Lab 3 Michael Anshelevich
October 26, 1997

S 248

Since I do not have a data set I am interested in analyzing, I will look at the most general time series I know of: Wilshire 5000 stock index. The advantage is that nobody else has a good model for it either.

The data set contains the daily values of the index from December 3, 1979 to October 11, 1996 (Figure 1). The data are given only for the trading days. However, as a first approximation we consider the series just as a function of the index, without taking weekends and holidays into account (Figure 1). The series has a long-term trend to grow. Also, a number of shocks are evident. The first difference is plotted in Figure 2. The series is clearly non-stationary, and shocks are apparent. Taking logarithms of the differenced data produces a series which is, except for the shocks, more stationary (Figure 2).

Since the index is in fact a multiplicative rather than an additive random process, the first thing to do is to take a logarithm of the series (Figure 3). The series has a trend which is amazingly close to being linear (Figure 3). Subtracting the linear fit, we obtain Figure 4. The first difference of the result is plotted in Figure 4. This series is significantly more stationary, again except for the shocks. Now, at the scale at which the shocks occur (decades) the series is reasonably short (17 years). Thus we consider the shocks as anomalies and remove the largest of them, corresponding to dates (± 1) October '87 (15, 16, 20, 23), January '88 (7), October '89 (12). We substitute for the values the mean of the series. The remaining series appears stationary (Figure 5). Its ACF, spectrum and PACF are plotted in Figures 5,6. The series is clearly not a white noise, thus the original series is not modeled well by multiplicative Brownian motion. We try to fit an ARMA model to the series. The plots suggest an AR(1) model. Fitting an AR model by AIC gives AR(1). Adding an MA component does not improve the fit. The diagnostics for ARMA(1,1) model, the residuals for AR(1) model, their ACF and PACF are plotted in Figures 7,8. For comparison, we simulate a series of 4000 iid normals, with the same variance as the estimated one for the residuals. This series, its ACF and PACF are plotted in Figures 9,10. The results are very similar to Figures 7,8. Also, the portmanteau test of whiteness (Figure 10) gives sufficiently high P-value. We conclude that the residual series can be considered as white noise.

We can also try to find periodicities in the series. Oversmoothing and overtapering (spans=(95, 105), taper at 50%) we obtain Figure 11. There appear to be components of periods 5 and 10. Since in the series the week is 5 days long, there is a weekly effect. The effect is not strong enough to warrant differencing (Figure 11). Nor does fitting a seasonal ARMA model of a small order improve the fit much (Figure 12). Thus we will not include the weekly effect in the model.

Since the true objects of interest are the shocks, which are excluded from the model here, the forecasting of the series would essentially give the linear trend, and thus does not tell us anything new.

Conclusions: the first difference of the logarithm of the series has a clear linear trend, and the residual series, after excluding a small number of events that lie outside the scope of this model, appears stationary gaussian. It is reasonably well modeled by a first-order autoregression process, with coefficient 0.14 and estimated variance of $6 \cdot 10^{-5}$. There is, of course, no reason why this model should in fact be true. There is also a suggestion of a small weekly effect.

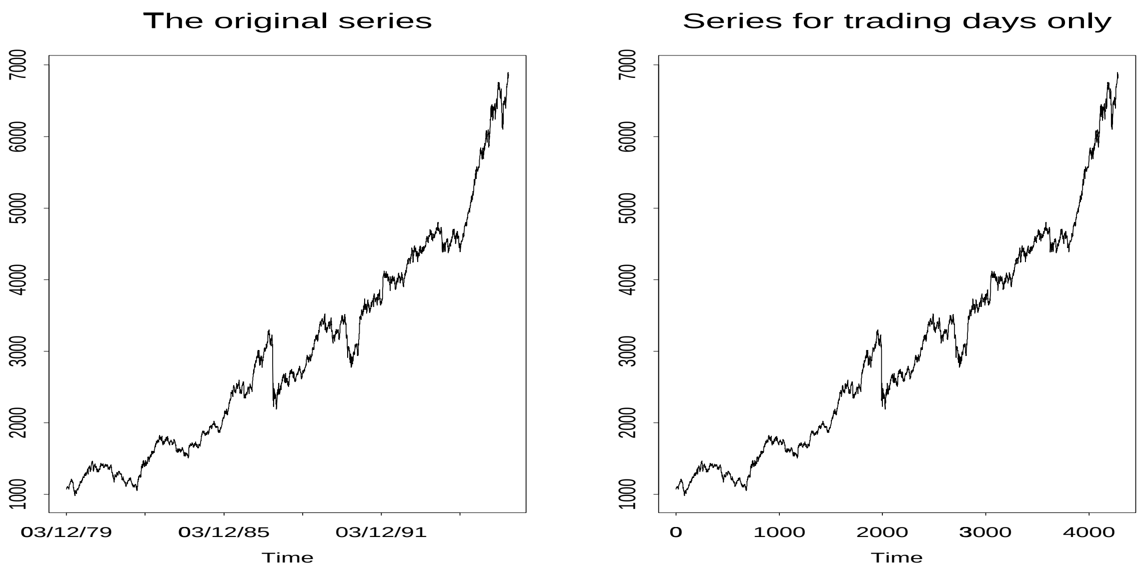


Figure 1: Wilshire 5000 stock index

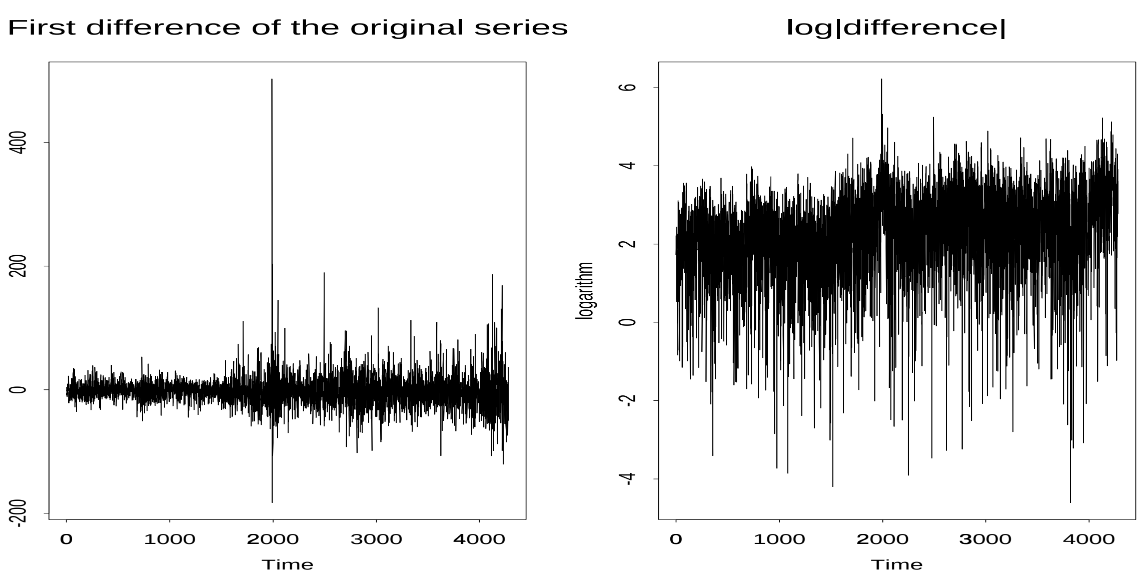


Figure 2: The first difference of the original series and logarithm of its absolute value

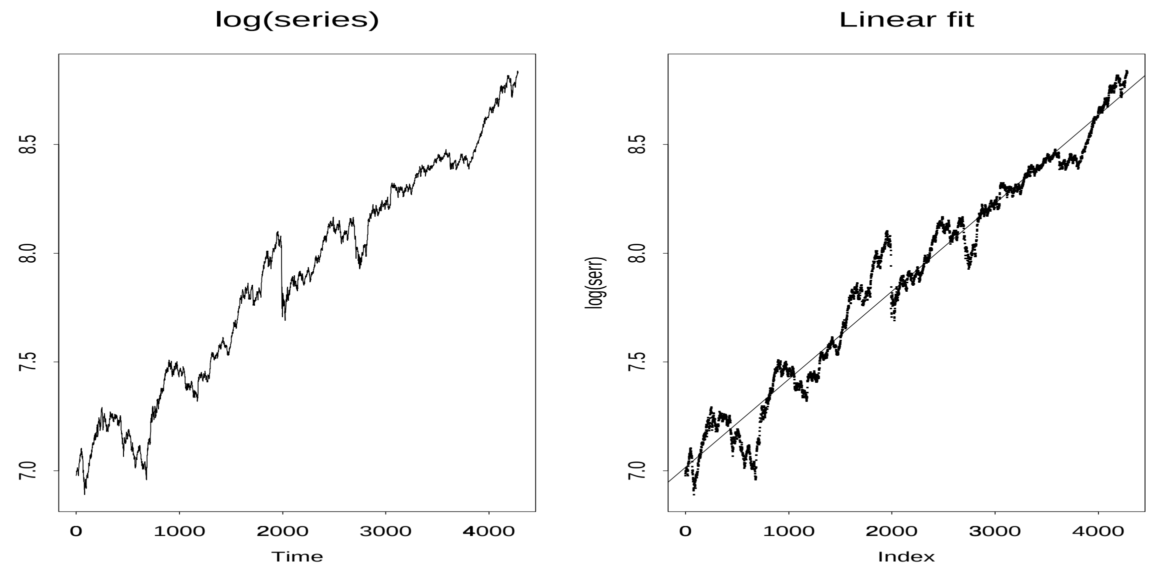


Figure 3: The logarithm of the series and the linear fit to it.

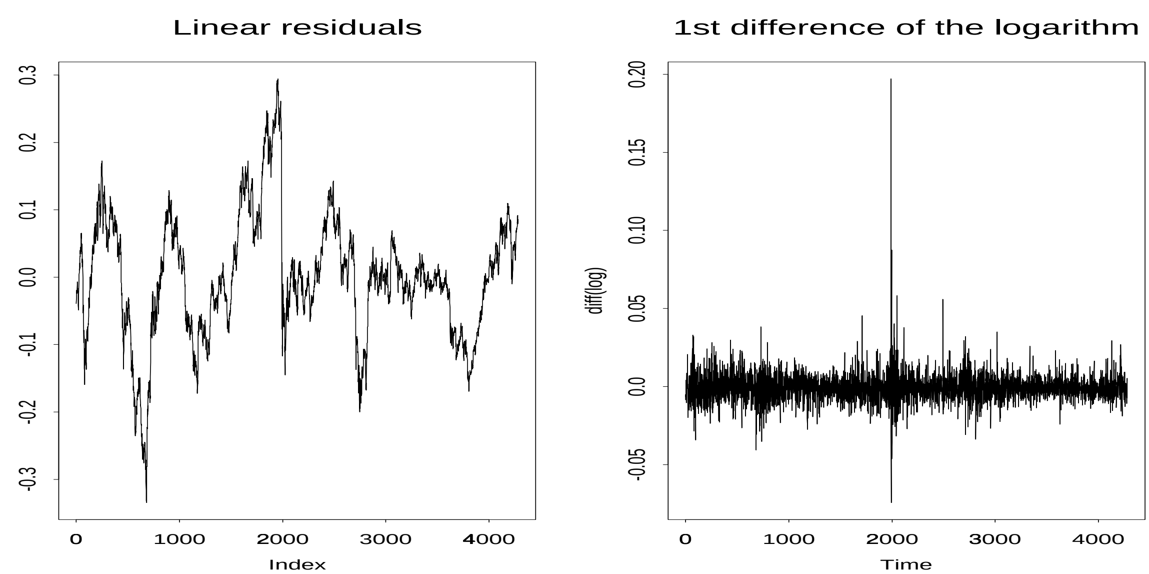
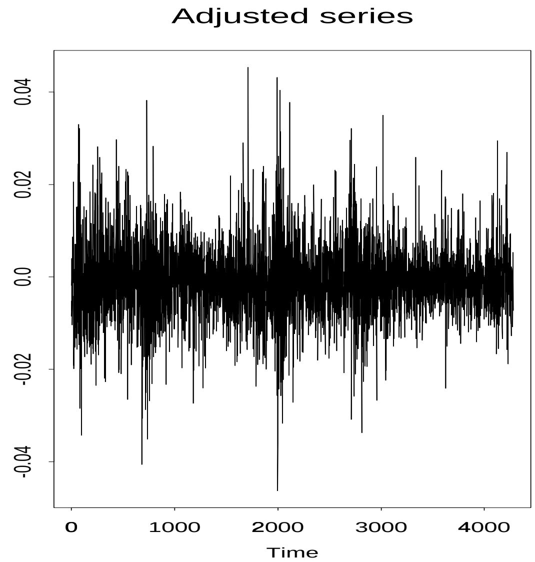


Figure 4: Linear residuals and their first difference.



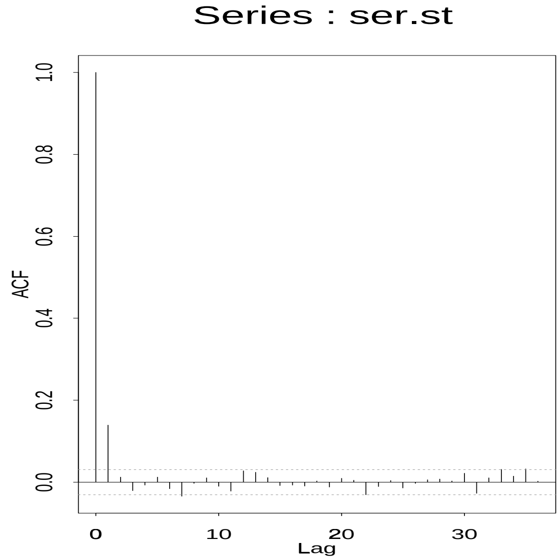
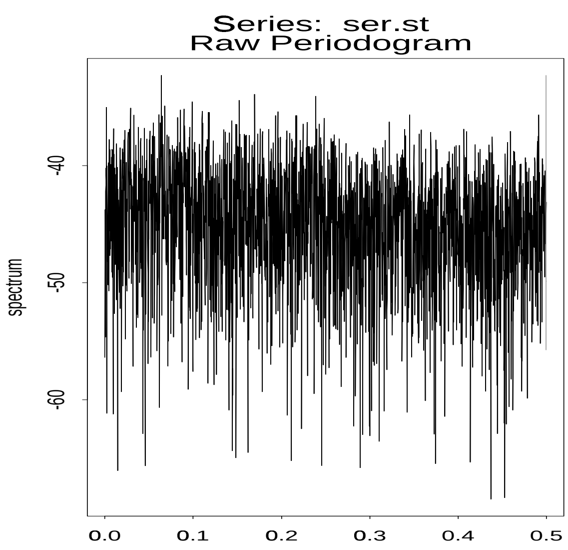


Figure 5: The series we will be working with: first difference of the logarithm of the data with shocks removed, and its ACF.



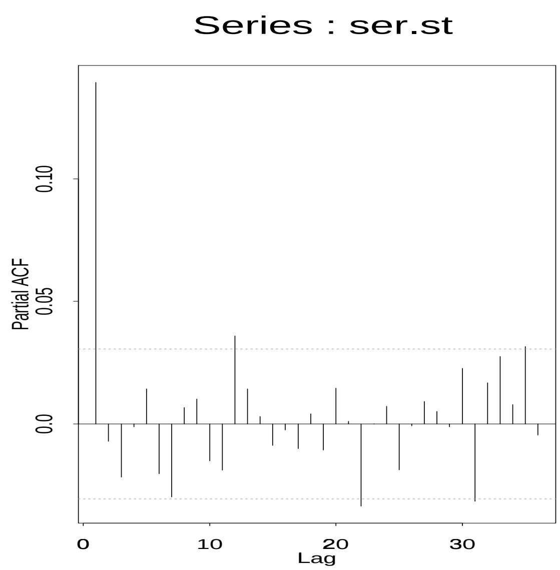


Figure 6: The spectrum and partial ACF of the series

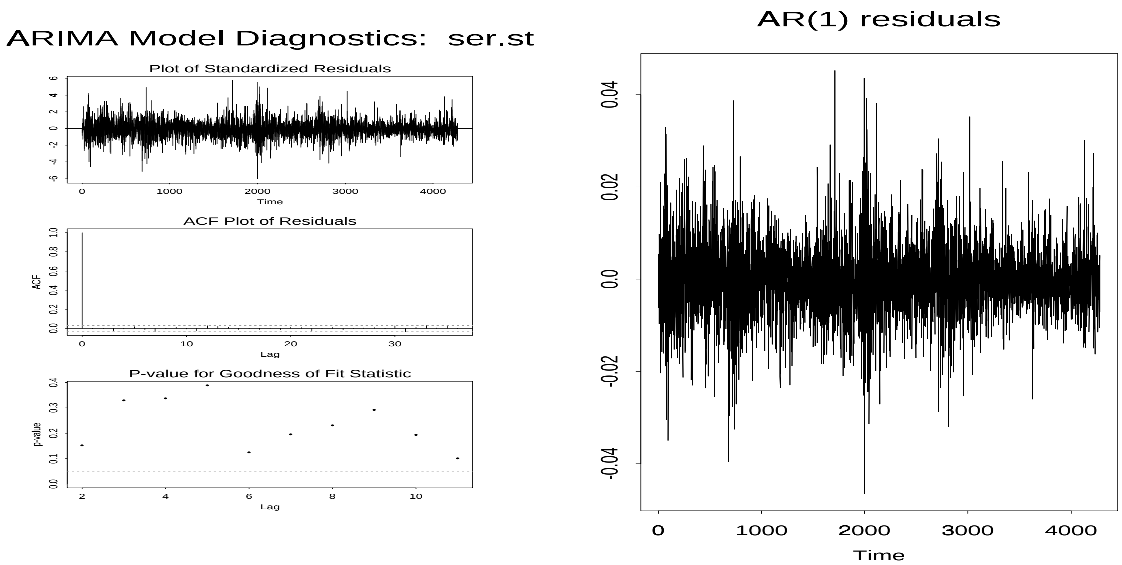


Figure 7: The diagnostics for ARMA(1,1) model and the residuals for AR(1) model.

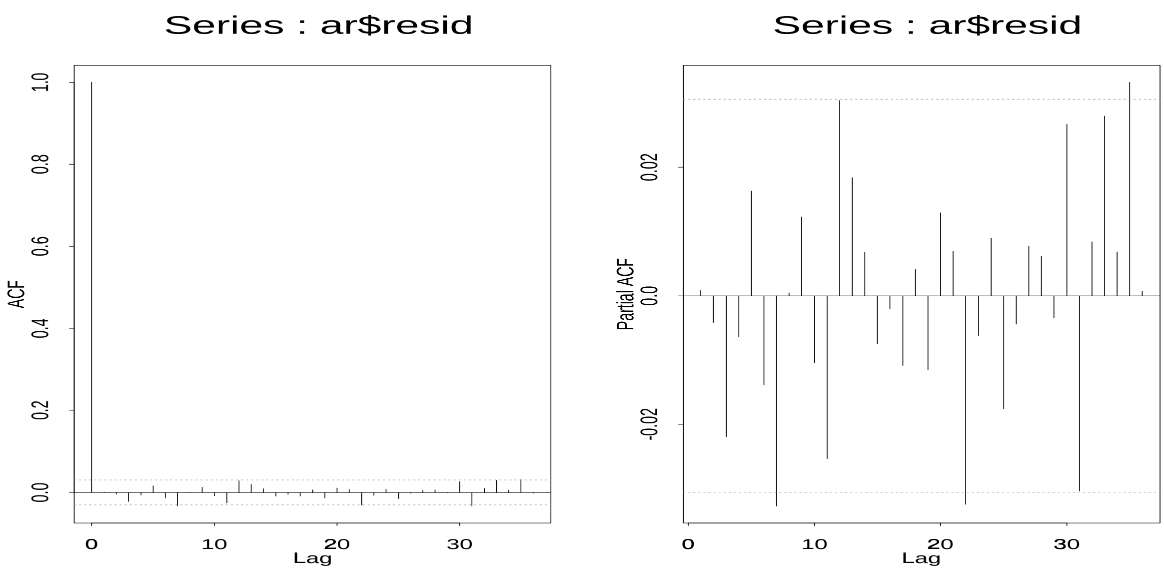


Figure 8: ACF and partial ACF for AR(1) residuals.

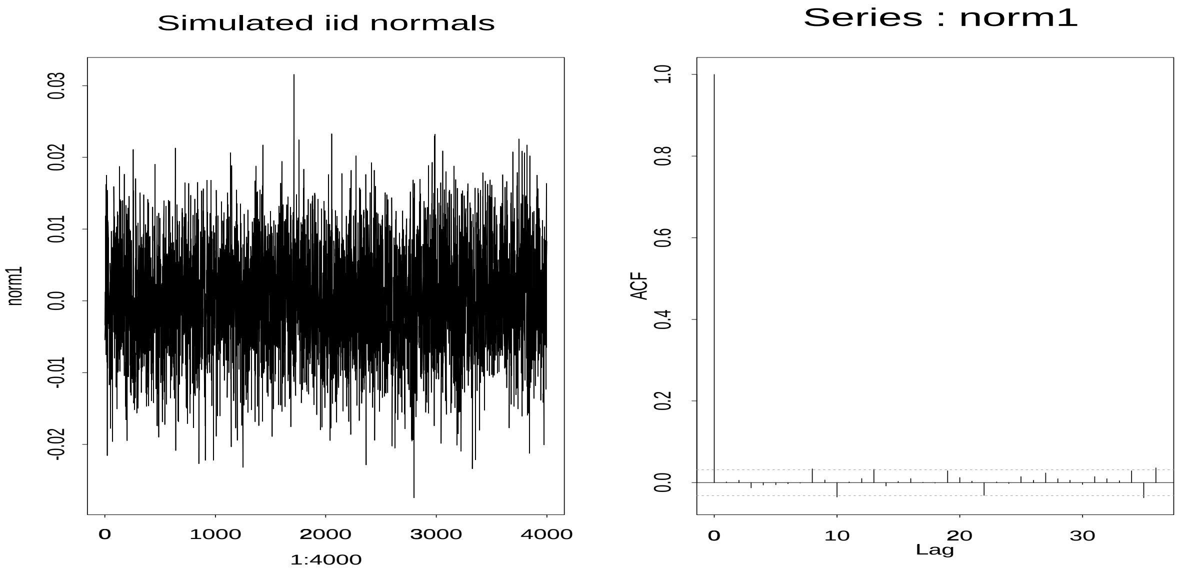


Figure 9: A simulated series of iid normals and its ACF

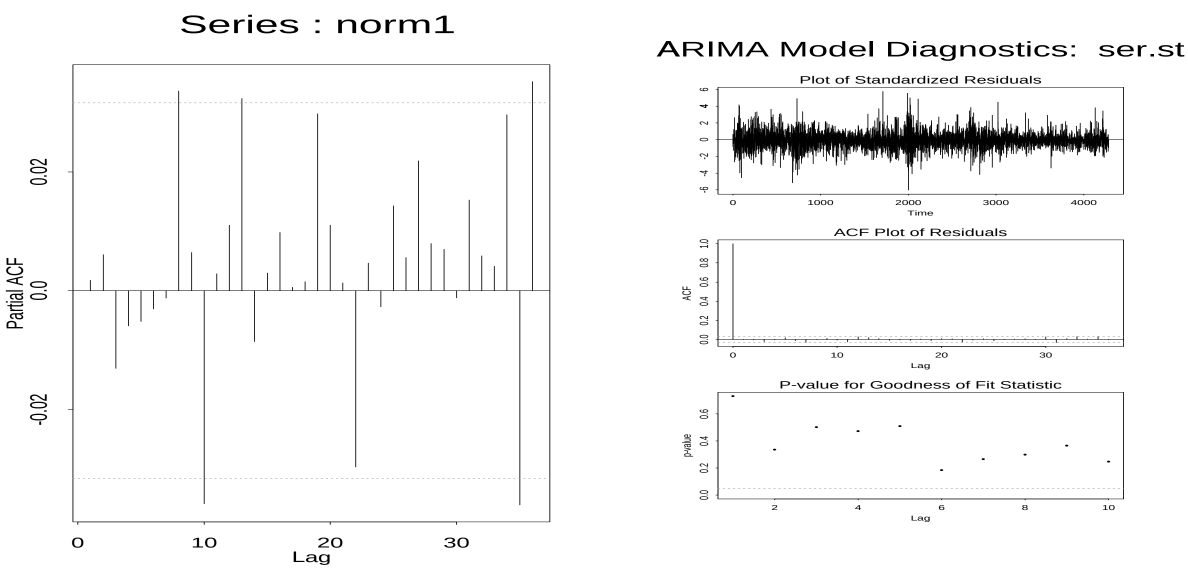


Figure 10: Partial ACF for the simulated normal series, and AR(1) diagnostics.

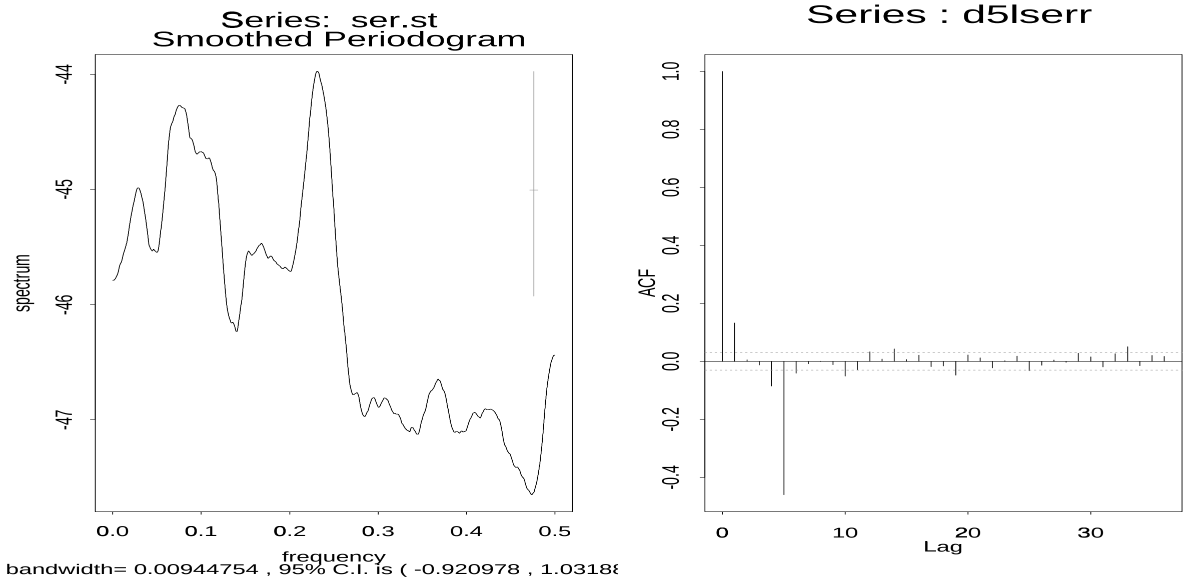


Figure 11: Smoothed tapered spectrum of the series and the ACF of the 5-day-differenced series.

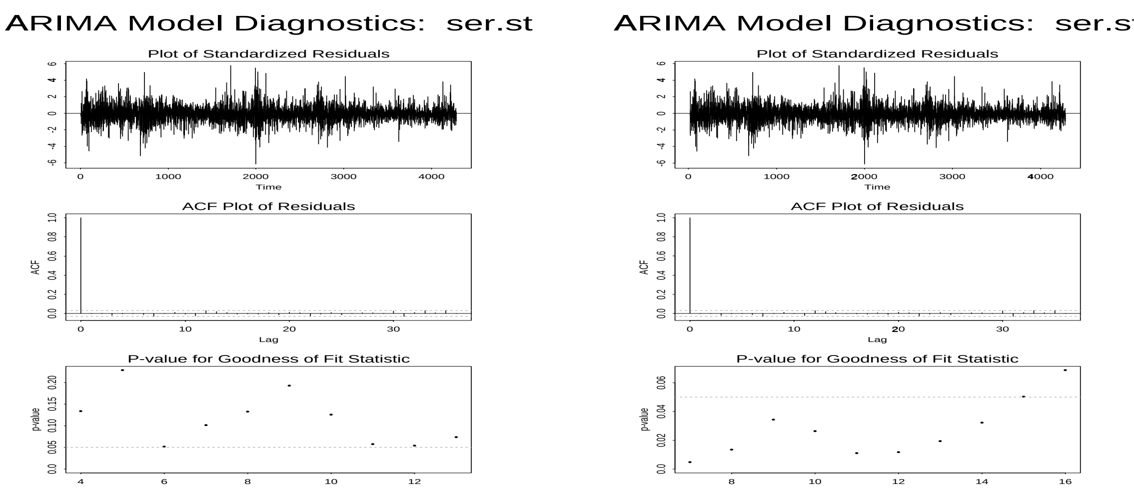


Figure 12: Diagnostics for ARMA(1,1)×ARMA(1,1) and ARMA(1,0)×ARMA(3,3)