POLS571 - Longitudinal Data Analysis
September 13, 2001

1 MA(q) Series

Moving average models are those in which current values of Y_t depend on current and past values of some random ("white noise") innovation. The key difference between MA models and those we've talked about so far is that the influence of a shock has *finite persistence*; the effect of a given innovation u_t lasts only as many periods as the model has lags, after which it vanishes entirely.

1.1 The MA(1) Model

The simplest *moving average* process is the MA(1):

$$Y_t = u_t + \theta u_{t-1},\ t = 1, 2, ...T \tag{1}$$

where u_t is i.i.d. with mean zero and constant variance σ_u^2. Its easy to calculate the properties of this series. The mean is:

$$\begin{aligned} \mu &= E(Y_t) \\ &= E(u_t) + \theta E(u_{t-1}) \\ &= 0 \end{aligned} \tag{2}$$

and the variance is:

$$\begin{aligned} \sigma_Y^2 &= E[(u_t + \theta u_{t-1})(u_t + \theta u_{t-1})] \\ &= E(u_t^2) + \theta^2 E(u_{t-1}^2) + 2\theta E(u_t u_{t-1}) \\ &= (1 + \theta^2)\sigma_u^2 \end{aligned} \tag{3}$$

The one-lag covariance is:

$$
\begin{aligned}
\gamma_1 &= E[(u_t + \theta u_{t-1})(u_{t-1} + \theta u_{t-2})] \\
&= E(u_t u_{t-1}) + \theta E(u_{t-1}^2) + \theta E(u_t u_{t-2}) + \theta^2 E(u_{t-1} u_{t-2}) \\
&= \theta E(u_{t-1}^2) \\
&= \theta \sigma_u^2 \qquad (4)
\end{aligned}
$$

and the two- and higher-lag covariances are all zero (why?).

For this series, then:

1. the means, variances and autocovariances are all independent of t,
2. the ACF is zero for all lags greater than one,
3. the ACF at one lag is dependent on the degree of dependence in the moving average process. In particular, since we earlier defined $\hat{\rho_1}$ as $\frac{\hat{\gamma_s}}{\hat{\sigma^2}}$, this suggests that

$$
\begin{aligned}
\hat{\rho_1} &= \frac{\theta \sigma_u^2}{(1+\theta^2)\sigma_u^2} \\
&= \frac{\theta}{1+\theta^2} \qquad (5)
\end{aligned}
$$

(see the figure). Note that there are two values of θ corresponding to every value of ρ_1; these values are reciprocals of one another (so ρ_1 has the same value at $\theta = 2$ and $\theta = 1/2 = 0.5$). This is because we can write $\rho_1 = \frac{1/\theta}{1+(1/\theta)^2} = \frac{\theta^2(1/\theta)}{\theta^2[1+(1/\theta)^2]} = \frac{\theta}{\theta^2+1}$. This means that, e.g.,

$$
Y_t = u_t + 0.5u_{t-1} \qquad (6)
$$

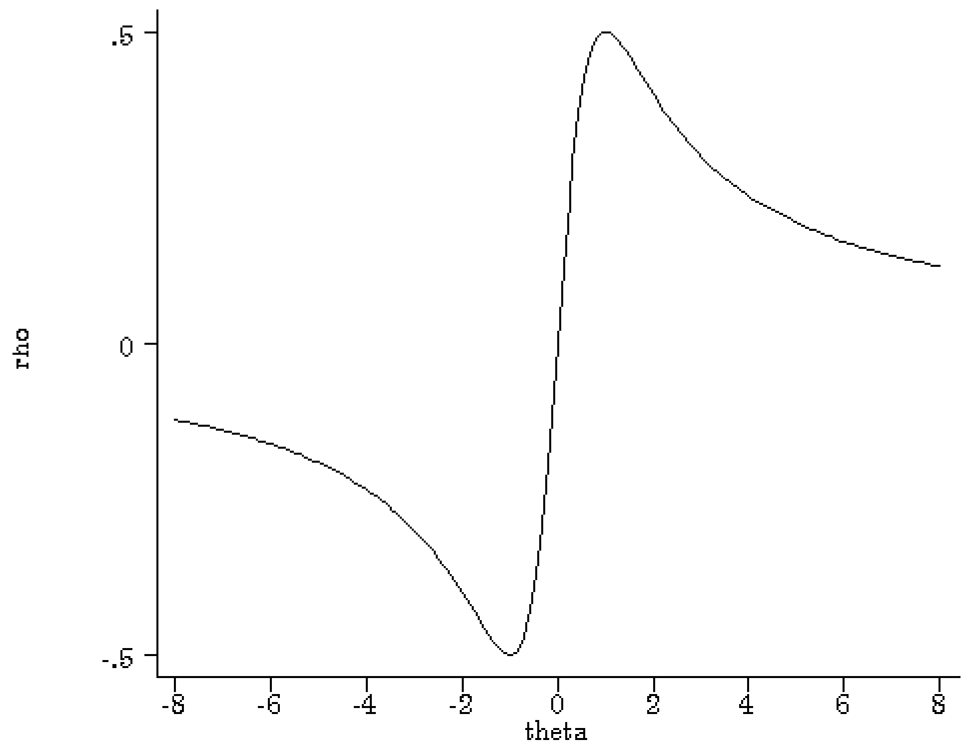
and

$$
Y_t = u_t + 2u_{t-1} \qquad (7)
$$

will have the same value for $\rho_1 = \frac{2}{(1+2^2)} = 0.4$.

More generally, for the MA(1) series given here,

Figure 1: Relationship between θ and ρ_1 for an MA(1) series



- $\theta = 0$ corresponds to $\rho_1 = 0$, and is a "white noise" process (i.e., no temporal dependence in Y_t).
- When θ is positive,
  1. ρ_1 will be greater than zero,
  2. successive values of Y_t will be *positively* related, and
  3. the series will be "smoother" than a white noise series.
- In contrast, when θ is less than zero,
  1. ρ_1 will be less than zero as well,
  2. successive values of Y_t will be *negatively* correlated, and

3. the series will be less "smooth" than a white noise sequence.

We can rewrite (1) by noting that $u_t = Y_t - \theta u_{t-1}$, and repeatedly substituting lagged values of u_t into the equation:

$$\begin{aligned} Y_t &= u_t + \theta u_{t-1} \\ &= u_t + \theta(Y_{t-1} - \theta u_{t-2}) \\ &= u_t + \theta Y_{t-1} - \theta^2 u_{t-2} \\ &= u_t + \theta Y_{t-1} - \theta^2 (Y_{t-2} - \theta u_{t-3}) \\ &= u_t + \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 u_{t-3} \\ &= \ldots \\ &= u_t + \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 Y_{t-3} - \ldots - (-\theta)^{(T-1)} Y_{t-(T-1)} - (-\theta)^T u_0 \end{aligned} \tag{8}$$

As $T \to \infty$, the last term in (16) drops out, and we can write the model as:

$$Y_t = -\sum_{j=1}^{\infty} (-\theta)^j Y_{t-j} + u_t \tag{9}$$

Note the similarity between (9) and the AR(p) model discussed before. In fact, a famous (and important) theorem in time-series analysis is the *Wold Decomposition Theorem*. This theorem states that:

Any weakly stationary, purely nondeterministic stochastic process can be written as a linear combination of a sequence of uncorrelated random variables.

That is, such a series can be written as:

$$Y_t = \sum_{j=0}^{\infty} \psi_j u_{t-j} \tag{10}$$

where the ψs are parameters. To ensure that the process has finite variance over time, we typically need to impose the condition

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty \tag{11}$$

In the MA(1) example we just discussed, it clear from (9) that the process meets these guidelines when $\theta < 1$. The same is true for the AR(1) process we discussed earlier, provided that $\phi < 1$.

1.2 MA(q) Models

A more general MA(q) model is:

$$Y_t = u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + ... + \theta_q u_{t-q},\ t = 1, 2, ...T \tag{12}$$

A finite moving average process of order q is always stationary, even without restrictions on the θs, because of the finite persistence of any single shock. If the model is expressed in AR form (as in (9) above), then the MA parameters must satisfy the stationarity conditions similar to those required of AR models. An MA model which satisfies these conditions is said to be *invertible*. An MA(1) model with $|\theta| > 1$ is thus not invertible, but remains stationary.

2 A General ARMA Model

We can put together everything we've learned so far by considering a model that has both AR and MA components, called an ARMA(p,q) model:

$$Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + u_t + \theta_1 u_{t-1} + ... + \theta_q u_{t-q} \tag{13}$$

Using the lag operator, we can rewrite this model in the form:

$$(1 - \phi L - \phi_2 L^2 - ... - \phi_p L^p)Y_t = (1 + \theta L + \theta_2 L^2 + ... + \theta_q L^q)u_t \tag{14}$$

Moreover, if the model requires differencing in order to achieve stationarity, we call it an ARIMA(p,d,q) model; e.g., the ARIMA(p,1,q) model:

$$\Delta Y_t = \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + u_t + \theta_1 u_{t-1} + ... + \theta_q u_{t-q} \tag{15}$$

The parameters $\hat{\phi}$ and $\hat{\theta}$ of an ARIMA model can be estimated using MLE; more on this in a little bit...

3 What do these series look like?

As an example, I generated some $N(0,1)$ white noise errors, and then created four different series, all starting from a value of zero and all *from the same errors.* The series follow:

$$Y_t = 0.9Y_{t-1} + u_t \quad (16)$$
$$Y_t = 0.9u_{t-1} + u_t \quad (17)$$
$$Y_t = Y_{t-1} + u_t \quad (18)$$
$$Y_t = 0.9Y_{t-1} + 0.9u_{t-1} + u_t \quad (19)$$

I plot these figures below...

Notice several things about the series:

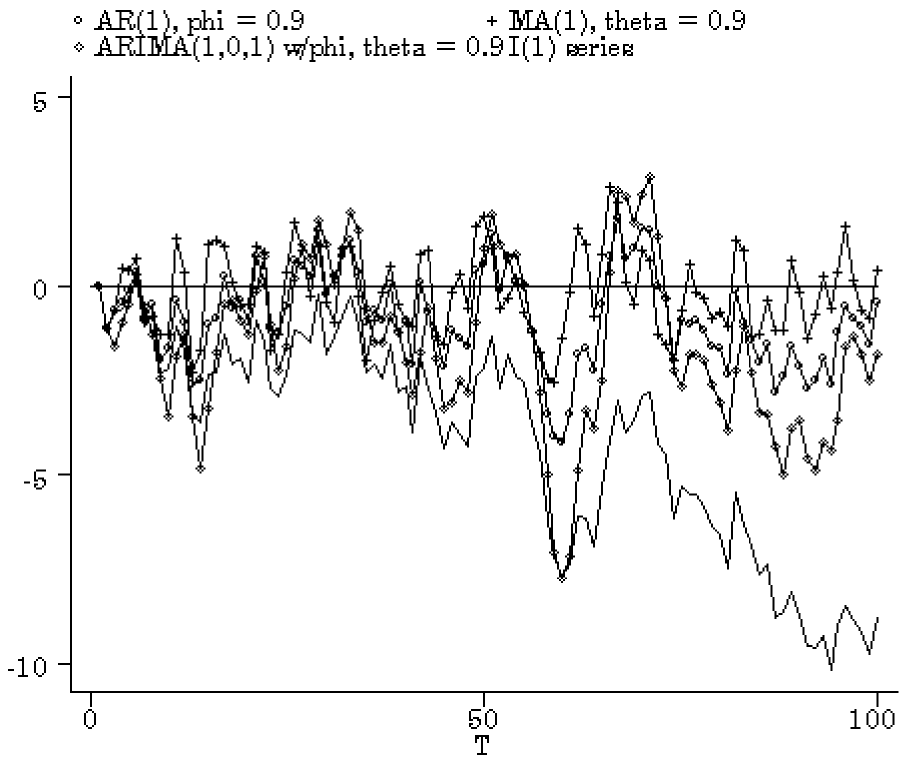
- The AR(1), MA(1) and ARMA(1,1) series are all mean-reverting, while the I(1) series "drifts" away from the mean (zero).
- The MA(1) series stays the "closest" to the mean – because the "shocks" don't persist, the series never strays far from its mean level.
- The ARMA(1,1) series "looks like" the AR(1) series, but has more variability and is less mean-reverting.

If we look at the values of the means and variances for the first 100 observations of these series, we see the following:

Variable	Obs	Mean	Std. Dev.	Min	Max
AR(1)	100	-.8392515	1.235517	-4.154195	1.811882
MA(1)	100	-.175531	1.096084	-2.555826	2.634273
ARMA(1,1)	100	-1.590666	2.176644	-7.756127	2.884106
I(1)	100	-4.071511	2.839775	-10.17822	.1636956

Next, let's take a look at the ACFs and PACFs of these series...

Figure 2: AR(1), MA(1), ARIMA(1,0,1) and I(1) series, $\phi = \theta = 0.9$



3.1 AR(1) identification

For AR(1) series with $|\phi| < 1$, the ACF should be a smoothly declining function of time; either positive (for $\phi > 0$) or alternating between positive and negative (for $\phi < 0$). The extent to which this will be true in practice, however, depends a lot on the length of the series.

Here's the ACF for our AR(1) series, for $T = 50$ and $T = 300$:

Figure 3: Correlogram, AR(1) series with $\phi = 0.9$, $T = 50$

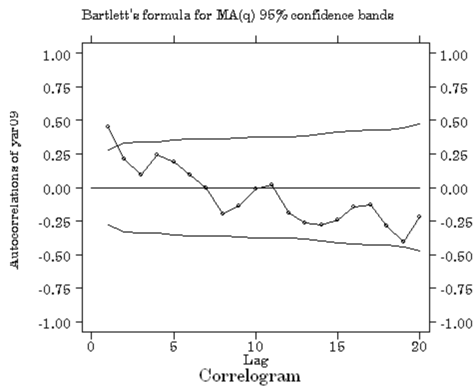
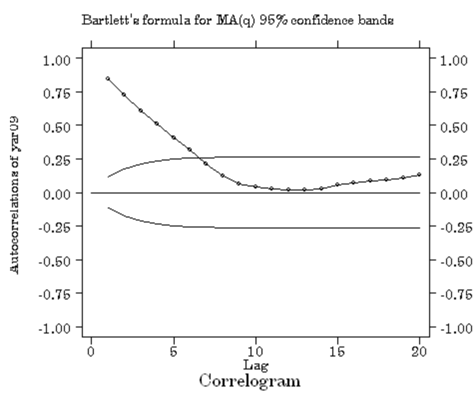


Figure 4: Correlogram, AR(1) series with $\phi = 0.9$, $T = 300$



Clearly, the more observations we have, the better...

Contrary to the ACFs, the PACFs for an AR(p) series should be zero beyond p lags. Again, however, this is true only as $T \to \infty$; in practice, the truth can be harder to find...

Figure 5: PACF, AR(1) series with $\phi = 0.9$, $T = 50$

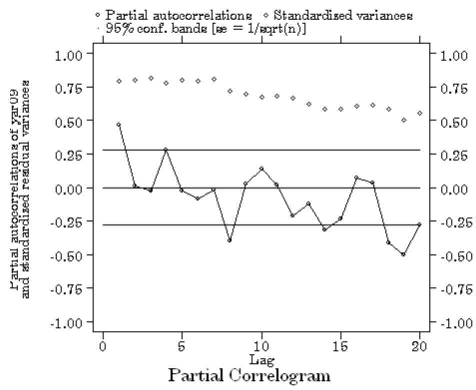
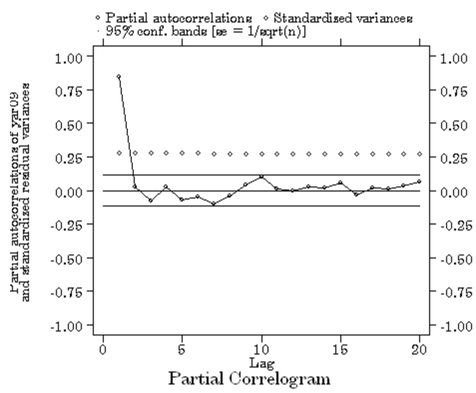


Figure 6: PACF, AR(1) series with $\phi = 0.9$, $T = 300$



It would be easy to identify this as a AR(1) series with 300 observations, less clear if we only had 50...

3.2 MA(1) identification

We follow a similar process for identifying MA(q) series. For such series, the ACF plot dies out quickly; beyond q lags, the ACF for an MA(1) series is zero. Again, however, these statements are true only asymptotically; in practice, the number of observations we have determine how easy it is for us to identify the series...

Figure 7: ACF, MA(1) series with $\theta = 0.9$, $T = 50$

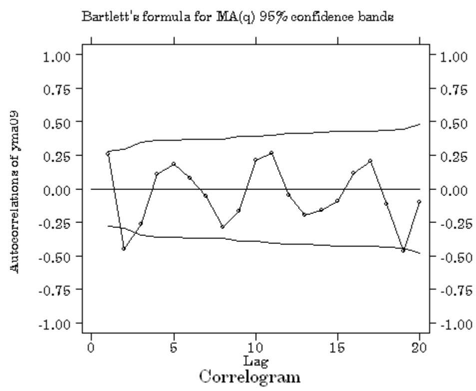
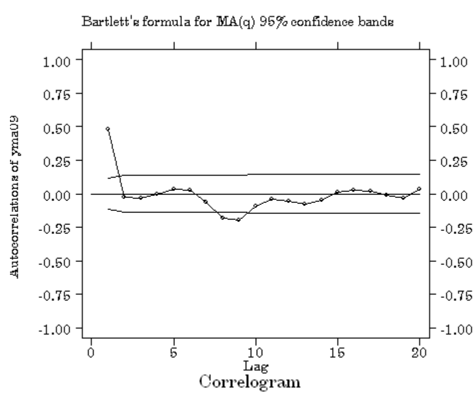


Figure 8: ACF, MA(1) series with $\theta = 0.9$, $T = 300$



In contrast to the AR(p) series, an MA(q) series will show a decaying PACF:

Figure 9: PACF, MA(1) series with $\theta = 0.9$, $T = 50$

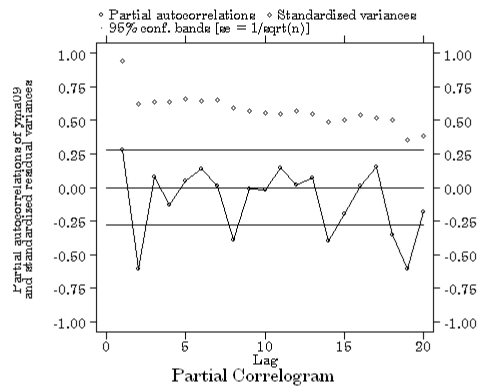
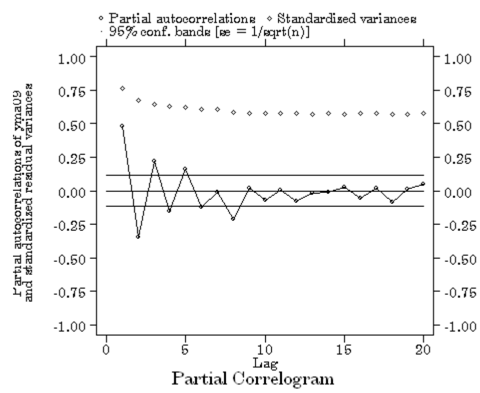


Figure 10: PACF, MA(1) series with $\theta = 0.9$, $T = 300$



Here, the decay is alternating between positive and negative values. The PACF is crucial for differentiating between, e.g., an AR(1) model and a higher-order MA(q) model. For example, the ACF of the two series

$$Y_t = 0.8Y_{t-1} + u_t \tag{20}$$

and

$$Y_t = 0.9u_{t-1} + +0.6u_{t-2} + 0.3u_{t-3} + 0.1u_{t-4} + u_t \tag{21}$$

will be almost identical in practice. The PACFs, however, will be very different, since there is no partial autocorrelation beyond one lag in the AR(1) model.

3.3 ARMA(p,q) model identification

For an ARMA(p,q) model, we expect to see "decaying" functions in both the ACF and the PACF. For illustration, here are the ACF and PACF for the ARMA(1,1) series with $\phi = \theta = 0.9$ and $T = 300$:

Figure 11: ACF, ARMA(1,1) series with $\phi = \theta = 0.9$, $T = 300$

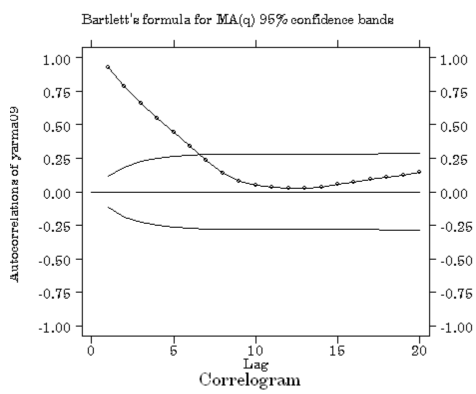
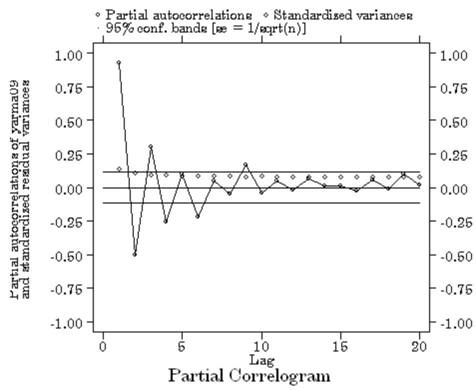


Figure 12: PACF, ARMA(1,1) series with $\phi = \theta = 0.9$, $T = 300$



3.4 Identifying Integrated series

Being variants of AR(1) series, integrated series exhibit many of the same patterns (that is, decaying ACFs and spiking PACFs). In fact, not surprisingly, its often hard to tell the difference between I(1) series and AR(1) series with ϕ close to 1.0. Here are the ACF and PACF plots for the I(1) series above (with $T = 300$).

Figure 13: ACF, I(1) series, $T = 300$

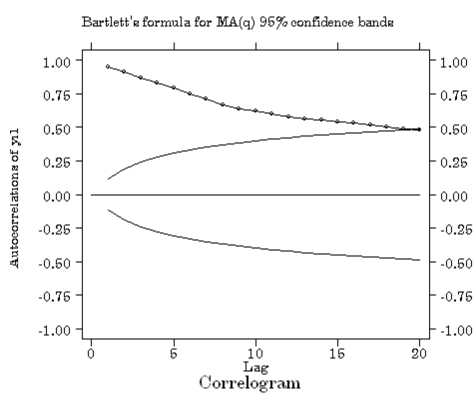
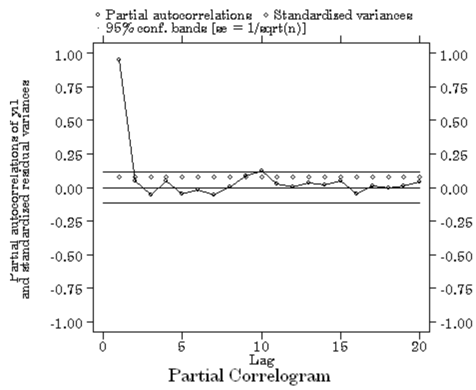


Figure 14: PACF, I(1) series, $T = 300$



Identification of I(1) series can be an especially difficult issue for ARIMA modelers, since assessing whether or not the series requires differencing is a key early decision. Getting it wrong can lead to *overdifferencing*, which leads to parameter redundancy and general ugliness (see McCleary and Hay for a good discussion of this).

4 Practical Issues - Identification, Estimation and Testing

The basic idea behind ARIMA modeling is a series of steps, usually along the lines of:

$$Identification \rightarrow Estimation \rightarrow Diagnostics$$

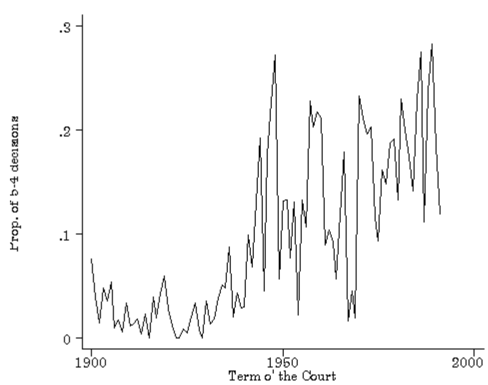
ARIMA modeling begins on the premise that you first need to identify what kind of data generating process is driving the data (e.g., AR, MA, integrated, etc.). Once that is accomplished, parameters may be estimated, after which diagnostic checks on the remaining residuals are run to make sure that the us are white noise. Only then can forecasts be made, etc. (If this seems a little *ad hoc*, it is...).

More formally, we can think of the process as:

1. Determine of the series is stationary. If so, proceed to (3).
2. If not, difference the series until it is.
3. Examine the ACFs and PACFs of the stationary series, in order to determine the data generating process (AR, MA, or some combination thereof).
4. Fit a model – starting simple – using ARIMA/MLE.
5. Examine and test the residuals to determine if they are white noise. If they are, proceed to (7).
6. , if they are not, go back to (3) and try again.
7. Proceed with inference, hypothesis testing, forecasting, and the like.

Consider the following example: the proportion of Supreme Court decisions decided by only one vote, by year, from 1900-1992. The series looks like this:

Figure 15: Proportion of One-Vote Supreme Court Decisions, 1900-1992



We begin by noting that the series doesn't really look very stationary, and its certainly not stationary in the variance; but that it also doesn't appear to be "drifting" over time either. For now, we'll adjust the variance issue by logging the series; we'll come back to the idea of modeling time series variances when we discuss ARCH/GARCH models. An examination of the ACF and PACF on the logged series reveals the following:

Figure 16: ACF, Supreme Court "One-Vote" series, 1900-1992

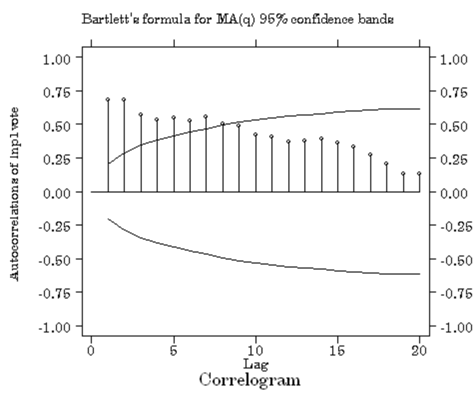
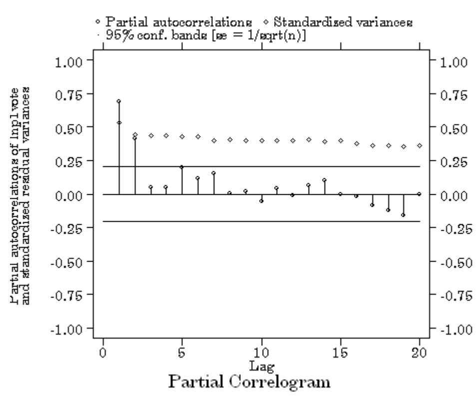


Figure 17: PACF, Supreme Court "One-Vote" series, 1900-1992



There's a generally decaying ACF, but its decaying relatively quickly; and there are big one- and two-lag spikes in the PACF. These things suggest an

AR(2) model. So, we can estimate such a model, using the `-arima-` command in Stata.

```
.  arima pr1vote, ar(1 2)
```

This gives us estimates as follows:

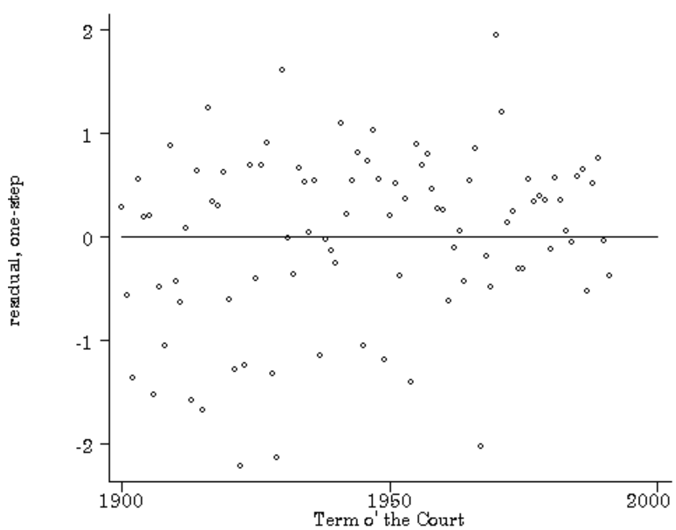
$$
\begin{array}{llllllll}
Y_t & = & \text{-2.856} & + & 0.400Y_{t-1} & + & 0.409Y_{t-2} & + \quad u_t \\
 & & 0.556 & & 0.084 & & 0.108 &
\end{array}
$$

which look pretty good (significant, etc.). Next, though, we have to determine if that is all we need to do. We do this by determining if the errors are white noise or not. The easiest way to do this is to generate and plot the residuals, and possibly to do some sort of test (e.g. a Q test).

```
.  predict ar2res, residuals

.  gra ar1res year, xlab ylab yline(0)
```

Figure 18: Plot of $\hat{u}$s, AR(2) model



The residuals look like white noise to me. A more formal test is the Q test: If the residuals are a white noise process, then:

$$Q = N(N+2)\sum_{j=1}^{s}\frac{1}{N-s}\hat{\rho}^2(j) \rightarrow \chi^2_s \tag{22}$$

where s is the number of lags specified and $\hat{\rho}_j$ is the estimated autocorrelation at lag s. Under the null hypothesis, the Q statistic converges to a chi-square distribution with j degrees of freedom. So, $\hat{Q}$ is a test for the null hypothesis of no autocorrelation in a series. The Stata command for this test is -wntestq-; its a good idea to estimate it for many values of s, just to be sure...

```
.  wntestq ar2res, lags(1)

Portmanteau test for white noise
---------------------------------------
Portmanteau (Q) statistic = 0.0118
Prob > chi2(1) = 0.9135

.  wntestq ar2res, lags(2)

Portmanteau test for white noise
---------------------------------------
Portmanteau (Q) statistic = 0.1310
Prob > chi2(2) = 0.9366

.  wntestq ar2res, lags(3)

Portmanteau test for white noise
---------------------------------------
Portmanteau (Q) statistic = 0.6555
Prob > chi2(3) = 0.8836

.  wntestq ar2res, lags(4)
```

```
Portmanteau test for white noise
---------------------------------------
Portmanteau (Q) statistic = 1.7582
Prob > chi2(4) = 0.7801
```

etc.

So, this is probably a pretty good model for the series we've got here.

Note several things about ARIMA models in general:

- Its rare, as an empirical matter, to have series that has more than two or three lagged effects. The exception are *seasonal* effects, which we aren't going to go into (for time reasons). (In practice, seasonal effects ought to be explained/modeled using covariates anyway...).
- Similarly, its rare to have an integrated series of higher order than I(1).
- In general, more parsimonious models are better. Its easy to "overfit" a model to the data using an ARIMA approach. The problem with over fitting is that the out-of-sample predictions are actually *worse* than those that would be obtained by a simpler model (since the model is essentially "fitting noise").

We're not going to go into this much more, even though there's a lot more we could do (diagnostics, prediction, etc.) We'll come back to these models later, however, and we'll discuss some multivariate Box-Jenkins models as well.