a stationary process, we propose a test of non-Gaussianity as G_0 that rejects the null hypothesis of Gaussian residuals if $G_0 > \chi^2_\alpha(J+1)$ where $\chi^2_\alpha(J+1)$ is the upper α percentile of the $\chi^2(J+1)$.

3. Simulation

We perform a limited simulation study to check the validity of the asymptotic approximation in Theorems 2.1 and 2.2. Figures 3 and 4 are histograms of a single component from $\Sigma_b^{-1/2}\hat{\mathbf{b}}$ with $T = 5$ and $T = 25$ for 10000 simulations of residuals that are IID standard normal. For simplicity we have only plotted a histogram for the third coefficient. The standard normal density is added to the plot when $T = 25$. All other coefficients have a similar looking shape. After the transform $\Sigma^{-1/2}\hat{\mathbf{b}}$ all histograms are standardized.

It remains to verify that calculating Σ_b under the assumption of a stationary process is necessary and the weighted distance between $\hat{\mathbf{b}}$ and $\mathbf{e}_1$ (instead of euclidean distance)

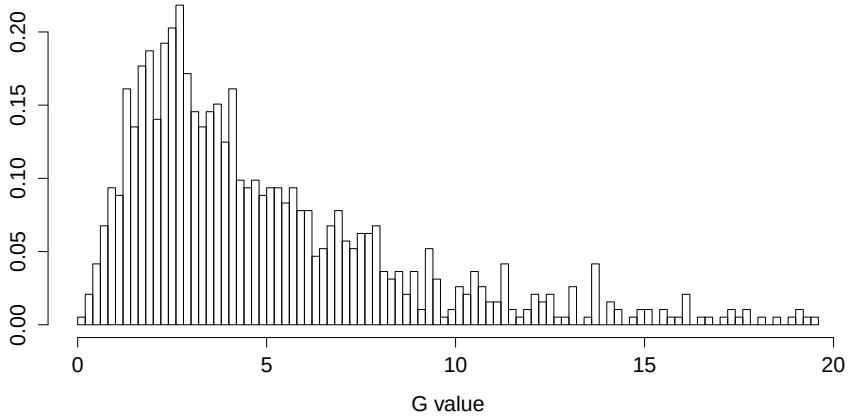

Figure 5: Histogram for 10000 simulations of G_0 statistic. Calculated under the assumption that residuals are a stationary process.

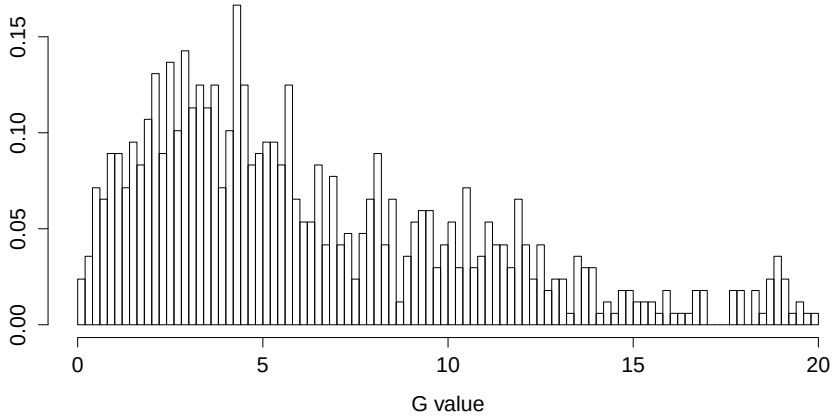

Figure 6: Histogram for 10000 simulations of G_0 statistic. Calculated under the assumption that residuals are uncorrelated.

is the appropriate measure to use. For this we simulated residuals from an autoregressive AR(1) process with $\phi = 0.5$. The sample autocovariances were used in our estimate of $\widehat{\Sigma}_b$. This process was repeated 10000 times, the resulting G_0 values are shown in Figure 5. This was done using $J = 5$ and the results fit Theorem 2.2 well. This same procedure was then repeated under the assumption that residual values were uncorrelated using the Euclidean distance between $\hat{\mathbf{b}}$ and $\mathbf{e}_1$ (all calculations are omitted from this report). The resulting G_0 values are plotted in Figure 6. Clearly, a very heavy tail indicates violation of a chi-squared distribution.

4. Concluding Remarks

In this proceedings, we have presented part of our ongoing research for developing a single statistic that can measure the departure of sample residuals from the idealized set of residuals following Gaussian white noise. We have only described a test for Gaussianity under the assumption residuals are possibly correlated. Further work must be done to both quantify the amount of correlation present and merge with ideas from section 2. In essence, we need to combine our test for Gaussianity under possibly correlated residuals with a test for serial

correlation of possibly non-Gaussian residuals. An initial power study of the test for normality yielded another interesting fundamental question. Namely, how can we generate a sample time series from a given marginal distribution with known ACVF $\gamma(\cdot)$? The answer is well known for the stationary Gaussian case, let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_T)' \sim N_T(\mathbf{0},\ I)$ then

$$\mathbf{X} = \Gamma_T^{-1/2}\mathbf{Z}$$

where $[\Gamma]_{ij} = \gamma(|i-j|)$ will be a stationary Gaussian time series with ACVF $\gamma(\cdot)$. This is not so easy for other marginal distributions. Our future work will also aim to make progress in this related area.

This report also focused on Gaussian marginals. While in practice this is the most widely utilized assumption, there are many interesting time series applications that have different marginal assumptions on residual values. As this research matures, other types of distributions will also be considered.

5. Appendix

In this appendix we provide explicit expressions for $E[e^{\beta x^2}e^{\beta y^2}x^j y^k]$ where $\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\mathbf{0},\ \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$. Let

$$E\left[e^{\beta x^2}e^{\beta y^2}x^j y^k\right] = \int x^j e^{\beta x^2} \underbrace{\int y^k \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}[y^2-2\rho xy+x^2]} e^{\beta y^2} dy}_{\zeta_{k,t,s}}\, dx.$$

The idea of the remaining calculation is to write $\zeta_{k,t,s}$ as a normal density with mean $\rho x/A$ and variance $(1-\rho^2)/A$ (with A given later) by completing the square i.e.

$$ay^2 + by + c = a\left(y + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c.$$

Let $A = 1 - 2\beta(1-\rho^2)$ and $D = \frac{1-\rho^2/A}{1-\rho^2}$. Then

$$
\begin{aligned}
&E\left[e^{\beta x^2}e^{\beta y^2}x^j y^k\right] \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}\int e^{\beta x^2}x^j \int y^k e^{\left\{-\frac{1}{2(1-\rho^2)}\left[(1-2(1-\rho^2)\beta)y^2-2\rho xy+x^2\right]\right\}}dydx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}\int e^{\beta x^2}x^j \int y^k e^{\left\{-\frac{A}{2(1-\rho^2)}\left[y^2-\frac{2\rho x}{A}\ y+\frac{4x^2\rho^2}{4A^2}-\frac{x^2\rho^2}{A^2}+\frac{x^2}{A}\right]\right\}}dydx \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}}\int e^{\beta x^2}x^j \int y^k e^{\left\{-\frac{A}{2(1-\rho^2)}\left[\left(y-\frac{x\rho}{A}\right)^2+\frac{x^2}{A}\left(1-\frac{\rho^2}{A}\right)\right]\right\}}dydx \\
&= \frac{1}{\sqrt{2\pi}\sqrt{A}}\int e^{\left\{\beta x^2\right\}}x^j e^{\left\{-\frac{1}{2(1-\rho^2)}(1-\rho^2/A)x^2\right\}}\int y^k \frac{\sqrt{A}}{\sqrt{2\pi}\sqrt{1-\rho^2}}e^{\left\{-\frac{A}{2(1-\rho^2)}\left(y-\frac{x\rho}{A}\right)^2\right\}}dydx \\
&= \frac{1}{\sqrt{2\pi}\sqrt{A}}\int x^j e^{\left\{-\frac{1}{2}\left(\frac{1-\rho^2/A}{1-\rho^2}\right)x^2\right\}}E\left[N\left(\frac{x\rho}{A},\ \frac{1-\rho^2}{A}\right)^k\right]dx \\
&= \frac{1}{\sqrt{2\pi}\sqrt{A}}\int x^j e^{\left\{-\frac{1}{2}\left(\frac{1-\rho^2/A}{1-\rho^2}\right)x^2\right\}}\sum_{\ell=0}^{k}\binom{k}{\ell}\left(\frac{\rho x}{A}\right)^{k-\ell}(\ell-1)!!\left(\frac{1-\rho^2}{A}\right)^{\ell/2}\mathbf{1}(\ell \text{ even})dx \\
&= \frac{1}{\sqrt{A}\sqrt{D-2\beta}}\sum_{\ell=0}^{k}\binom{k}{\ell}\rho^{k-\ell}(1-\rho^2)^{\ell/2}A^{\ell/2-k}(\ell-1)!!\mathbf{1}(\ell \text{ even})\int x^{j+k-\ell}\frac{\sqrt{D-2\beta}}{\sqrt{2\pi}}e^{\left\{-\frac{1}{2}(D-2\beta)x^2\right\}}dx \\
&= \frac{1}{\sqrt{A}\sqrt{D-2\beta}}\sum_{\ell=0}^{k}\left[\binom{k}{\ell}\rho^{k-\ell}(1-\rho^2)^{\ell/2}A^{\ell/2-k}(\ell-1)!!\mathbf{1}(\ell \text{ even})(D-2\beta)^{-\frac{j+k-\ell}{2}}\right. \\
&\qquad\qquad \left.(j+k-\ell-1)!!\mathbf{1}(j+k-\ell \text{ even})\right]
\end{aligned}
$$

References

[1] Box, G. E. and Jenkins, G. (1970). *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.

[2] Box, G. E. and Pierce, D. A. (1970). "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models". *Journal of the American statistical Association*, 65(332), 1509-1526.

[3] Brockwell, P. J. and Davis, R. A. (2009). *Time series: theory and methods*. Springer.

[4] Cardoso, J. F. (2003) "Dependence, correlation and gaussianity in independent component analysis." *The Journal of Machine Learning Research*, 4, 1177-1203.

[5] Ljung, G. M. and Box, G. E. (1978). "On a measure of lack of fit in time series models". *Biometrika*, 65(2), 297-303.

[6] Puuronen, J. and Hyvärinen, A. (2011) "Hermite Polynomials and Measures of Non-gaussianity." *Artificial Neural Networks and Machine Learning–ICANN*, 205–212.

[7] Shapiro, S. S. and Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika*, 591-611.

[8] Shumway, R. H. and Stoffer, D. S. (2000). *Time series analysis and its applications* (Vol. 3). New York: Springer.