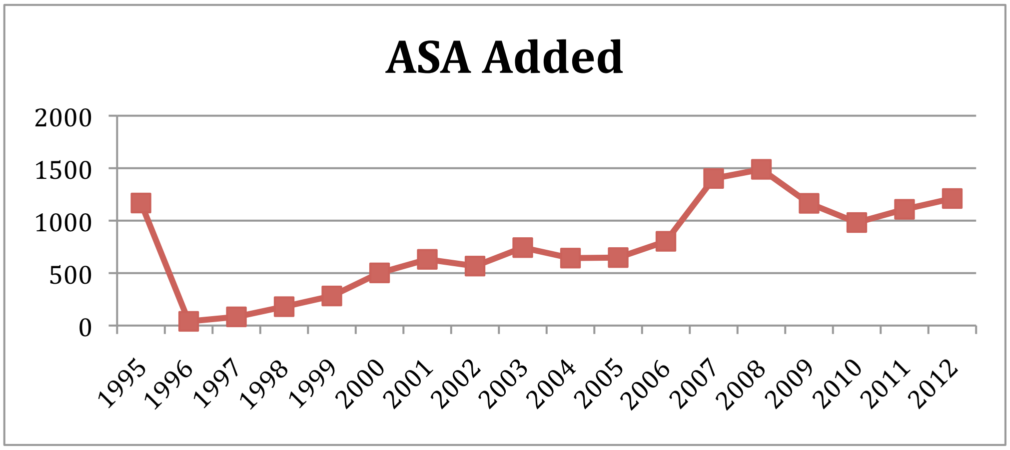
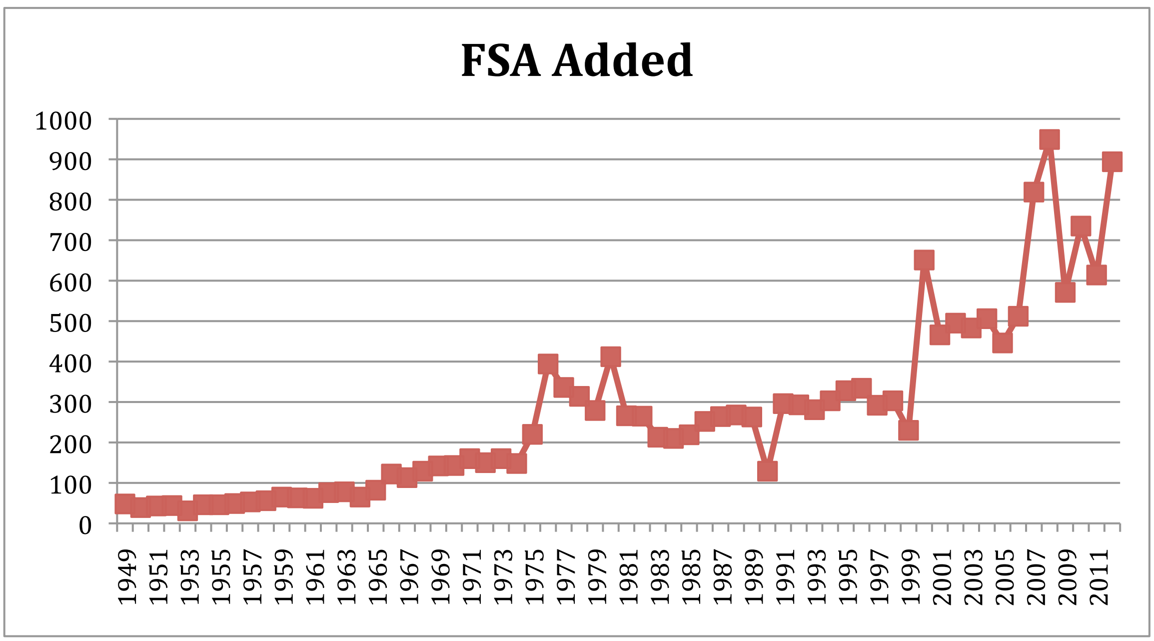
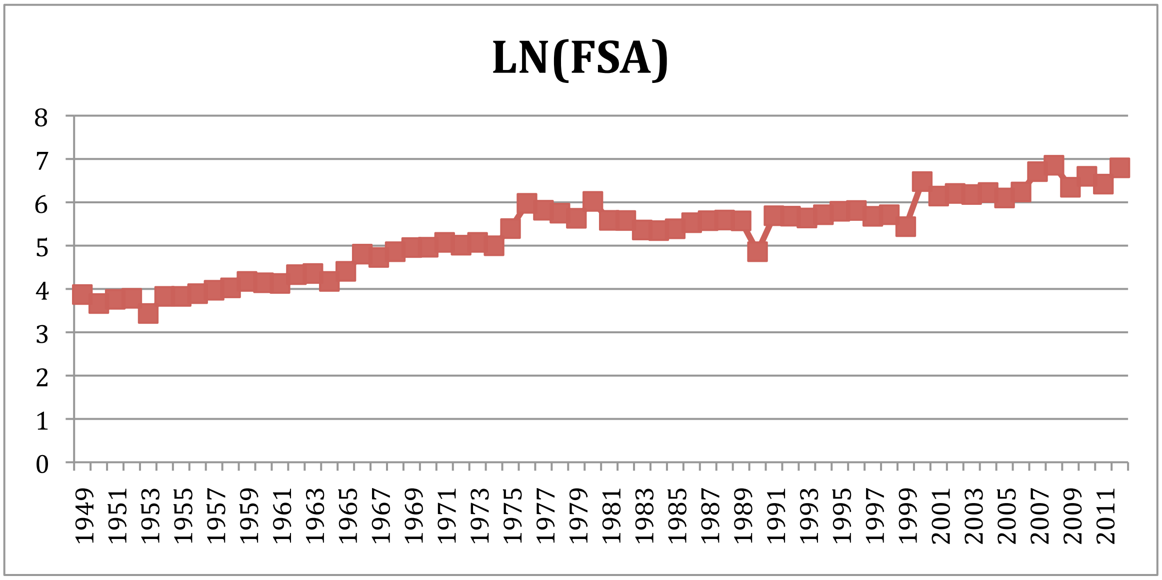
Students of time series analysis may be particularly interested in the number of new actuaries added to the Society of Actuaries every year. While the number of new ASAs added may seem more relevant for this course, there are more years of data available for the number of new FSAs added. This data is provided on the SOA website: http://www.soa.org/About/Membership/about-members-by-year.aspx.

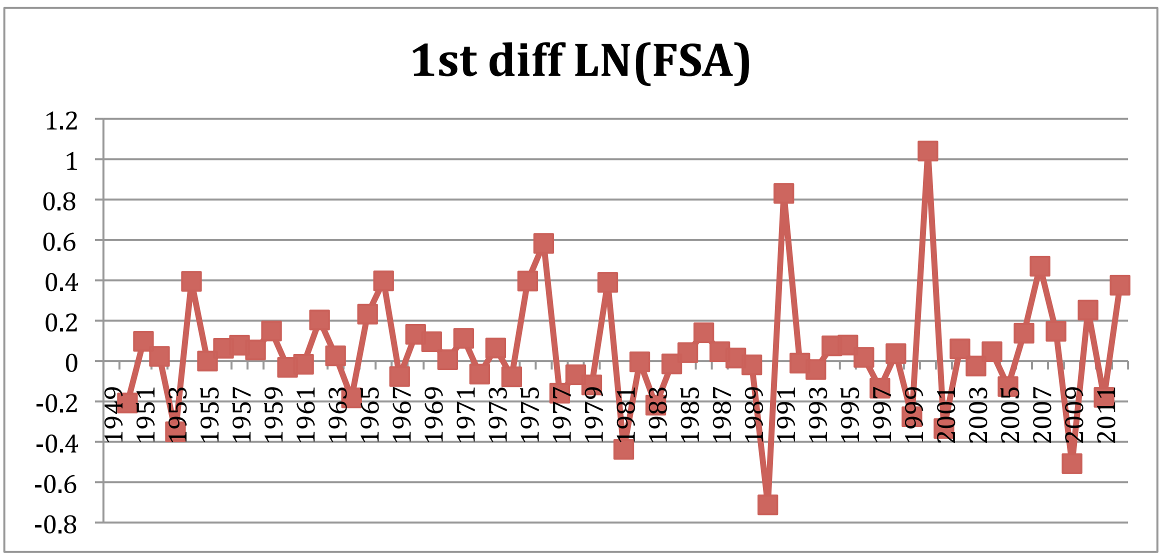




Since the variance for new FSAs added each year seems to increase as the level of the series increases, and since the level of the series does appear to be increasing exponentially, we transform the data with the natural log.

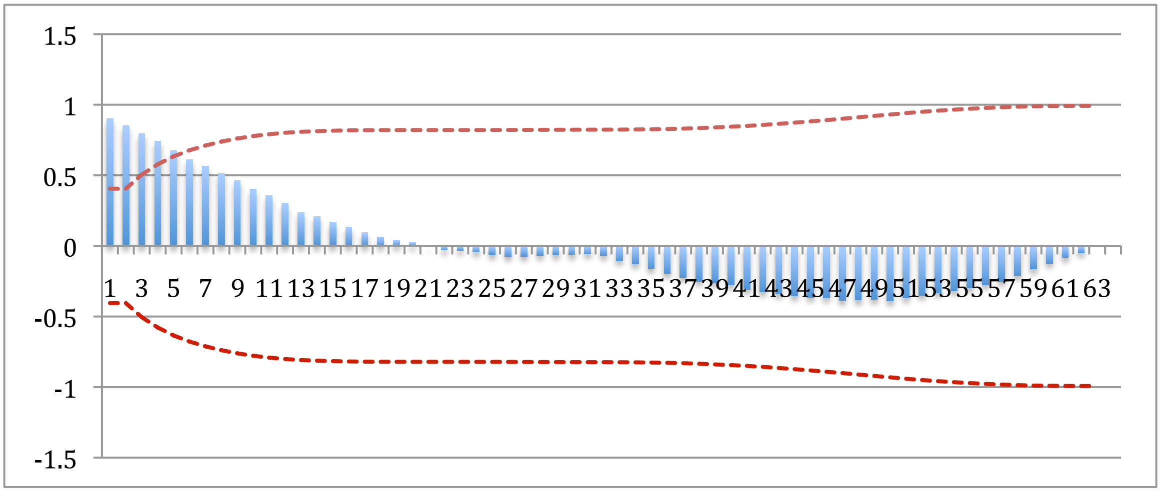


This is much more uniform although it's still increasing. The first difference of the natural log looks better but we may not need to go that far.



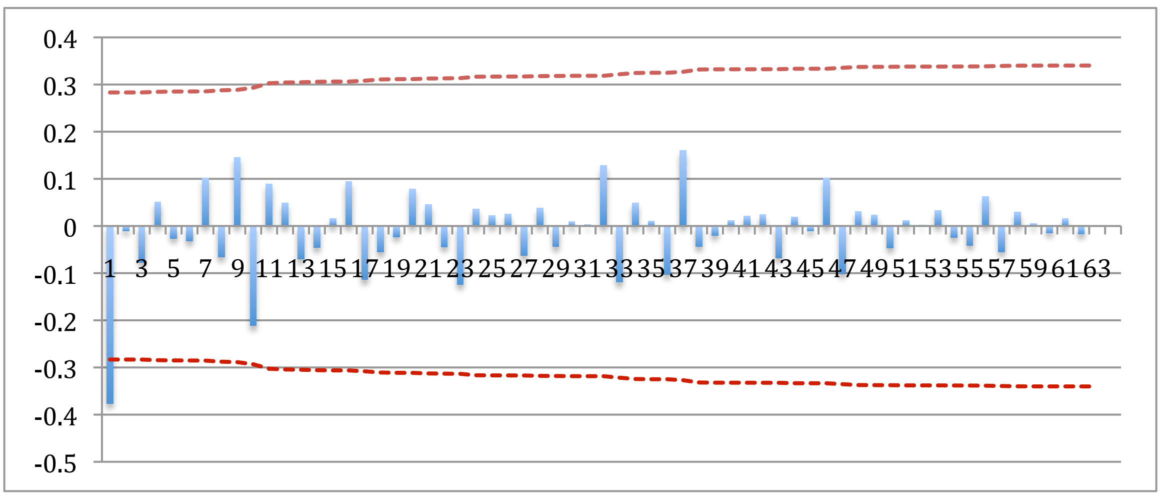
To choose the appropriate model we must look at the sample autocorrelation (ACF) and the sample partial autocorrelation (PACF). From here on out, analysis is based on the data transformed by the natural log.

The following chart shows the ACF as well as plus and minus two standard errors based on variance formula 6.1.11 in the Cryer and Chan textbook: $(1/n)*[1+2\ \Sigma\ r_j^2]$, summing r_j^2 from j=1 to j=k-1. This standard error is appropriate if the assumption is that the series is a moving average.

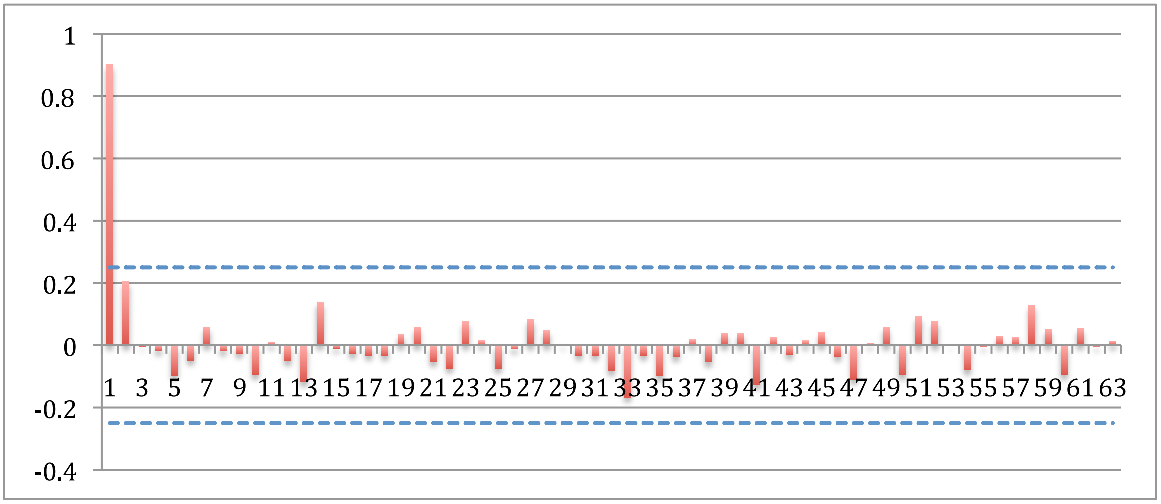


Although the ACF values fall within the error lines by lag 6, they decrease slowly and do not drop to zero. It seems likely that the values could continue sinusoidally for some time. This graph indicates that the series is not based on a moving average formula, or at least not solely on a moving average formula.

The ACF for the first difference of the log-transformed data looks more appropriate for a moving average model with order 1. But again, this may not be an appropriate choice due to over-differencing.



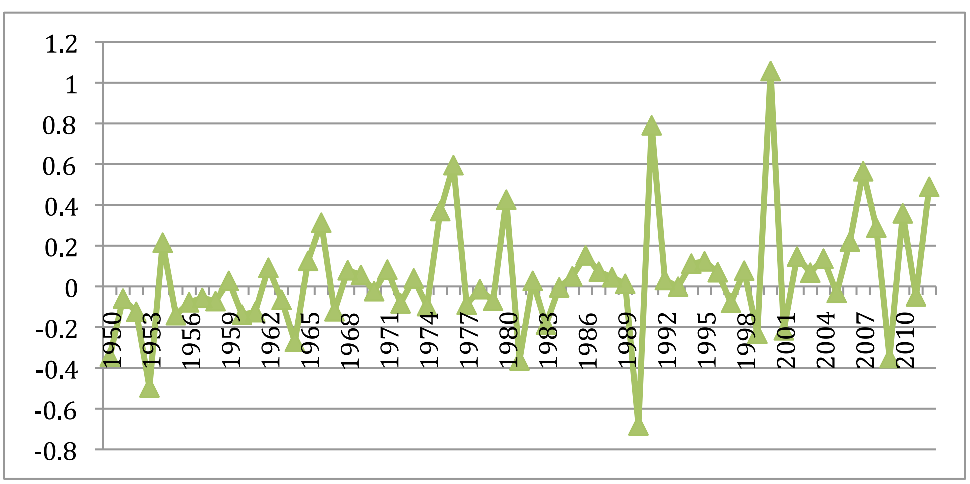
The graph below for the PACF shows critical value lines at plus and minus $2/\sqrt{n}$. This seems to indicate pretty clearly that an autoregressive model is appropriate. Given how clearly significant lag one is, a first difference of the data seems unnecessary.



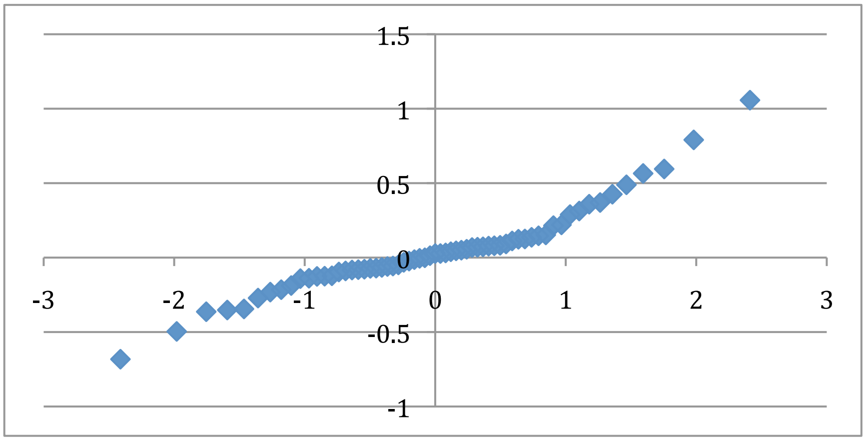
To estimate the model parameters we use the method of moments. Since we have determined that the time series can be modeled with a AR(1), we estimate Φ with r_1 which is 0.9018. To estimate the noise variance, σ_e^2, we start by estimating the process variance with the sample variance: $\gamma_0 \approx s^2 = (1/(n-1)) \Sigma (Y_t - \overline{Y})^2 = 0.7937$. The noise variance for an AR(1) model is then estimated as $(1-r_1^2)s^2$ which is 0.1482.

Using least squares estimation, we first estimate μ as $\overline{Y}$, or 5.2667. Then the estimate for Φ is different from r_1 by one term in the denominator. For the new actuary data we calculate 0.8377. We can also calculate θ_0 as $\mu * (1 - \Phi)$ which is 0.8546.

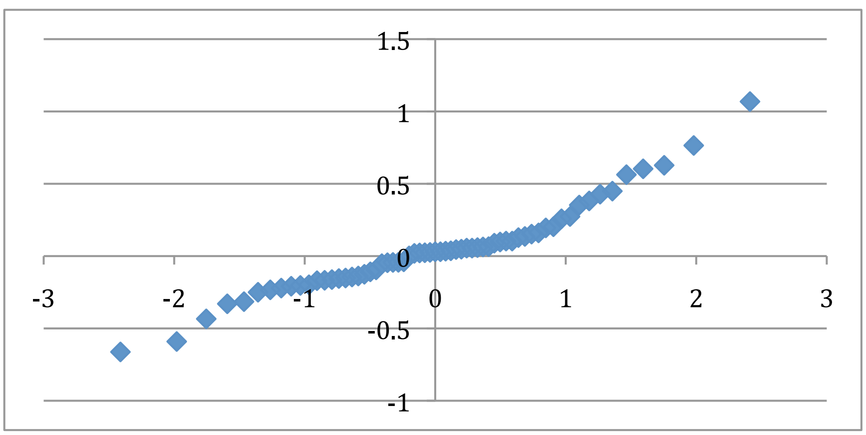
For model diagnostics we first look at the model produced by the method of moments. We will also include a θ_0 term here although the textbook does not explicitly discuss non-zero averages for method of moments estimations. For a Φ value of 0.9018, $\theta_0 = 0.5172$. The residuals are defied as actual values minus predicted: $\hat{e}_t = Y_t - \Phi Y_{t-1} - \theta_0 = Y_t - 0.9018Y_{t-1} - 0.5172$.



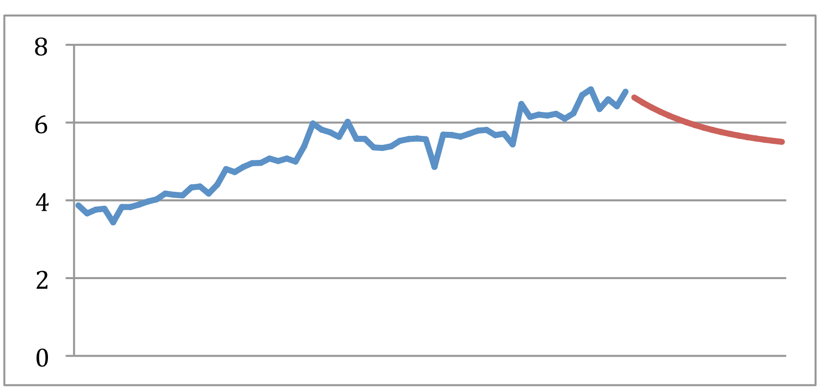
It doesn't look particularly normal but it's not horribly off. There does seems to be a trend of negative residuals in early years, increasing throughout the data to mostly positive residuals later. A quantile-quantile plot also helps to assets the normality. There seem to be long tails, but the majority of the residuals fit the normal curve.



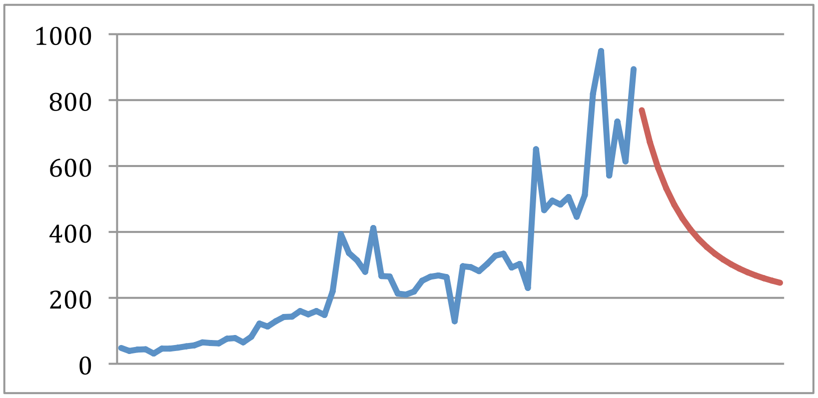
The quantile-quantile plot is not better for the parameters based on the least squares estimation.



The forecast of the model highlights the issue. It doesn't make sense for the number of new actuaries to drop down to the average of all years from 1949 through 2012.

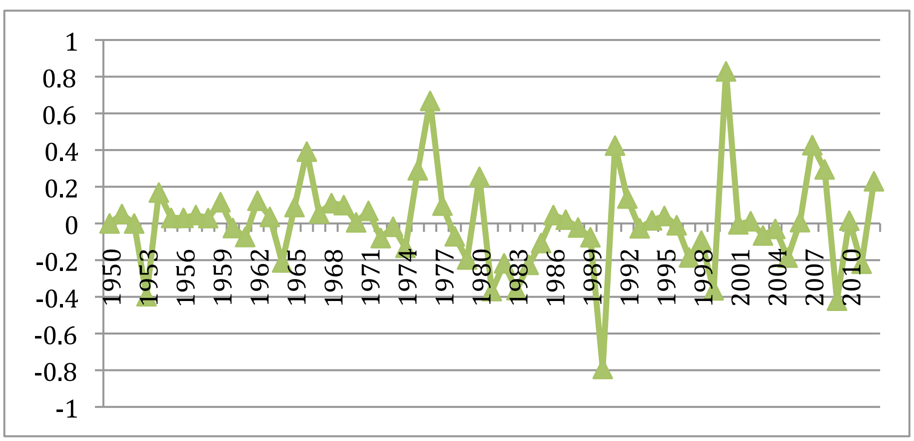


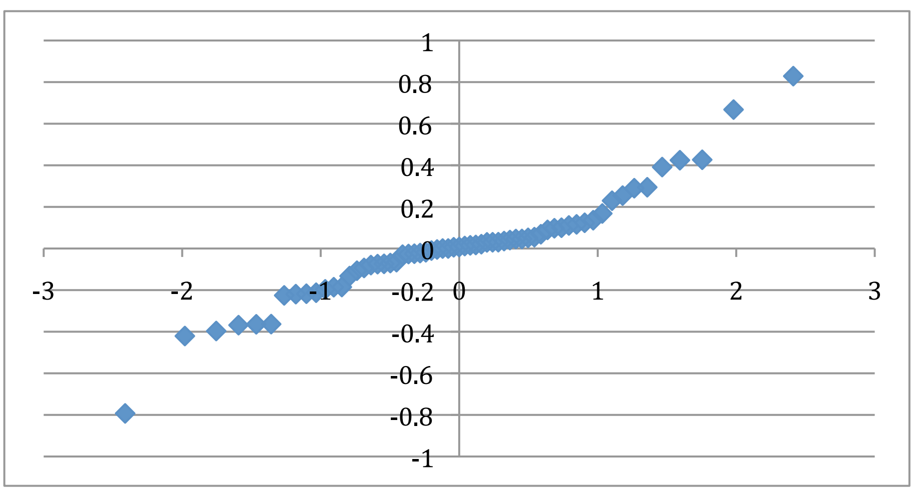
Transforming the data back to before the log adjustment shows the same thing.



It seems necessary to revisit the integrated moving average model. The method of moments for a first order moving average models uses $\rho_1 = -(\theta/(1+\theta^2))$. We have $\rho_1 = -0.3762$, which results in θ of either 2.2043 or 0.4537 and only the later satisfies the invertibility condition. There is also a non-zero average of 0.0464 so the time series model is $\Delta Y_t = 0.0464 + e_t - \theta e_{t-1} = 0.0464 + e_t - 0.4537e_{t-1}$.

Residuals for this model are more symmetrical although the q-q plot isn't perfect.





The forecast for both the transformed and pre-transformation data also appears a lot more useful. It may be a little optimistic beyond the first few years. But for the near future, this model produces a much better forecast than the one indicating that the count of future new FSAs will decrease. This is positive news for all actuarial students.