Investigation of Long Memory in Quarterly Housing Price Index Data

Vikas Garg, Yasin Khan, Shirley Lu, Timothy Roberts, Yifan Tang, Wesley Tillu

I. Introduction

This project seeks to understand the housing price process in a discrete time series setting. As key drivers of mortgage rates and valuation, addressing house price evolution is critical to investing in mortgage backed assets. Fundamental considerations relevant to our project scope will include, but will not be limited to, the effect of default and prepayment risk on mortgage pool valuation and the significance of geography in the housing market. Additionally, this project will investigate the presence of long memory in the housing price process, with a specific focus on long memory in housing price volatility.

In general, an understanding of the housing price process should lend itself to a model from which time series forecasts can be derived using appropriate methods and techniques.

II. Data

A. Interpolation

The Office of Federal Housing Enterprise Oversight (OFHEO) Housing Price Index (HPI) provides quarterly housing price data on a state-level starting from 1975. This quarterly data was interpolated to achieve monthly housing price data. Interpolation gives two benefits: first, monthly data matches the frequency of borrowers' payment streams, which is the basis for asset-backed securities, and second, interpolation increases our data set from 135 quarterly observations to 402 monthly observations, thus increasing the sample size significantly. Interpolation does have its drawbacks, though, of increasing estimation errors. The selection of appropriate interpolation techniques will aim to minimize such errors.

Two interpolation techniques were analyzed. First, a linear interpolation between each quarterly data point was used to estimate monthly housing prices. Second, a cubic spline interpolation method was implemented on the entire data set. For both methods, HPI's quarterly observations were fixed so that the interpolated data set matches the quarterly for the corresponding months as to reduce error. A plot of the quarterly and monthly time series shows the interpolation methods to be accurate. See Figure 1.1.1. The more granular zoomed-in plots show that a cubic spline is continuous and smooth as compared to the linearly interpolated data.

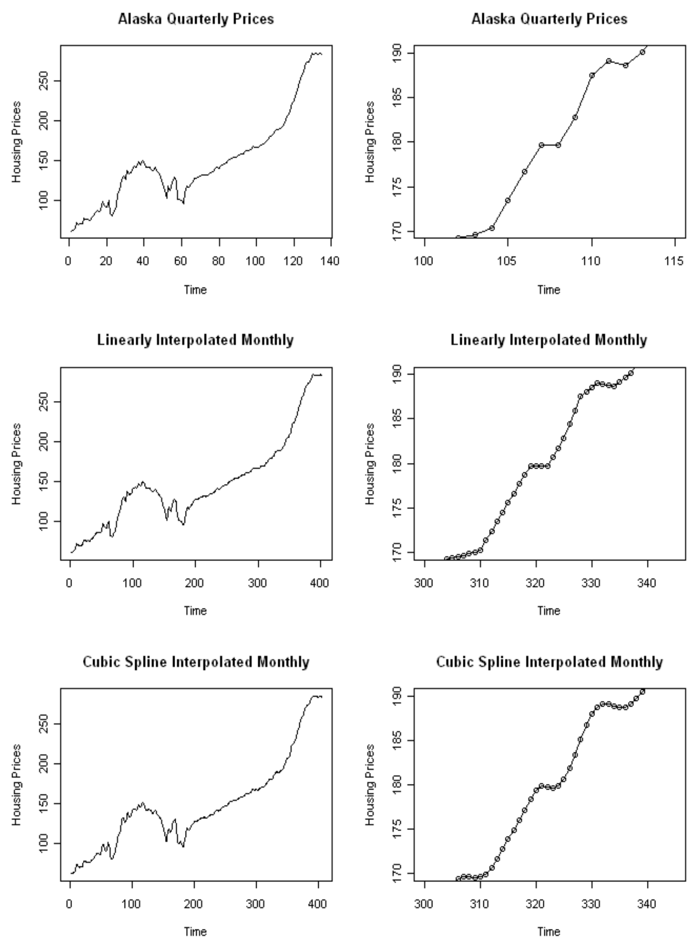


Figure 1.1.1: Comparison of quarterly with linearly and cubic-spline interpolated monthly prices for Alaska. The plots on the right-hand side shows a zoomed-in version of the original and interpolated data.

Furthermore, the two-sample Kolmogorov-Smirnov (K-S) test was used to statistically check whether the quarterly and the interpolated data sets are drawn from the same distribution. The hypothesis for the K-S test is defined as follows.

H_0: The two samples come from a common distribution.
H_A: The two samples do not come from a common distribution.

The p-value for both the linear and cubic-spline interpolation methods was 1, verifying that the data interpolation did not fundamentally alter the original observed data set.

The quarterly and monthly autocorrelation functions (ACF) for the simple returns were calculated and compared to validate that the monthly ACF at lag i approximately equals the quarterly ACF at lag $i+3$ for all i. Figure 1.1.2 shows the ACF plots of the quarterly and monthly simple returns. As expected, the monthly autocorrelations peak at every third multiple of where the quarterly autocorrelation peaks. This confirms that no additional autocorrelation was induced in the interpolation process. The monthly autocorrelations also exhibit the same pattern and shape as the quarterly, except with more granularity and precision. The same can be said for the squared simple returns series.

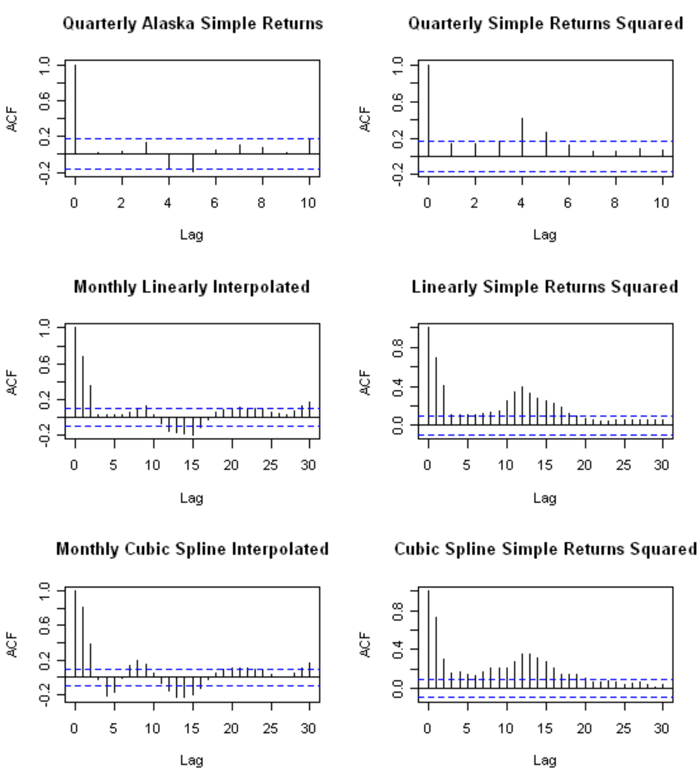
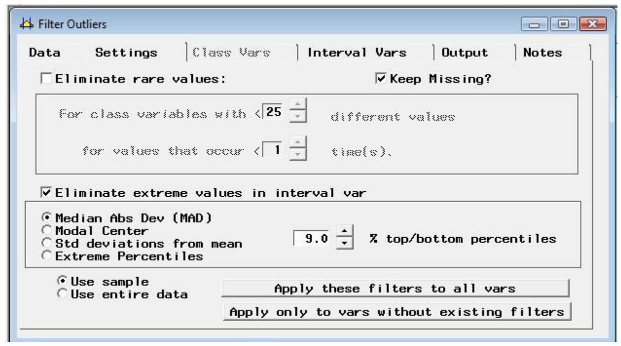


Figure 1.1.2: Autocorrelation plots of quarterly and interpolated monthly simple returns (left-side plots) and of the squared simple returns series (right-side plots) for Alaska.

B. Data Cut-Offs

1) SAS Treatment of Outliers

The treatment of outliers by the Filter Outlier node in SAS Enterprise Miner was investigated. In essence, SAS will eliminate observations based on the empirical quantiles of the ranked data. Different methods of choosing the empirical quantiles to be eliminated are available, including the Median Absolute Deviation, Modal Center, Standard Deviation, and Extreme Percentiles. Additionally, the desired percentage of the top and bottom quantiles to be eliminated can be specified.



Since our data is a time series, it is not appropriate to intermittently eliminate data points. Our goal is to determine a cutoff point from which we remove prior data for model construction.

2) Grubb's Test

A common method of identifying outliers in a data set is the Grubb's test (also known as the maximum normal residual test). One assumption of the test is that the data set is normally distributed, which is not the case for all the state level simple returns of the OHFEO HPI. Another shortfall of the test is that Grubbs' test detects one outlier at a time at certain significance levels. This outlier is expunged from the dataset and the test is iterated until no outliers are detected.

Grubbs' test is defined for the hypothesis:

H_0: There are no outliers in the data set

H_A: There is at least one outlier in the data set

The test statistic is:

$$G = \frac{\max\limits_{i=1,\ldots,N} \left|Y_i - \overline{Y}\right|}{s}$$

For the two-sided test, the hypothesis of no outliers is rejected at significance level α if:

$$G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t^2_{\alpha/(2N),N-2}}{N-2+t^2_{\alpha/(2N),N-2}}}$$

with $t_{\alpha/(2N),N-2}$ denoting the upper critical value of the t-distribution with N – 2 degrees of freedom and a significance level of α/(2N).

The Grubb's test was performed on the Alaska OFHEO HPIO simple returns. At the 10% confidence level, the returns for second quarter 1980 and second quarter 1989 are considered outliers. Identical results hold for the 5% confidence level. However, at the 1% confidence level, only the simple return for second quarter 1989 is considered an outlier by the Grubb's test.

3) Cut-off Determination

It was speculated that the OFHEO housing price index may have varying degrees of data quality. More specifically, the data collection methods before 1990 may have not been consistent, leading to incorrect index levels. This may be evidenced by the extremely high volatility in early periods of the simple return data – up to around quarter 40.

To learn more about any changes in data collection, OFHEO was contacted by phone. Based on the understanding of the OFHEO representatives, there has not been a change in the methodology of data collection for their HPI. It receives quarterly data electronically from the enterprises (ie. Fannie and Freddie), which in turn, is fed directly into their indexing methodology. There is no knowledge or documentation regarding discrepancies in data collection with Fannie and Freddie data; however, any change in data collection might be from the Fannie and Freddie side.

OFHEO noted that as a result of their repeat-transaction methodology, the HPI continues to change to reflect newer data. For example, if a home is sold just once in first quarter 1975, nothing happens to the index. If however, in 2008, the house is sold again, the HPI might be updated both in 1975 and 2008. This might show the index converges to a more "refined" estimate, and gives less cause for error.

Changes in Fannie and Freddie's data collection process will be further investigated. OFHEO will be providing a knowledgeable contact at Fannie and Freddie, where any data collection changes made have been made.

4) Treatment of Outliers

While no conclusive evidence was found of any change in data collection methodology, some type of cut-off seemed to be needed. The high variance in the simple returns implies either time-specific volatility or an error / change in data collection methodology. Since there was no evidence of increased volatility in housing prices from 1975 to early 1980s, it was concluded that the volatility was due to the latter reason.

Therefore, looking at the simple return charts for all states, a cut-off point of quarter 40, or the 1st quarter of 1985 was chosen. To be consistent, this cut-off point was used for all states under investigation. The shortened data set for states was used for the remainder of the analysis.

III. METHODS

A. Initial Analysis

1) Exploratory Data Analysis

The initial analysis of the data was carried out as follows:

The simple return series was created from the HPI data and time series plots were generated for all the states. If the simple return series created from the HPI data is stationary then the appropriate AR/MA or the ARMA order is determined (using the Partial Autocorrelation Function (PACF) and the Auto Correlation Function (ACF) plots) and then the model parameters are estimated. If the series is not stationary, it's tested for unit root stationarity. The first differenced series of the simple returns is created, for series which show the presence of a unit root. Next step is to determine the orders of an ARIMA model for this series. Box-Ljung test is used on residuals to test the adequacy of the model. The test is performed separately for the residuals, and on the squared residuals to test the adequacy of the mean equation and the presence of ARCH effects respectively. Given below are charts prepared for AR, which show a non stationary behavior in the simple return series, and a slowly decaying ACF plot. But the differenced series appears to be stationary.

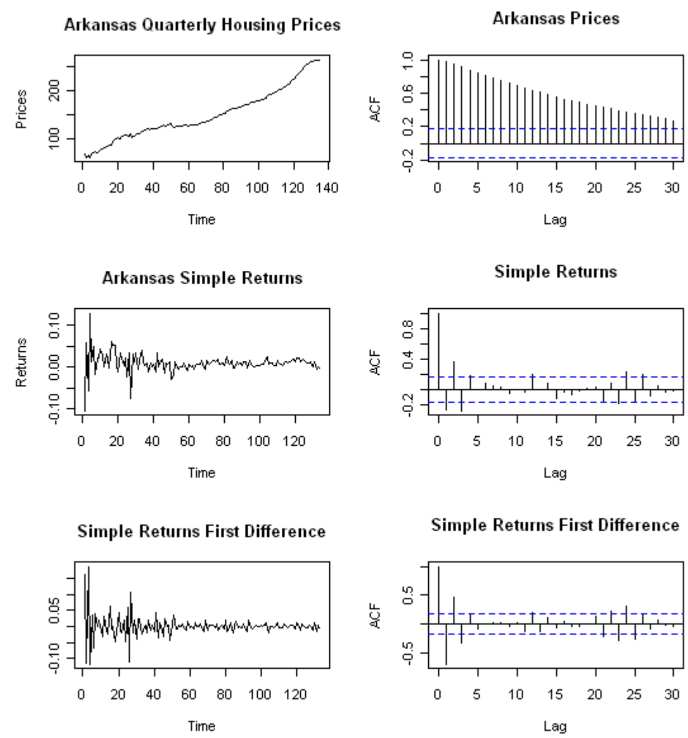


Figure 2.1: Plots and ACF of Arkansas quarterly housing prices, their simple returns, and the simple returns first difference.

2) Data Transformations

Many parameter estimators are based on the assumption of Gaussian distribution. In particular, the Whittle estimator for the fractional differencing parameter is one such estimator. The Whittle estimator is generally preferred over other fractional differencing parameters as it is consistent, unbiased, and asymptotically efficient.

The Box Cox transformation aims to apply a power transformation to ensure that the data follows an approximately normal distribution. The Box Cox transformation stabilizes variance and induces homoscedacity in the time series. It is defined as

$$y^{(\alpha)} = \begin{cases} \dfrac{y^{\alpha}-1}{\lambda}, \lambda \neq 0 \\ \log(y), \lambda = 0 \end{cases}$$

In practice, this transformation is applied to the data before any necessary differencing. Hence, the Box Cox was tested on the original housing price index data rather than their simple returns.
If initial plots of the OFHEO HPI data do not exhibit normality, the Box Cox test was used to estimate the appropriate power of λ along with its standard error. Then, the interval (λ +/- std. error) was evaluated to determine the power to apply. If the interval contained 0, a log transformation was used. Otherwise, the estimate or its closest fraction was used to transform the data.

The effectiveness of the data transformation was tested graphically with density plots and statistically using formal normality tests.

A weakness of the Box Cox transformation is that if the true distribution is far from Gaussian, there will not exist a sufficient λ that will make the data normal. However, even in cases where no power transformation could bring the distribution to exactly normal, the usual estimates of λ will lead to a distribution that satisfies certain restrictions on the first 4 moments, and thus will usually be symmetric.

3) I(1) Testing

To test for unit root stationary, primarily, the Augmented Dickey-Fuller (ADF) test was used. The ADF test checks for the presence of unit root in the time series. The hypothesis tests are summarized as follows:

H_o: $y_t \sim I(1)$, series contains a unit root
H_a: $y_t \sim I(0)$, series is stationary
where $y_t \sim$ i.i.d

The ADF test was implemented using adf.test within **R** (using the package "tseries"). To determine the lag order, AIC and BIC values were determined for multiple AR models. The lag associated with the lowest of AIC and BIC values was selected.

Additionally, the Phillips-Perron test was added, using pp.test within **R**, to substantiate test results from the ADF test. The hypotheses are the same as those from the ADF test.

4) Normality Testing

As stated before, testing the univariate data for normality was an essential part of this analysis, since Gaussian MLE's such as the Whittle estimate have parametric assumptions about the process under test. Graphical and statistical tests were employed in this study.

B. Parameter Estimation for Model Building

The current methodology estimates the model orders and the fractional differencing parameter of the model by a two step process:

In the first step the ARMA orders are determined to fit the mean equation by inspecting the PACF/ACF/EACF (Extended Auto Correlation Function) plots. Using these orders the value of "d" (fractional differencing parameter) is estimated by the GPH methodology and a fractionally differenced series si created. The model parameters are estimated by trial and error approach to determining an appropriate model that best fits the data.

Box-Ljung test is applied at each stage to check for model adequacy and to investigate the presence of ARCH effects. Model selection is based on BIC (Bayesian Information Criterion).

$$BIC(\ell) = \tilde{\sigma}_{\ell}^{2} + \frac{\ell \ln(T)}{T}$$

The model with the lowest BIC value is accepted as the final model. BIC was chosen over AIC for model comparison as BIC is found to give more superior results as compared to AIC for autoregressive model order selection [6]; also BIC enjoys the consistency property in terms of selecting the true model. [7]

1) Fractional Differencing Parameters

An appropriate investigation of fractional differencing includes entertaining several estimators for a fractional differencing parameter in an ARFIMA (p,d,q) framework. According to Beran, a linear process with i.i.d. innovations can exhibit long memory, however, it is difficult to distinguish the behavior of linear ARMA-type processes from fractionally integrated processes for small values of n, where n is the length of the univariate process under question.[1]

With these considerations in mind, it was necessary to include a fractional integration determination based on the following:

- An efficiency study of different estimators for the fractional differencing d, including "removal of fractional differencing."
- A formal statistical hypothesis test for fractional differencing. Specifically, Dolando *et. al.*[2] has proposed a Fractional Dickey Fuller test, in which, in the following regression setup,

$$\Delta y_t = \mu + \varphi \Delta^d y_{t-1} + A(L)\Delta y_t + \varepsilon_t$$

the null hypothesis of unit-root non-stationarity ($d=d_0=1$) is tested against fractional integration ($d=d_1\ \varepsilon\ (-1/2,1/2)$).

$$H_0: \varphi = 0,\ y_t \text{ is } FI(d_0)$$
$$H_A: \varphi < 0,\ y_t \text{ is } FI(d_1)$$

Initial results of this test are promising in simulation, but further investigation has to be carried out to validate statistical accuracy with respect to MARS-derived critical values (see Sephton[3]).

- An objective comparison of ARFIMA(p,d,q) (long memory) with an alternative model, using forecast errors.

2) Joint Estimation of ARFIMA(p,d,q) Parameters

The current methodologies for estimating the model parameters follow a two step approach; the first step is to obtain the estimate of fractional differencing parameter "d" and an ARMA estimation technique is applied to the fractionally differenced time series in the second.

However a one step approach would give better parameter estimates. We are currently researching the methods available for implementing a one-step joint estimation framework of the ARIMA (p,d,q) model to the return series, once the ARCH effects are confirmed.

A couple of methodologies are proposed in the scientific literature for a one step estimation of the ARIMA (p,d,q) process. These are based on Bayesian inference for autoregressive fractionally integrated moving average (ARFIMA) models using Markov chain Monte Carlo methods [5] and a pseudo linear method for the estimation of Fractionally Integrated ARMA models [4], but a feasible implementation of these methodologies is yet to be confirmed in the context of our analysis.

IV. Results

A. Initial Analysis

1) Data Transformations

Analysis of the states demonstrated that the distribution of the returns deviates too far from Gaussian and thus, any Box Cox transformation attempt fails. Figure 3.1.1 shows graphical results from before and after the transformation for Arkansas and Table 1.2 shows numerical results.

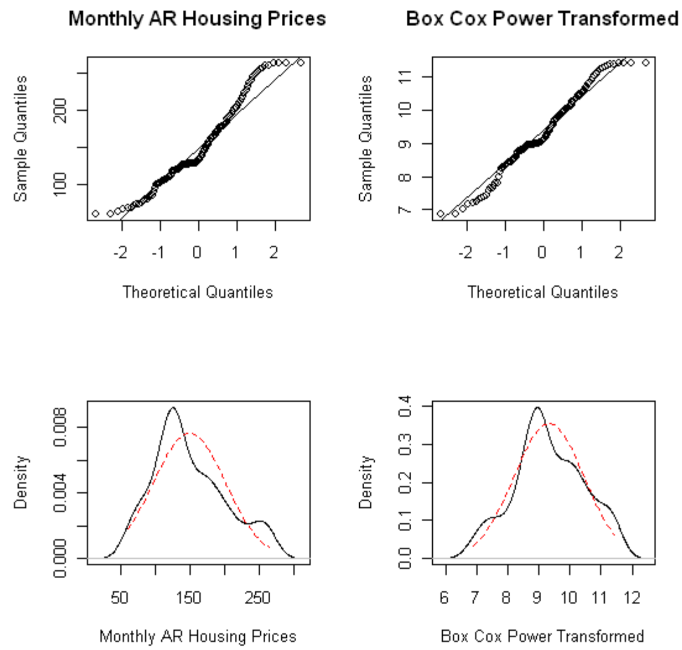


Figure 3.1.1: QQ plots and density plots of the Arkansas HPI data (left) and the Box Cox transformed series with λ=0.23 (right)

p-value	AR	AR Box Cox transformed (λ = 0.23)
Shapiro	1.15E-04	0.022
JB	0.022	0.452
AD	3.82E-05	0.042
CVM	8.61E-05	0.042
KS	2.20E-16	2.20E-16

Table 4.A.1.2: p-values from normality tests for the Arkansas housing price series and the Box Cox transformed series with λ=0.23 (right)

The QQ plots and density plots show significant departures from normality. While the Jarque-Bera test does not reject normality in the transformed series, the p-values of the other normality tests strongly suggest non-normal distributions. The Jarque-Bera tests deviations from normal skewness of 0 and kurtosis of 3. Thus, a symmetric and short-tailed distribution with sample skewness and excess kurtosis coefficients near 0 can incorrectly pass the JB test, as was the case here. It is concluded that a Box Cox transformation is inadequate in making the housing price series Gaussian.

2) I(1) Testing

Results of unit root tests are shown in Table 1. As shown, the p-values all show the rejection of the H_o for the presence of a unit root; thus, these states were determined to be stationary.

	Test Lag	ADF	PP
AK	6	0.01	0.01
AL	6	0.01	0.01
AR	4	0.03	0.01

Table 4.A.2.1: p-values obtained from applying the Augmented Dickey-Fuller and Phillips-Perron tests to a sample of states

3) Normality Testing

134 data points, 1000 simulations

p-value	$N(0,\sigma^2)$	σ(1)	σ(2)	σ(3)
Shapiro	2.5%	0.028	0.023	0.018
	50.0%	0.509	0.491	0.523
	97.5%	0.973	0.979	0.974
JB	2.5%	0.016	0.013	0.017
	50.0%	0.550	0.550	0.568
	97.5%	0.978	0.976	0.969
AD	2.5%	0.021	0.019	0.026
	50.0%	0.498	0.486	0.511
	97.5%	0.964	0.968	0.966
CVM	2.5%	0.018	0.020	0.021
	50.0%	0.486	0.496	0.526
	97.5%	0.968	0.977	0.975
KS	2.5%	0.029	0.021	0.019
	50.0%	0.510	0.503	0.486
	97.5%	0.977	0.979	0.971

Table 4.A.3.1: Normality test p-values for simulated data, 134 data points

662 data points, 1000 simulations

p-value	$N(0,\sigma^2)$	σ(1)	σ(2)	σ(3)
Shapiro	2.5%	0.027	0.022	0.031
	50.0%	0.494	0.483	0.492
	97.5%	0.972	0.964	0.972
JB	2.5%	0.018	0.015	0.026
	50.0%	0.522	0.515	0.529
	97.5%	0.968	0.974	0.972
AD	2.5%	0.031	0.028	0.023
	50.0%	0.506	0.516	0.494
	97.5%	0.968	0.962	0.976
CVM	2.5%	0.022	0.034	0.023
	50.0%	0.513	0.513	0.500
	97.5%	0.981	0.978	0.982
KS	2.5%	0.024	0.027	0.023
	50.0%	0.499	0.499	0.498
	97.5%	0.965	0.965	0.975

Table 4.A.3.2: Normality test p-values for simulated data, 662 data points

Data was simulated from a normal distribution, of both samples sizes 134 and 662, in Table 3.3.1 and 3.3.2, respectively. Normality tests were applied to the simulated data and the p-values were recorded. The simulation was repeated 1,000 times and the empirical quantiles of the p-values were recorded.

The purpose of this exercise is to determine the accuracy of the normality tests. The results were expected, with an approximately uniform distribution of p-values. This will be investigated further before conclusions are drawn.

B. Parameter Estimation for Model Building

1) Fractional Differencing Parameters

Fractionally integrated processes of sample sizes 1000 were generated using fracdiff.sim (**R** package "fracdiff") from the ARIMA(0,d,0) distribution. The 2.5%, 50%, and 97.5% quantiles for the d values are indicated in Table 2. The GPH estimator (**R** package "fracdiff") was used to estimate the fractionally integrated parameter d from the simulated data. Simulations of 1000 were used to construct the empirical quantiles, with the purpose of understanding the accuracy of the GPH estimator.

d = 0.4	2.5%	0.321
	50.0%	0.420
	97.5%	0.659
d = 0.3	2.5%	0.227
	50.0%	0.318
	97.5%	0.555
d=0.1	2.5%	0.016
	50.0%	0.111
	97.5%	0.357
d=-0.2	2.5%	-0.283
	50.0%	-0.185
	97.5%	0.098
d=-0.6	2.5%	-0.632
	50.0%	-0.526
	97.5%	-0.233

Table 4.B.1.1: Quantiles of d-parameter of simulated fractionally integrated processes, determined though the GPH estimator (n=1,000; Burn-In = 1,000; Simulations = 1,000; ARIMA(0,d,0))

Selected Density Plots:

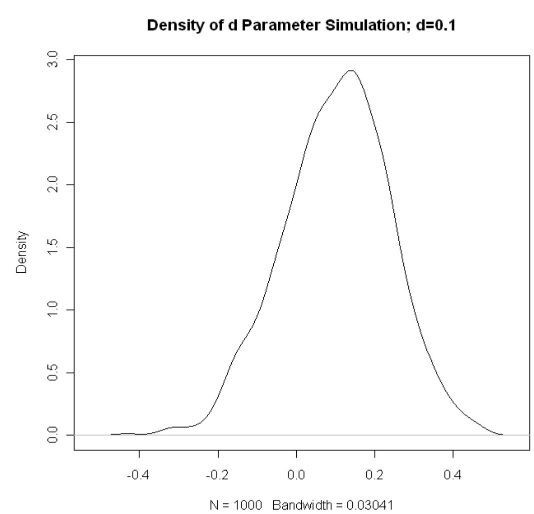


Figure 4.B.1.2: Density of GPH estimates of d Parameter, Simulated data has d=0.1

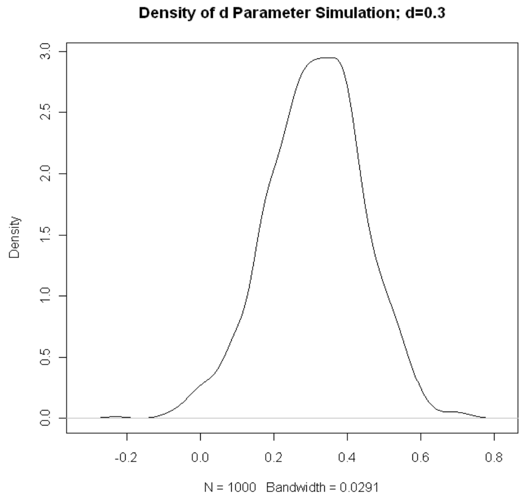


Figure 4.B.1.3: Density of GPH estimates of d Parameter, Simulated data has d=0.3

The density plot indicates that the GPH estimator appears to be a reasonable estimate of the fractionally integrated parameter. This is verified by the relatively small variance of the density plots. The accuracy of the GPH estimator will be investigated further by comparing fractionally integrated parameter estimates to the results from other estimators.

	GPH	Sperio
AK	0.276	0.121
AL	0.215	0.209
AR	0.219	0.201

Table 4.B.1.4: d-parameter estimates using the GPH and Sperio Estimates for a select sample of states

Using the GPH estimator as described above, d-parameters were estimated for the selected states. The results are shown above in Table 3. Additionally, d-parameters were estimated using the Sperio estimator, to gauge the accuracy of GPH. Results using the Sperio estimator confirmed the d estimates for Alabama and Arkansas. However, the results for Alaska verified the need for further investigation.

To supplement these findings, a Monte Carlo simulation has been included, where a random ARFIMA (0,d,0) process is simulated 1000 tines, for a series length of 134, where d is the point estimate derived from the corresponding estimation method. The following plot includes point estimates, in **black**, for the parameter itself, as well as standard errors derived from the asymptotic distributional assumptions of the estimation method. The results of the simulation study are superimposed and shown in **blue**.

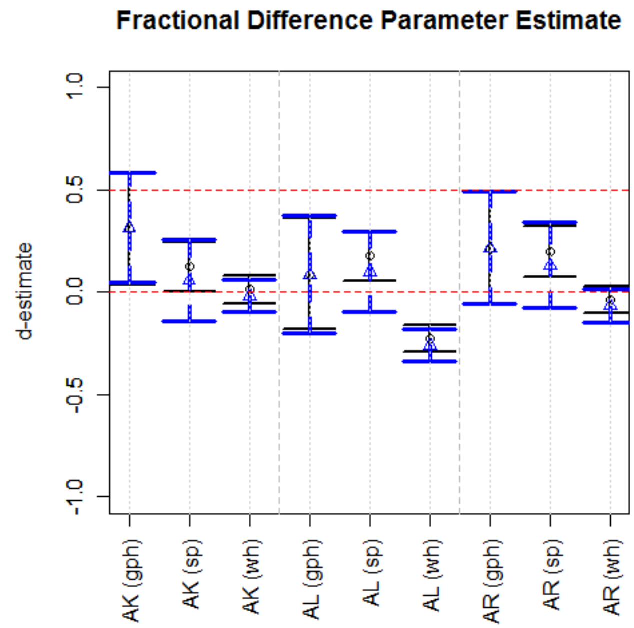


Figure 4.B.1.5. d-parameter estimates for AK, AL, AR.

Additionally, included point estimates were included for the fractionally differenced series, anticipating that any degree of fractional integration should be “filtered-out” by taking a fractional difference, using the previous d parameter estimates.

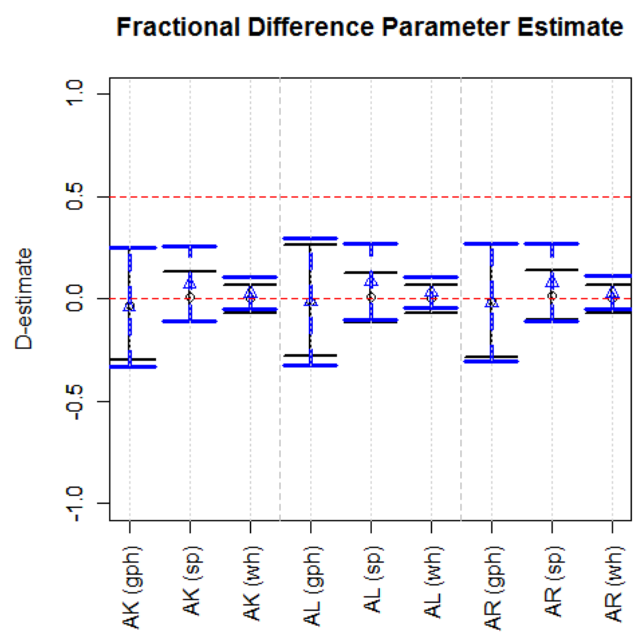


Figure 4.B.1.6. D-parameter estimate for fractionally differenced series.

Here, “D” denotes the fractional difference parameter after fractionally differencing each univariate series with the corresponding estimate of d. As expected, all D’s are close to zero.

It is noteworthy, however, that in the case of the most stationary estimate, the Whittle estimate of AK, taking a fractional difference also produces a D=0. Thus, this method of inferring fractional integration vs. unit root or some other type of integration, and may not be sufficient to estimate a suitable estimate for d as well.

This research shows that Whittle Gaussian MLE (Whittle estimators have the lowest estimation errors from simulation. For a complete picture, the following simulation results show ranges of estimates for simulated time series, with corresponding standard deviation. For this reason, we use the Whittle estimator to infer long memory parameters, which are shown in the model specification section.

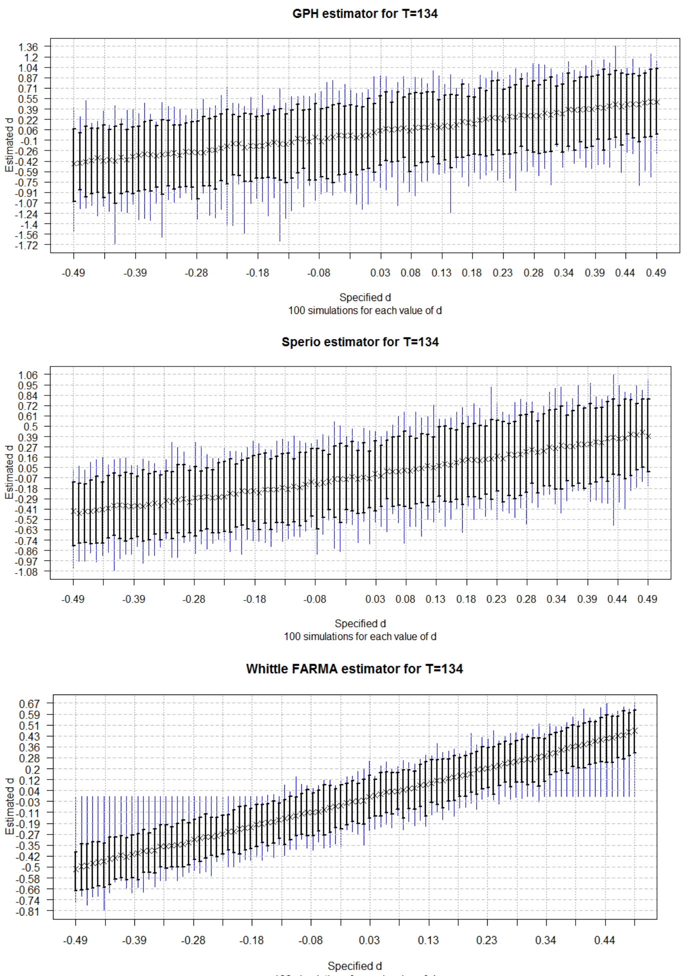


Figure 4.B.1.7. Simulated range for d-parameter estimates, showing smallest estimation error for Whittle Gaussian MLE.

2) ARFIMA-GARCH Model Building

This section shows a typical steps taken for fitting an adequate univariate time series model to OFHEO state simple returns. This sections differs from the preceding analysis, in that we've used interpolated and scrubbed data in our analysis for model building.

Of course, checking for unit-root non-stationary is a necessary first step, but this example ("AK") aims to provide methods for fitting parameter values for a given model. From this analysis, the three states' simple returns from their respective scrubbed, interpolated indices are best fit with an ARFIMA-GARCH model.

The following provides a summary of the analysis used to derive parameter estimates for the AK, AL, and AR models.

i. Investigate the ACF of the simple returns and squared returns.
ii. If there is persistence in the ACF, take a fractional difference using the Whittle Gaussian MLE and work with this series.
iii. Check for adequate removal of long-term dependence by inspecting the ACF of the fractionally differenced series and squared fractionally differenced series. Otherwise, try a different estimate for "d".
iv. Model the mean equation in an ARMA(p,q) framework.
v. Check the ACF of the residuals and squared residuals from iii and check to see if a volatility model is necessary using the Box Ljung test.
vi. Model the conditional mean and volatility equations jointly in a ARMA(p,q) + GARCH(m,s) framework, using (p,q) from iii.
vii. Repeat step iv for this new set of residuals, and proceed to iii if necessary.

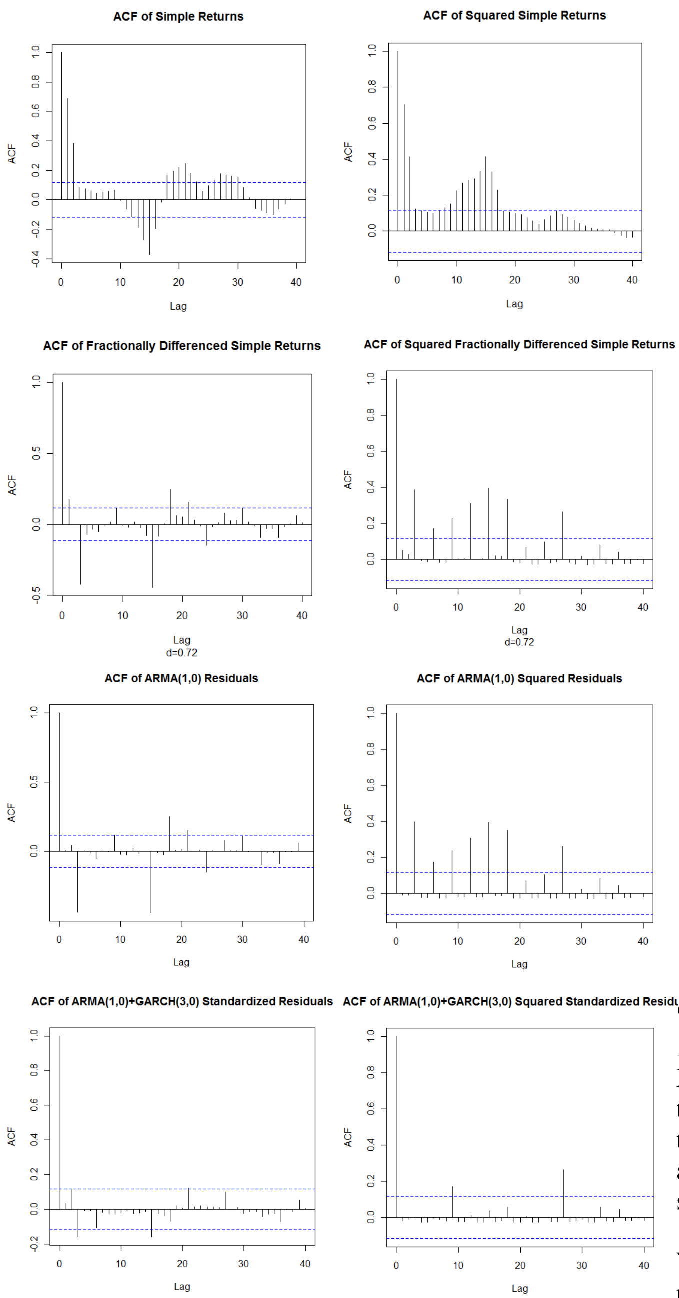


Figure 4.B.2.1: The ACF plots of steps i, iii, v, and vii for AK.

The results of the estimation procedure for all three states' models are summarized as follows, and are of the form:

$$\Phi(B)r_t = \Theta(B)a_t + \phi_0$$

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2$$

where m,s is the order of a GARCH(m,s) model.

State	Interpolated/Scrubbed Data Models				
	Lag	ϕ	θ	α	β
AK d=0.72	0				0.50
	1	0.20			
	2				
	3			1.00	
	Lag	ϕ	θ	α	β
AL d=0.77	0				
	1		-0.21		
	2		-0.13		
	3		0.35	1.00	
	Lag	ϕ	θ	α	β
AR d=0.70	0				
	1		-0.30		
	2		-0.19		
	3		0.34	1.00	

4.B.2.2: Parameter values for fit ARFIMA-GARCH models.

C. Model Checking

1) Temporal Holdout and Box Ljung Test

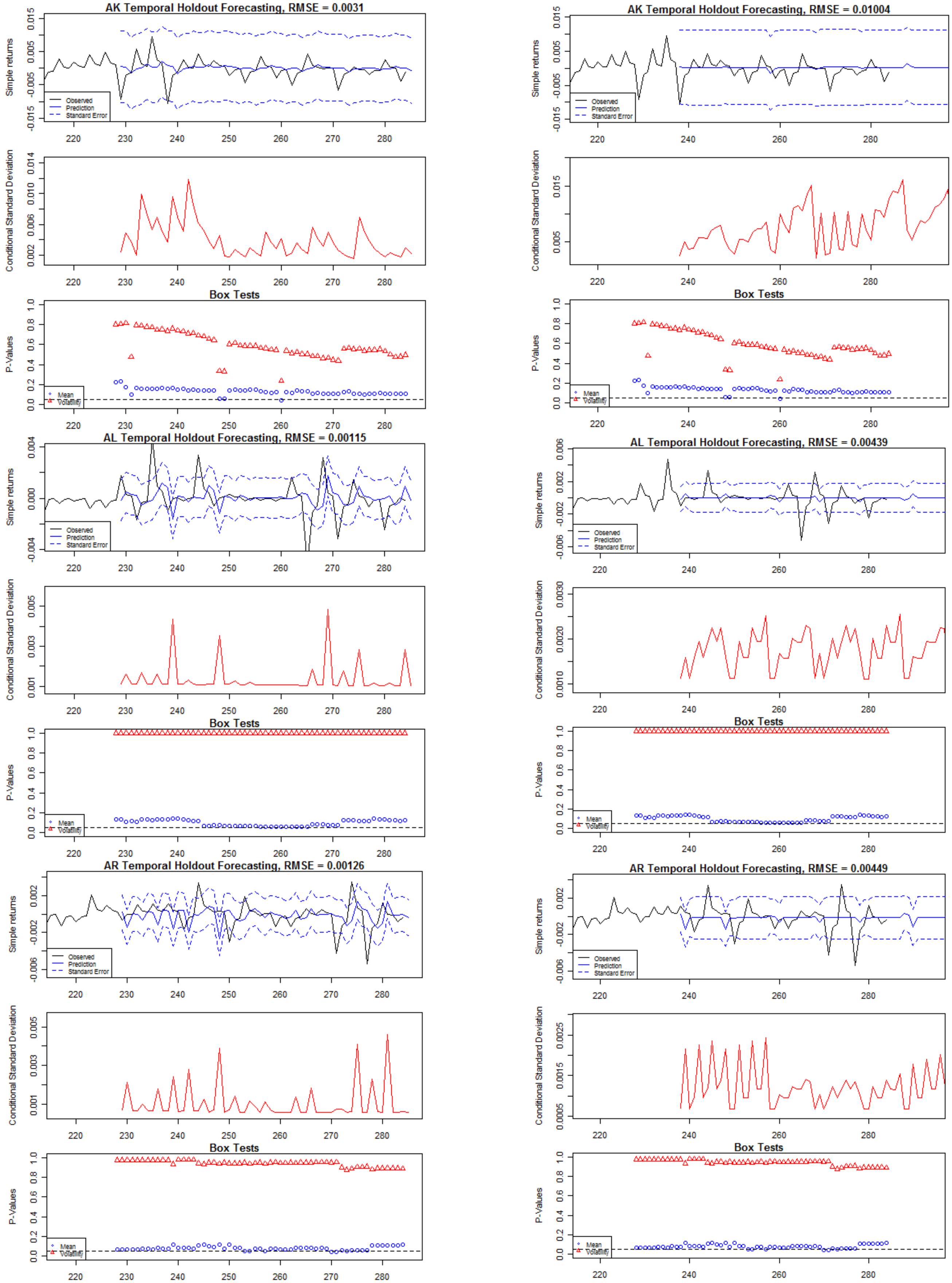
Model checking was performed using temporal holdout, in which the model is partitioned at some initial percentage of the data, and this data is used to estimate a new model to create a forecast. The analysis was repeated for the three states for forecasts of 1 and 10 steps ahead, starting with 80% of the original data.

We can see a significant jump in the RMSE for the two respective prediction intervals, as well as consistent order of RMSE among the 1 and 10 step ahead forecasts, for all three models.

At each point after 80%, the model is re-estimated and a new forecast is derived. The model is checked for adequacy using the Box Ljung test, which tests for the existence of zero serial correlation (white noise) in the following hypothesis:

$$H_o: \rho_i = 0,\ i \in \{1..m\}$$
$$H_a: \rho_i \neq 0,\ i \in \{1..m\}$$

where ρ_i is the autocorrelation at lag i, and m is the total number of lags tested.



4.C.1.1: Temporal holdout forecasting error and Box Ljung Model Verification. Prediction interval 1 on left, 10 on right.l

V. Conclusion

The majority of the states' simple return series were determined to be stationary based on Augmented Dickey-Fuller and Phillips-Perron test results. Furthermore, the simple returns do not follow a Gaussian distribution and a Box Cox transformation is ineffective in making the return series normal. Some of the exhibited non-normality may be a result of ARCH effects; however, this may be remedied by GARCH modeling, which will be further investigated. Simulation shows that the density of the GPH-estimated fractional differencing parameter d appears to be light-tailed and symmetric about its true value.

In analyzing each state's simple returns, interpolated linearly from quarterly to monthly, and cutting-off data before T=40, we observe adequacy of ARFIMA(p,d,q)+GARCH(m,s), for m=3,s=1. Using temporal holdout and Box Jenkins, we can approximate the adequacy and performance of these models over time

References

[1] Beran, Jan. *Statistics for Long Memory Processes.* p.144.

[2] Dolando, J. "A Fractional Dickey Fuller Test for Unit Roots."

[3] Sephton, P. "Critical Values for the Augmented Factional Dickey Fuller Test".

[4] G.N. Fouskitakis and S.D.Fassois. "A Pseudo Linear Method for Fractionally Integrated Arma (ARFIMA) Model Estimation."

[5] Jeffrey S. Pai, Nalini Ravishanker. "A Pseudo Linear Method for Fractionally Integrated Arma (ARFIMA) Model Estimation."

6 Zazli Chik Pusat Asasi Sains, "Performance of Order Selection Criteria for Short Time Series"

7 Yuhong Yang, "Can the Strengths of AIC and BIC Be Shared"