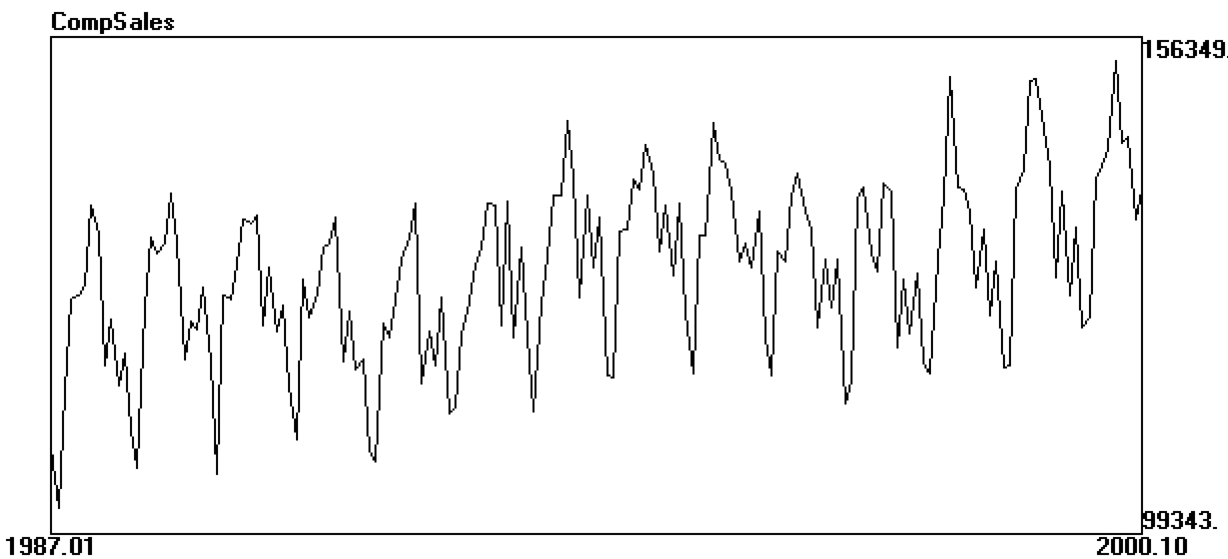
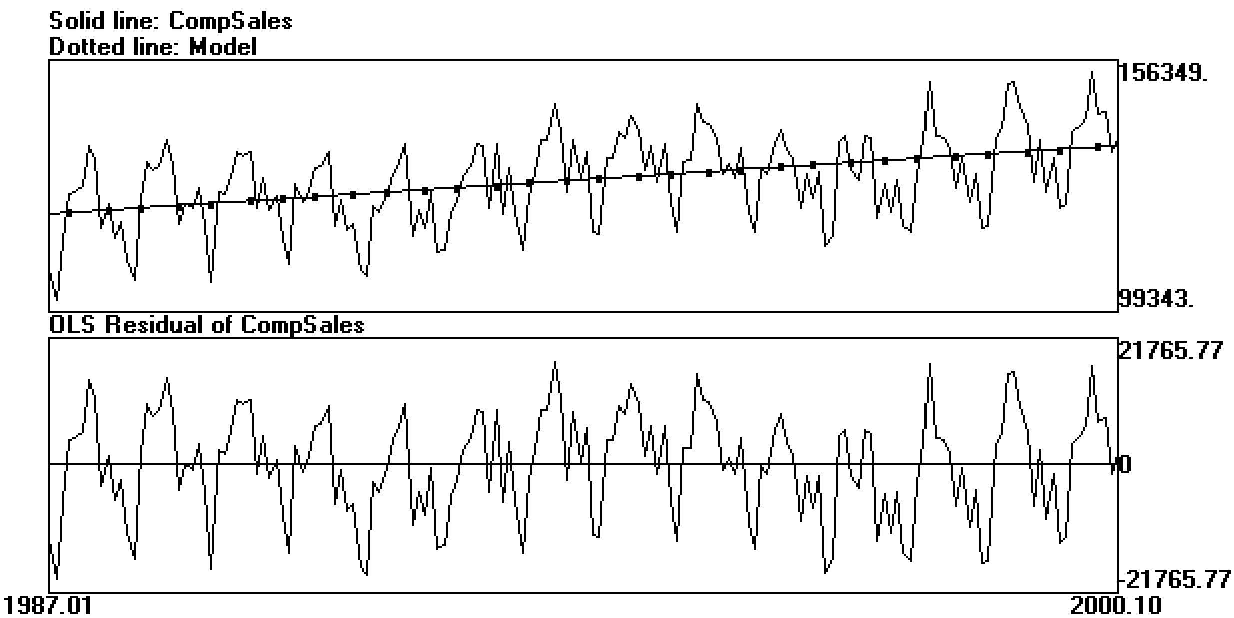
Time Series Forecasting

Example: Comp Store Sales



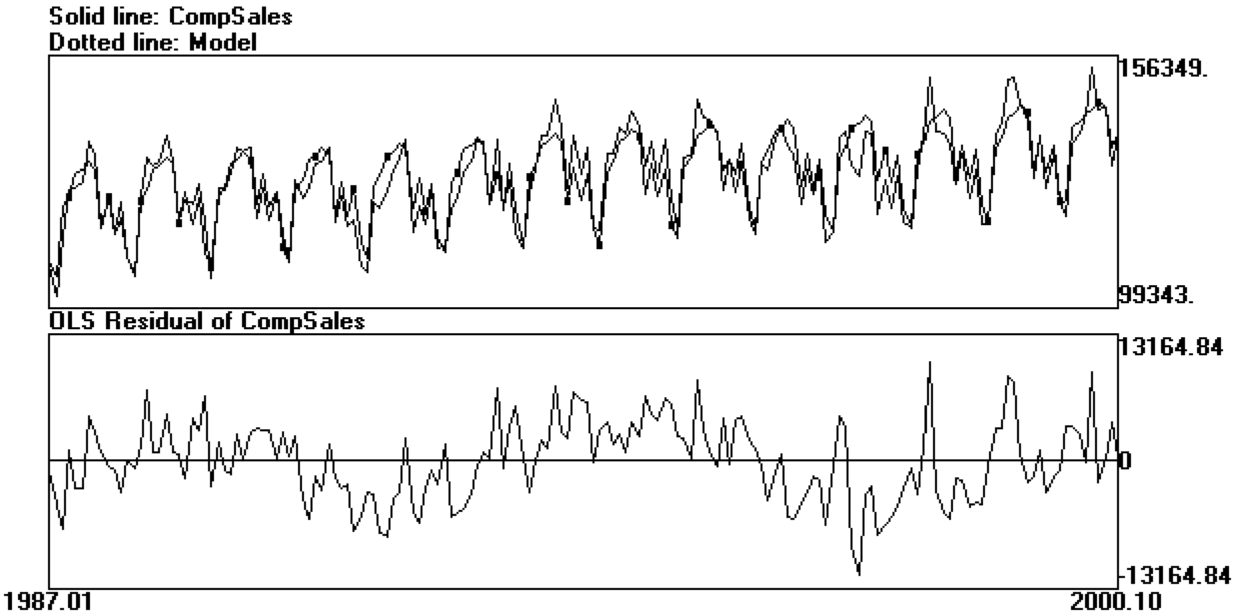
Clearly there is an upward trend

Remove a linear trend by regression on a constant and "t"



The linear trend is significant (t-stat about 70). A quadratic term is insignificant (t about 0.85).

There appears to be seasonality.



The seasonal dummies are jointly significant (F(11,153) = 50).

There is still clear structure in the residuals.

Some principles of time series analysis:

1. The Autocorrelation Function for a time series y_t the auto-correlation function $\theta(k)$ is the correlation between y_t and y_{t+k} the "correlation at lag k"

2. The Partial Autocorrelation Function, p(k) is the correlation between y_t and y_{t+k} controlling for all the y's between y_t and y_{t+k}. This is more like a regression coefficient than a simple pairwise correlation.

Both are graphed as functions of k.

Models for time series - AR and MA

AR(1): $y_t = \alpha\, y_{t-1} + \varepsilon_t$

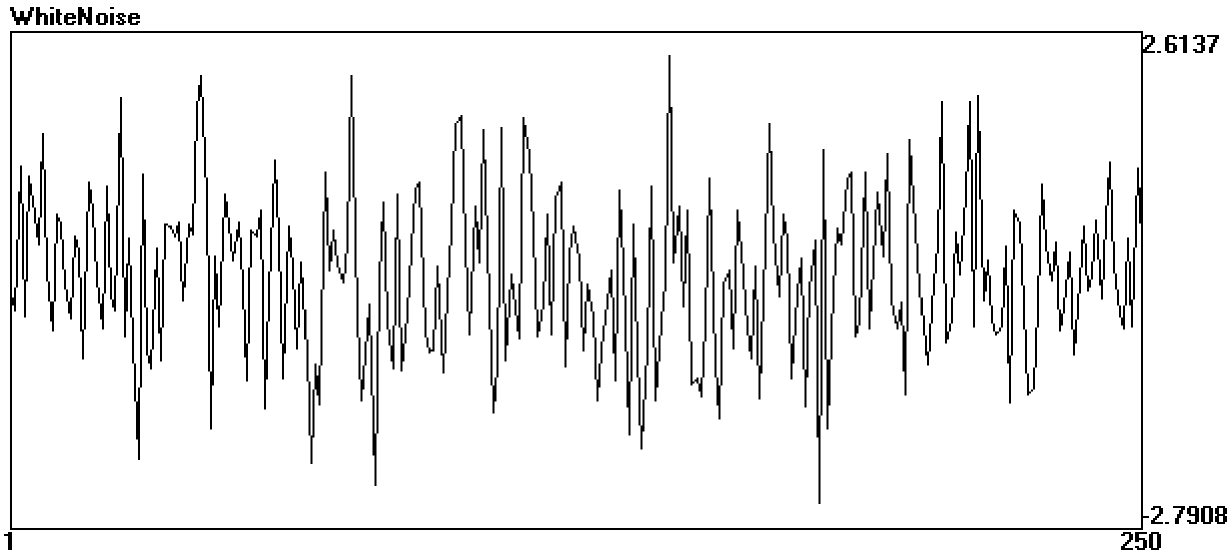
where ε_t is white noise. This gives a geometrically declining auto-correlation function (think of powers of α) and a partial autocorrelation function with zeros for k>1 (why?)

MA(1): $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$

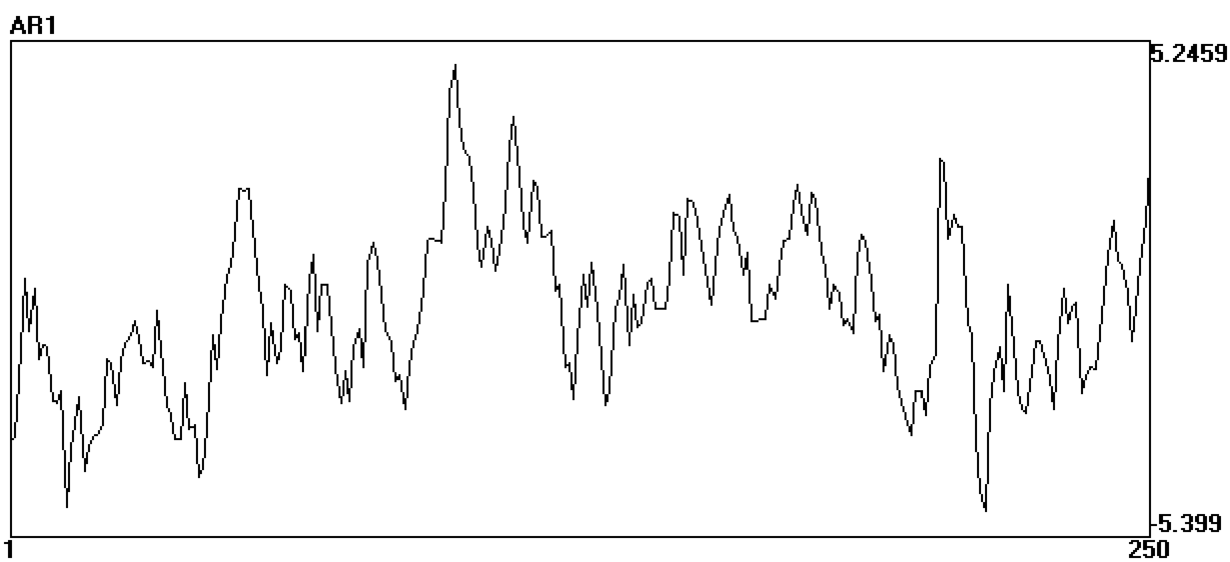
with ε_t white noise. The autocorrelation function is zero for k>1 and the pac function declines geometrically in absolute value but has positive values at odd lags (for θ positive) and negative values at even lags.

White noise has zero correlations of all types.

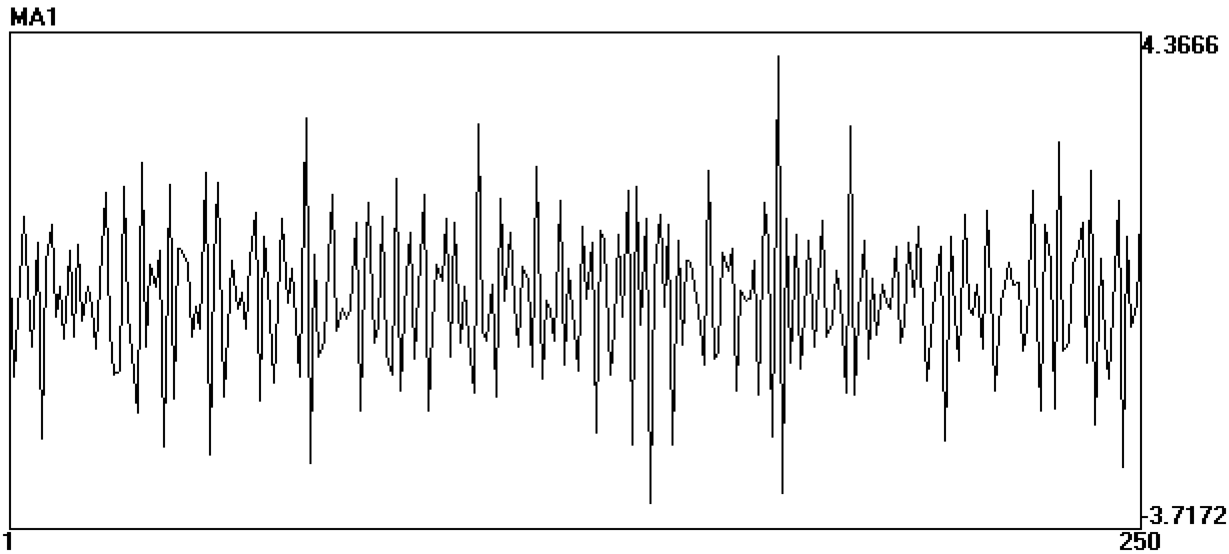
White Noise



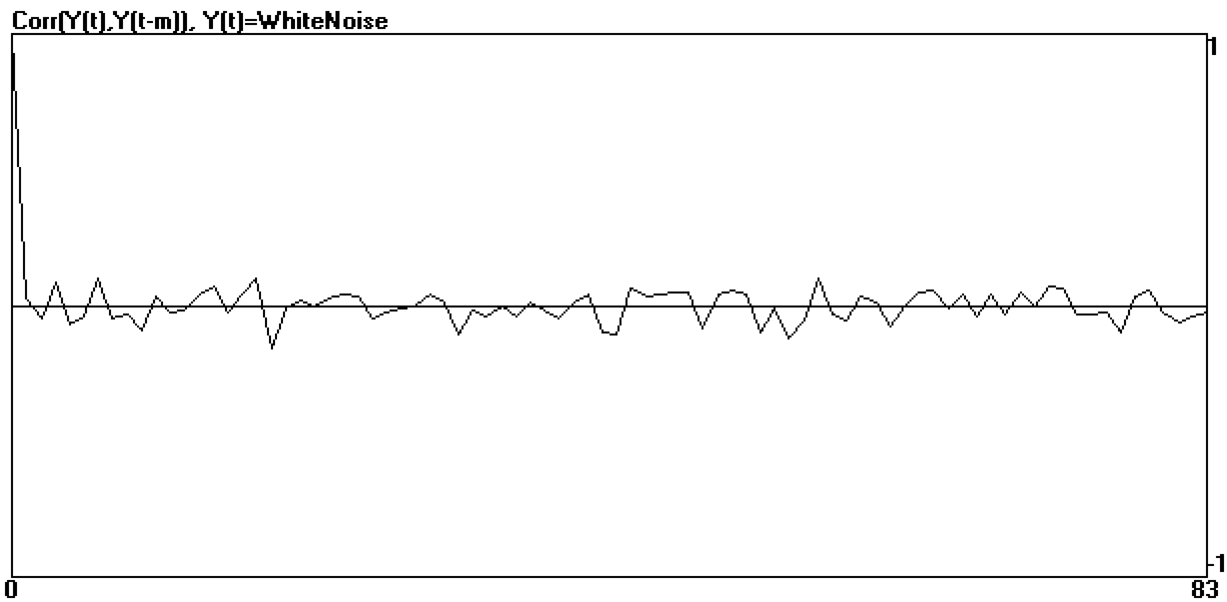
AR(1), parameter = 0.9

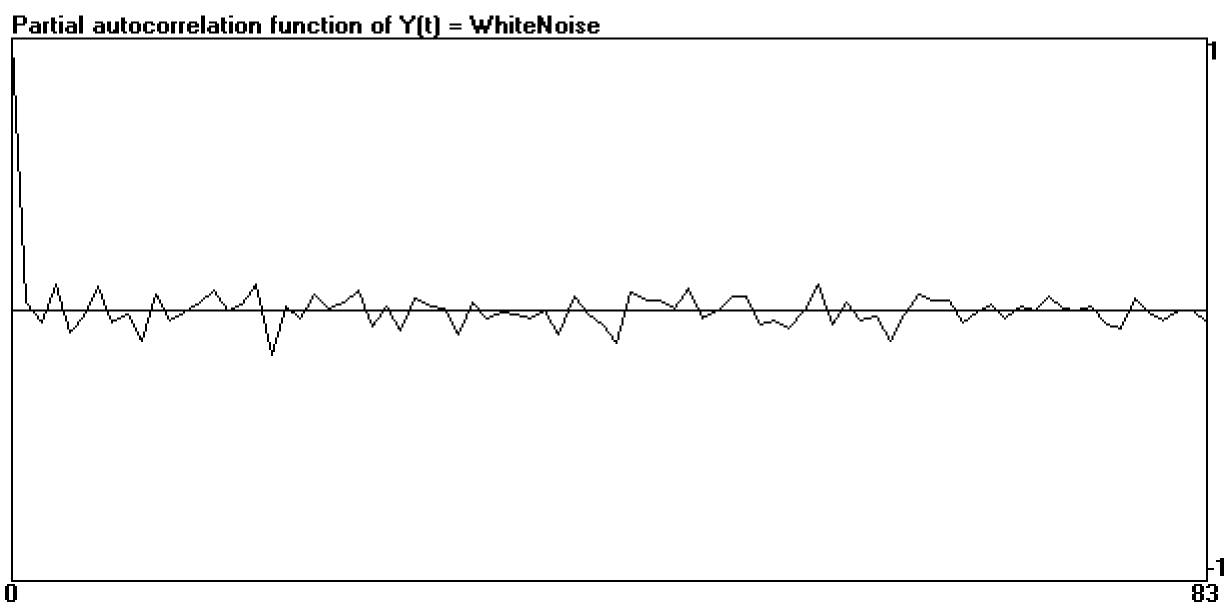


MA(1), parameter = 0.9

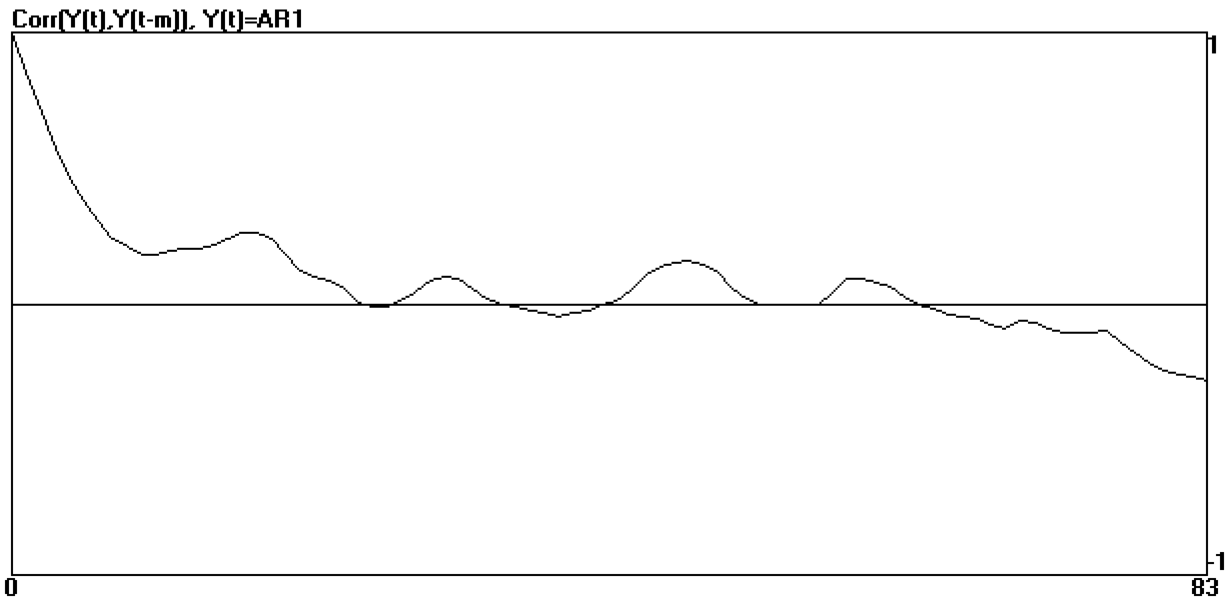


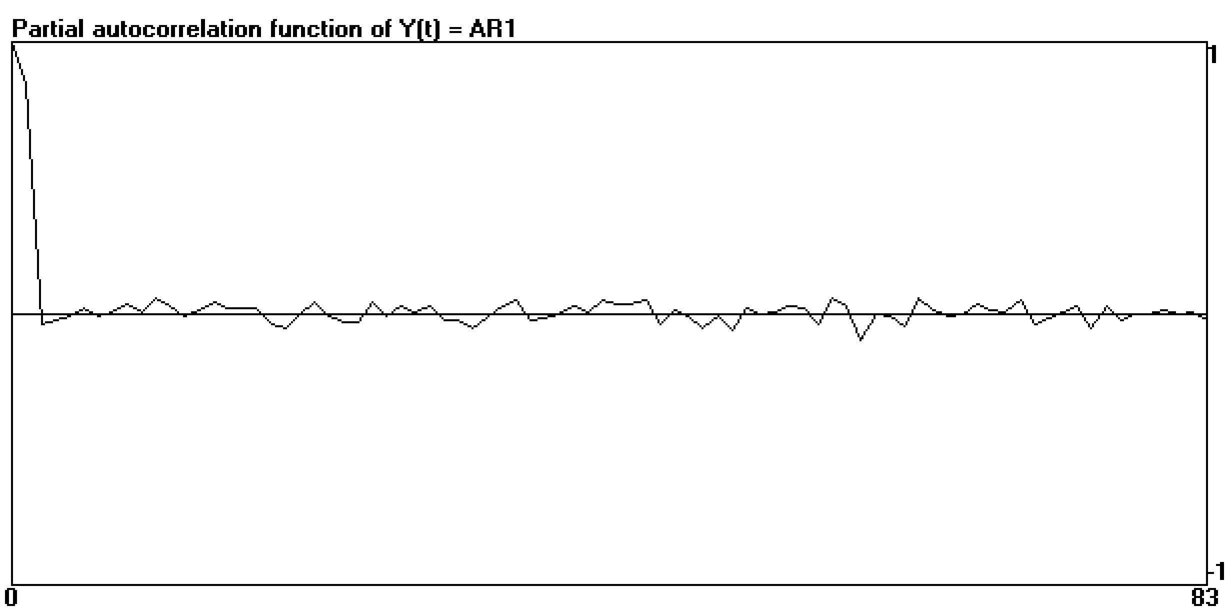
Autocorrelations and pac, White Noise



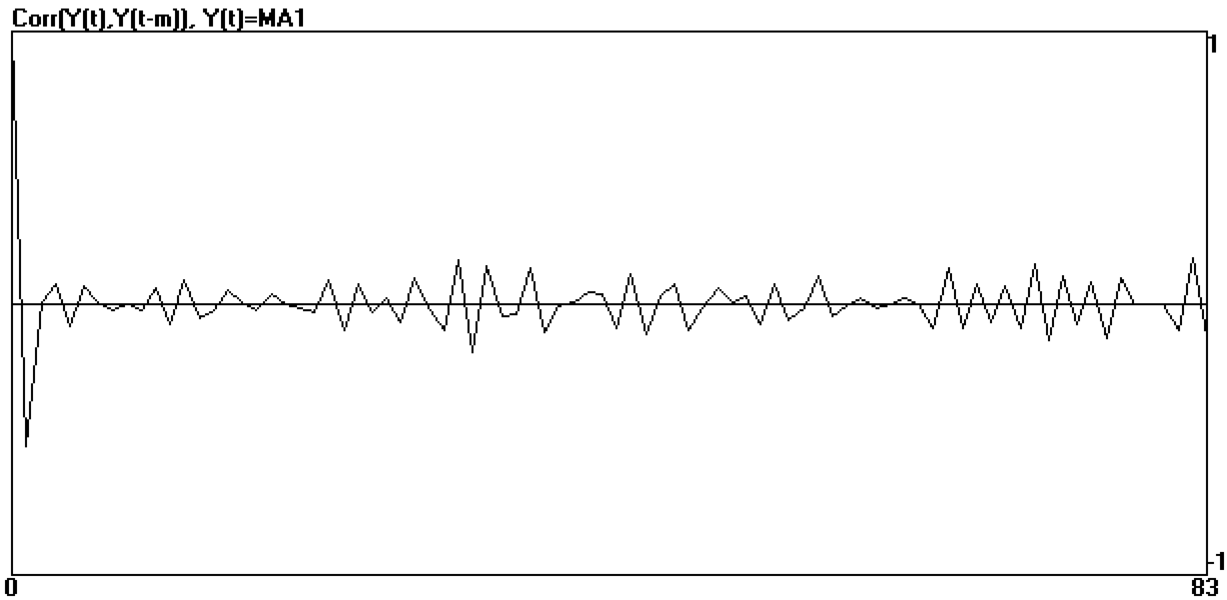


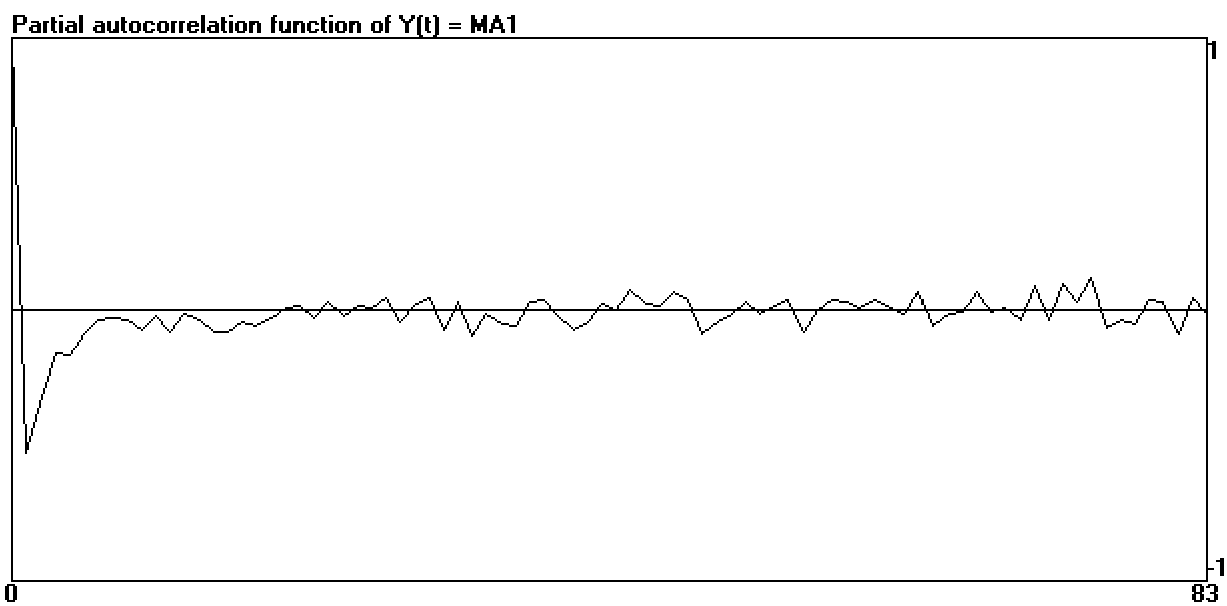
Ac and pac for AR(1) (0.9)





Ac and pac for MA(1) (0.9)





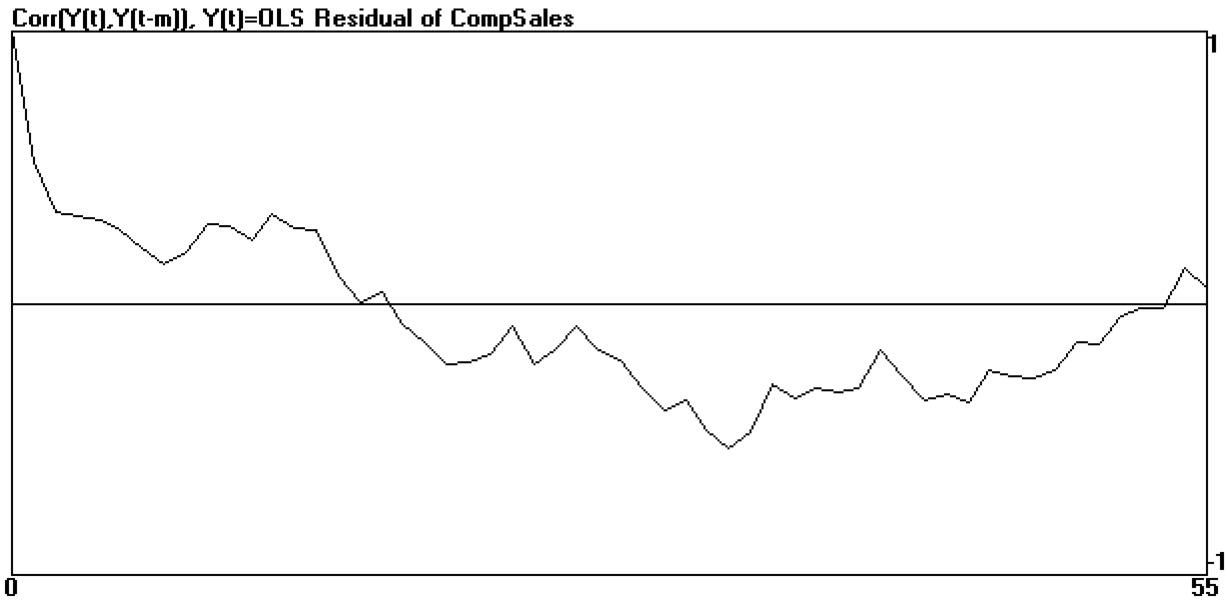
These models can be extended to higher order - AR(p) or MA(p). These give more complicated patterns of auto-correlations. Also, the models can be mixed: ARMA models

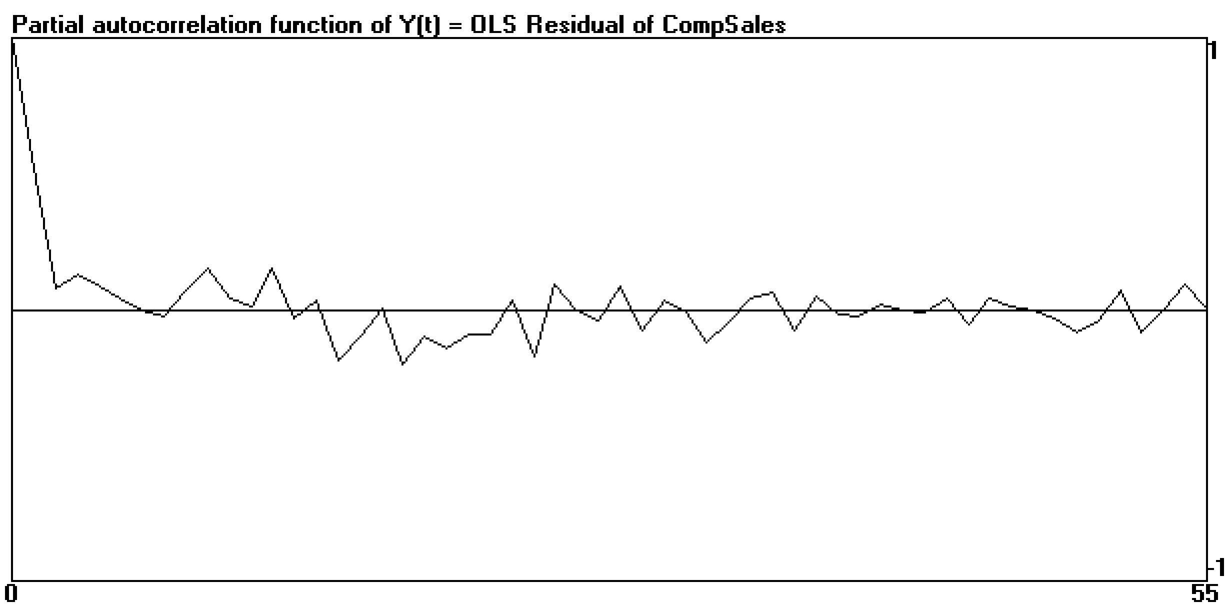
For example: ARMA(1,1)

$$y_t = \alpha y_{t-1} + \varepsilon_t + \theta\, \varepsilon_{t-1}$$

This can get out of hand - ARMA(p,q), ARIMA(p,d,q), etc.

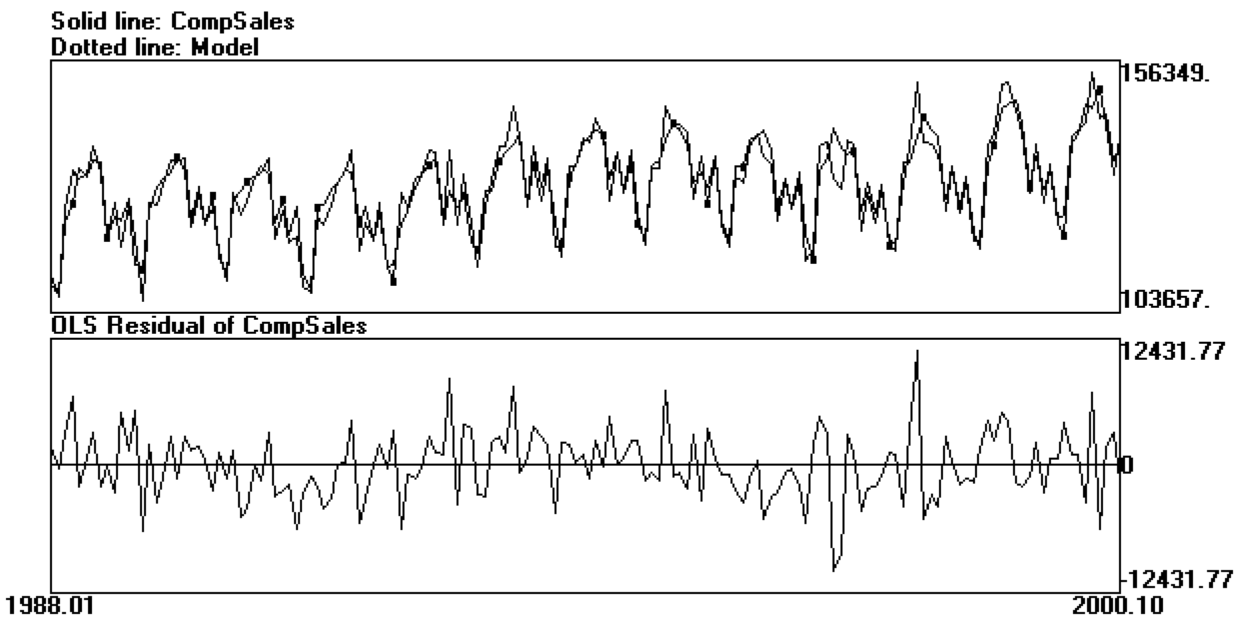
Can get at stochastic seasonality with lagged variables - eg. In monthly data, we often use lags at 1,3, and 12, corresponding to monthly, quarterly, and annual dependence.



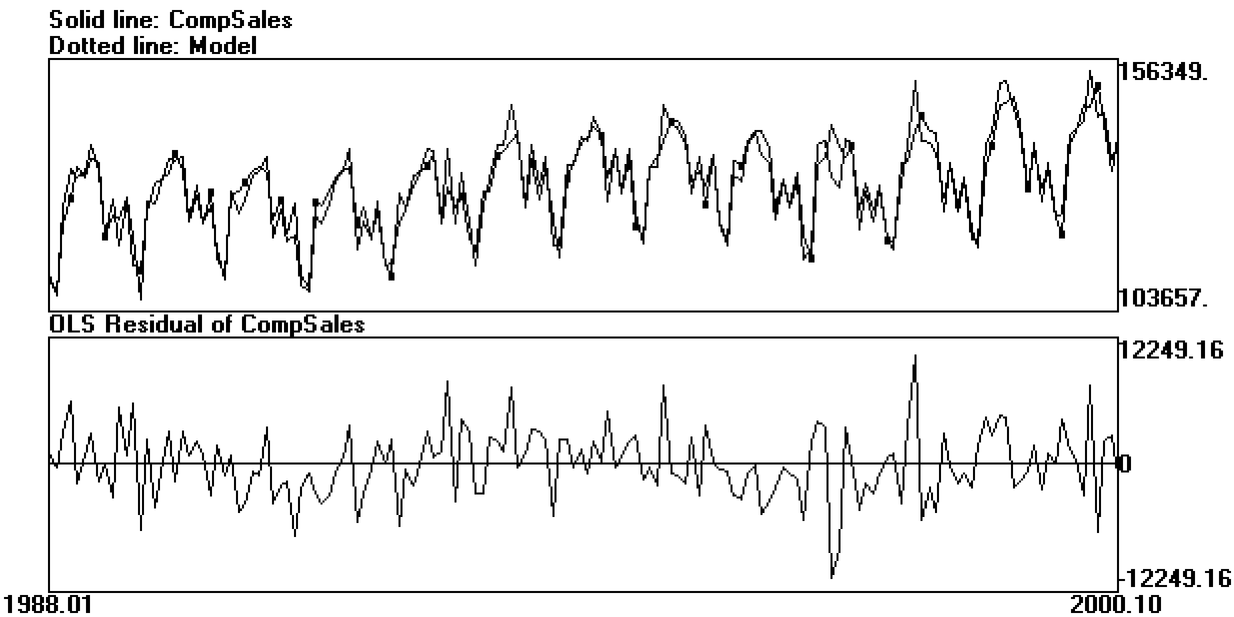


It looks like most of the action is AR

An AR specification with lags at 1,2,3,and 12 gives a better fit, with lag 2 insignificant

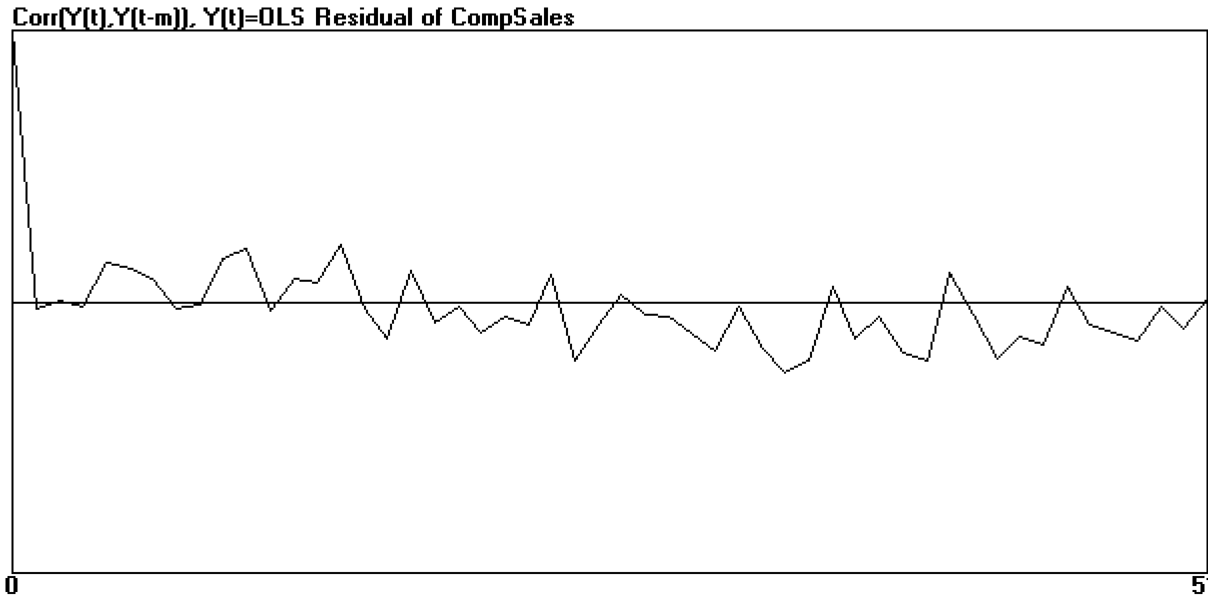


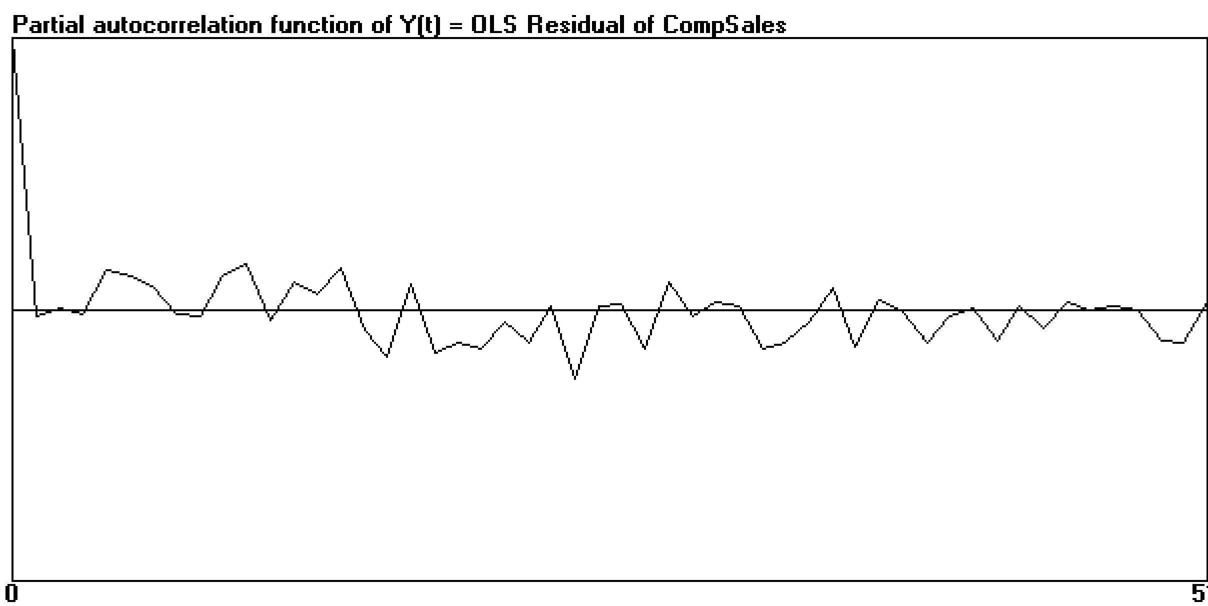
The series is still quite volatile - we now drop the lag at length 2 and add the number of saturdays in the month as an explanatory variable.



The standard error is 3946$ (compare with the mean of 129552$) and the R-squared is 0.88.

The residual autocorrelation and pac functions are pretty flat ...

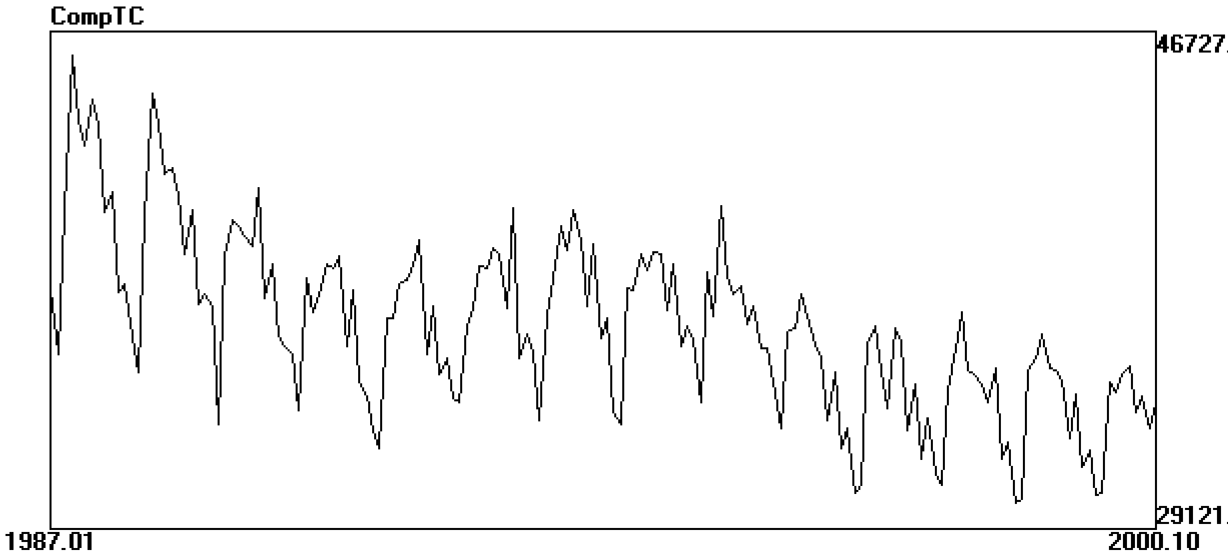


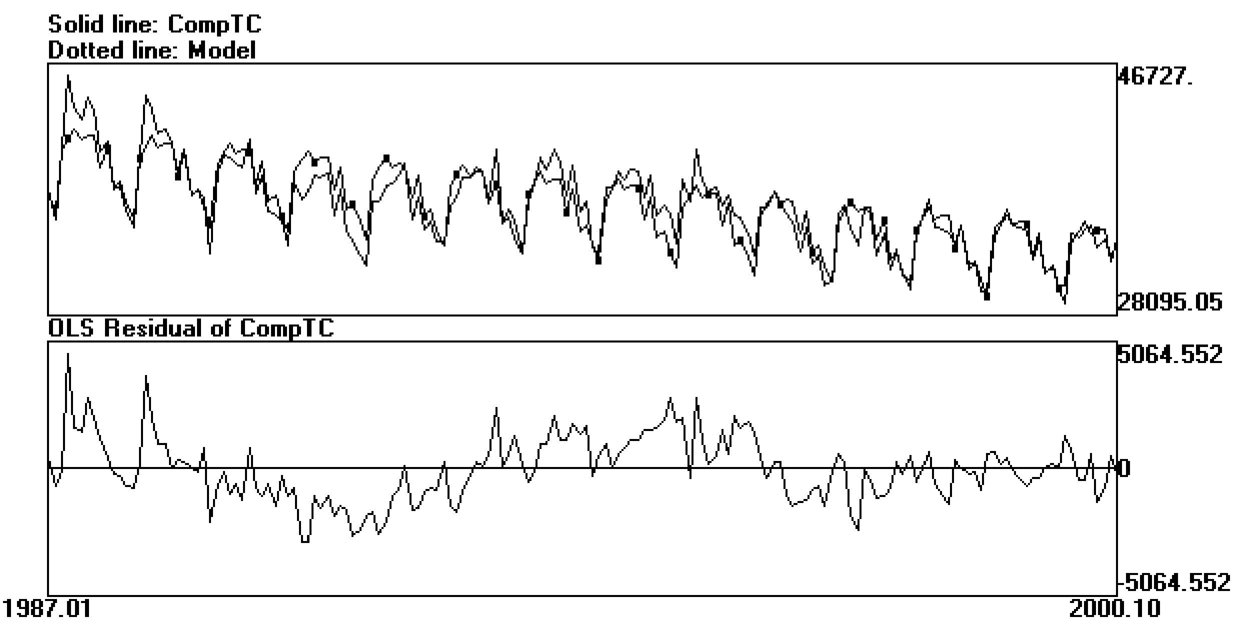


As a check on the specification, a model (ARMA) with lags at 1,3, and 12 was fit to the residuals. None of the coefficients were important and they were jointly insignificant, indicating that there is not much discernable structure remaining.

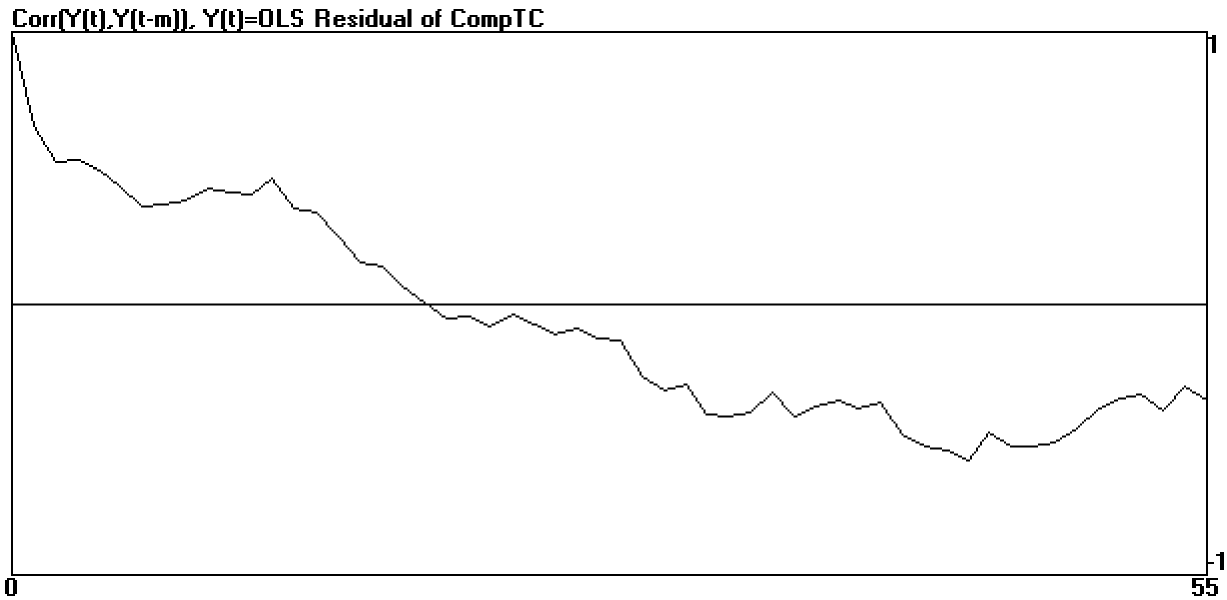
The forecast for 2000.11 is 130656$ with a standard error of about 4000$.

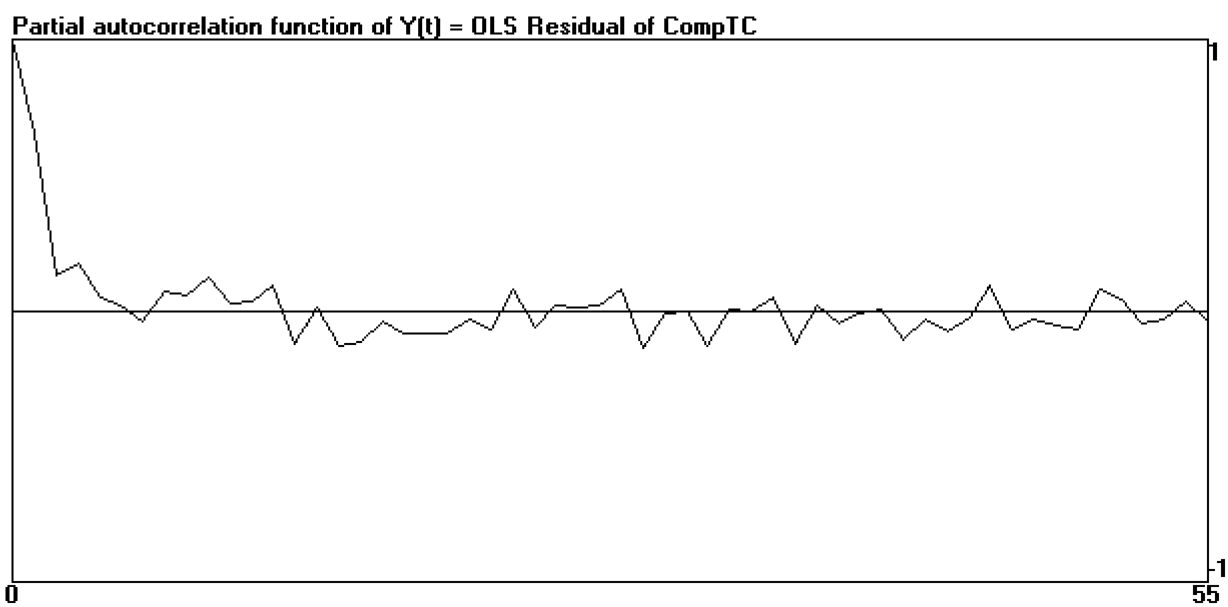
Series 2: Transaction Counts
- negative trend
- clear seasonality



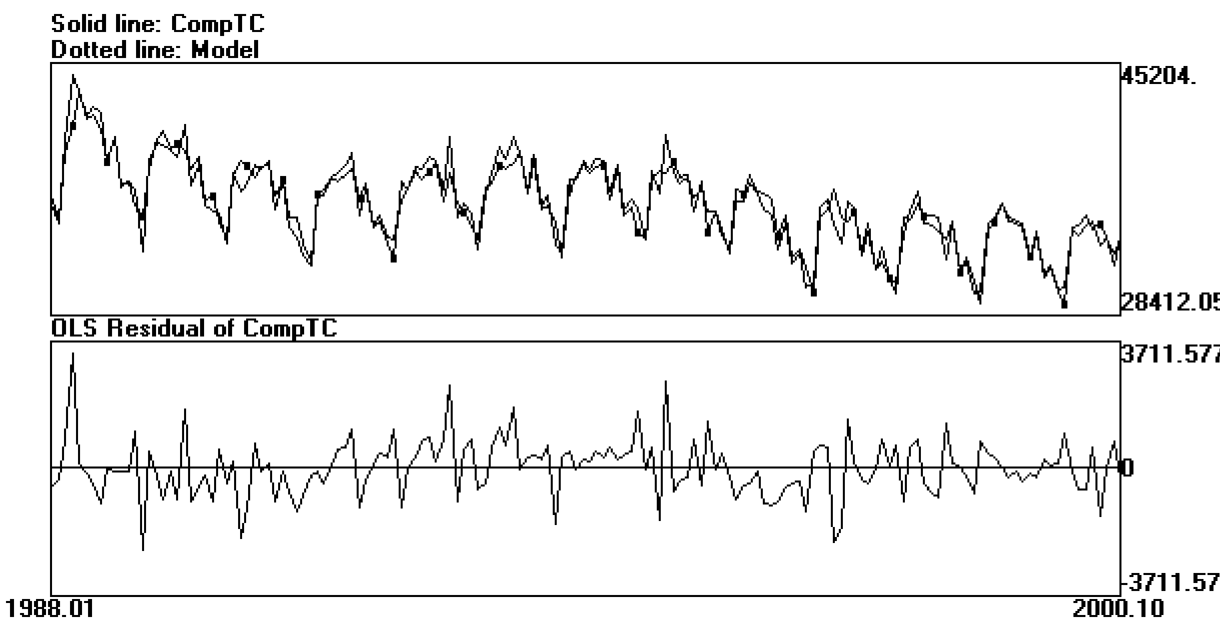


Autocorrelation and pac functions (for the residuals from last slide) -looks like AR dominates again

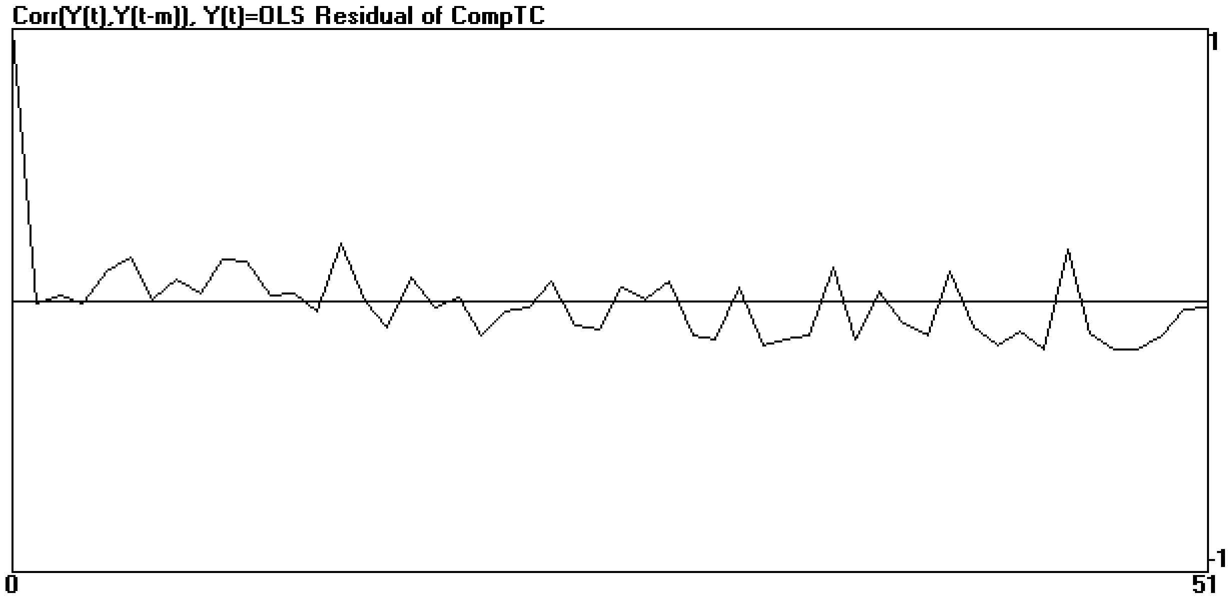


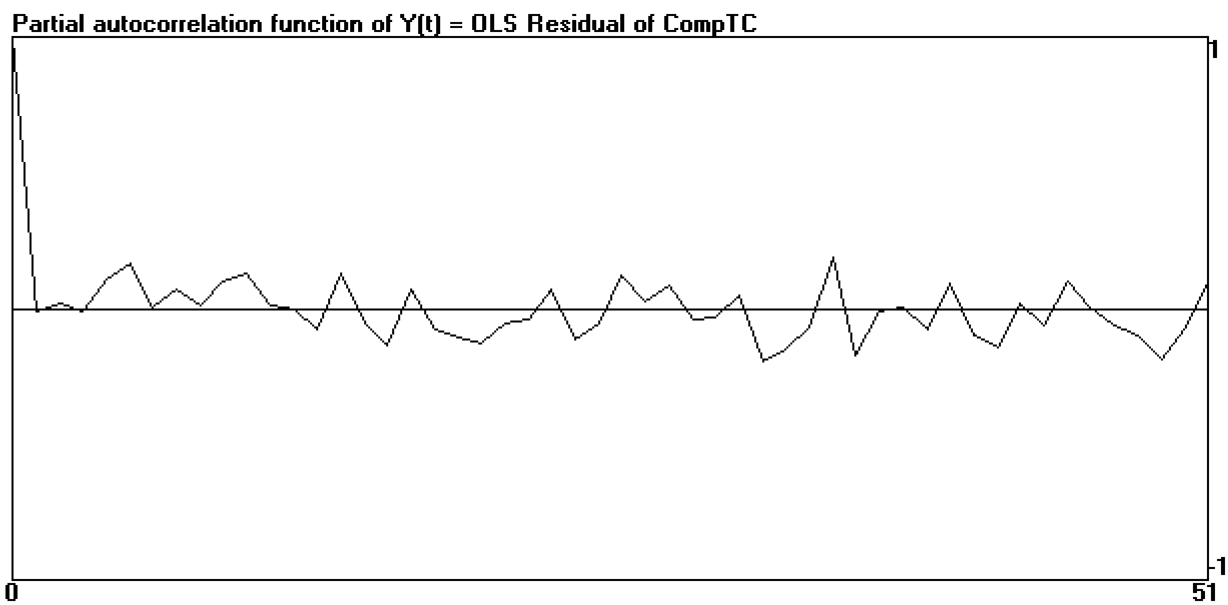


An AR(12) model shows important dynamics at lags 1, 3 and 12. Including the number of saturdays as well the results are:



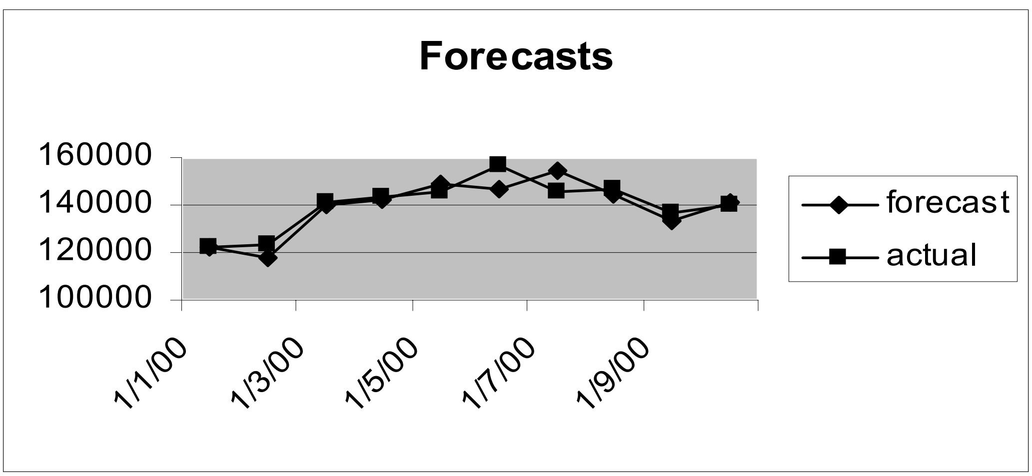
The residual autocorrelation and pac functions show little remaining structure

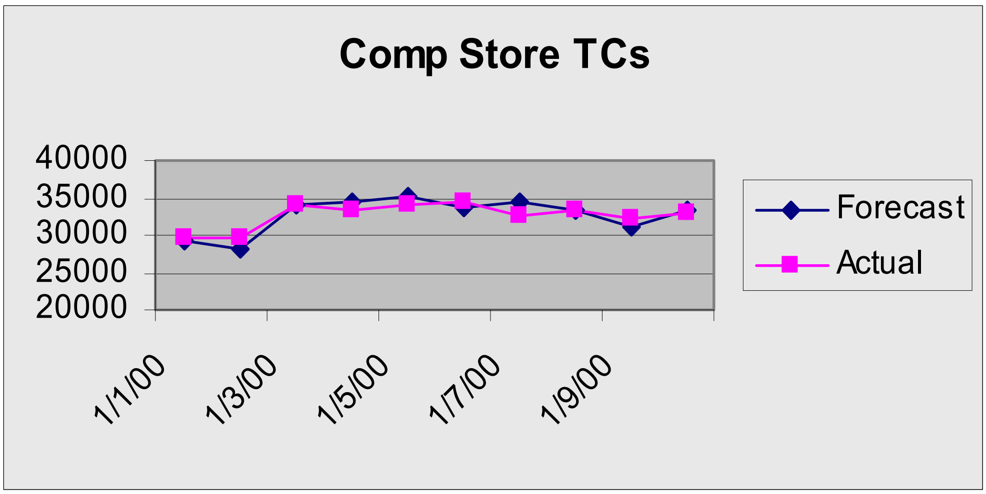




The prediction for Nov. is 29865, se 1032

Sales Forecasts





Comp Store TCs
40000
35000
30000
25000
20000
1/1/00
1/3/00
1/5/00
1/7/00
1/9/00
Forecast
Actual