8
Modeling Variation In Time

Key terms and concepts: time series; trend; seasonality; mean reversion (autoregression)

As spring follows winter

This chapter covers a variety of techniques that are useful for analyzing time series, or data collected over time. (This is also called *longitudinal data*.) The key issue here is time dependence. Up to now, we've assumed that the ordering of our observations is irrelevant. We can no longer do that when our data is longitudinal: spring follows winter, and not the other way around. The ordering of events in time is not random—a fact we cannot foolishly ignore when doing data analysis.

Time series have many interesting features. For example, economies usually grow, and sometimes shrink, but rarely stay the same. Put simply, they show trends. Moreover, time series of economic variables often reflect seasonal patterns of consumption and human behavior. Americans, for example, heat their homes in January and go to the beach in July; they buy gifts in December and school supplies in September. These patterns recur year after year, and are non-ignorable in many longitudinal data sets.

Trends and seasonality

Let's get a sense of what time-series data can look like. In Figure 8.1 we see an example of data where we see a long-term increasing trend—sales are growing over time. There also seems to be a seasonal pattern; if you look closely, you'll notice that Quarters 3 and 4 tend to be higher, on average, than quarters 1 and 2 (though not always).

There are two simple ways for modeling long-term, year-by-year differences. First, we could treat the year as a categorical

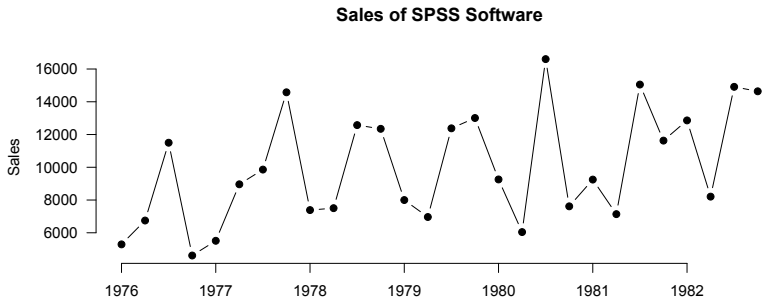

Figure 8.1: Sales in thousands of dollars of SPSS software over a period of seven years.

predictor and introduce dummy variables. Alternatively, we could introduce a time index as a predictor. That is, let $t = 1$ for your first observation, $t = 2$ for your second observation, $t = n$ for your n-th observation, and so forth. Then fit a model where the time index t enters linearly as a predictor:

$$\mathrm{E}(y_t) = \beta_0 + \beta_1 t + \epsilon_i \,.$$

This regression equation can easily accommodate other predictors.

In general, if the year-by-year differences seem random, and not part of an overall trend, then the dummy-variable approach is best. The dummy-variable approach will work, of course, even if there is a trend. It's disadvantage, however, is that it doesn't allow us to forecast future years, for it leaves us with no clear way of relating next year's dummy-variable coefficients to those from past years. There is no trend, just random year-to-year fluctuations.

If, on the other hand, the year-on-year changes exhibit a clear trend—as with the seemingly linear trend above—then it's best to use a time index as a predictor in the model. Not only does this fit the data with few coefficients to estimate, but it also allows you to forecase growth or decay into the future. We wouldn't want to extrapolate too far into the future using a simple linear model, of course. No trend holds indefinitely.

Periodicities, too, can be handled in at least two different ways. First, we can use dummy variables corresponding to the granularity of our observations. With quarterly data on a time series with an annual period, we would treat the quarter as a categorical variable and introduce three dummy variables. For monthly data on a time series with an annual period, we'd have twelve categories and

eleven variables. For annual data on a time series with a 17-year period, we'd have 17 categories and 16 dummy variables.[1] This logic holds regardless of the period and regardless of how granular your observations are. In general, if you have k observations per time period, then you will have $k-1$ dummy variables.

[1] A 17-year period may sound completely bizarre, but it does happen in nature: en.wikipedia.org/wiki/Magicicada

The second way of handling periodicities is to use periodic functions to transform the time index. The simplest periodic functions are sine and cosine, each of which have period 2π. If we have k observations per period, then we can describe the periodic component of the time series using the following transformed predictors:

$$\begin{aligned} x_{\cos}(t) &= \cos\left(\frac{2\pi t}{k}\right) \\ x_{\sin}(t) &= \sin\left(\frac{2\pi t}{k}\right). \end{aligned}$$

Remember that t is your time index, equal to 1 for your first observation, 2 for your second, and so on. Then we can fit a linear model incorporating both the trend and periodic component:

$$\mathrm{E}(y_i \mid t) = \beta_0 + \beta_1 t + \beta_2 x_{\cos}(t) + \beta_3 x_{\sin}(t) + \epsilon_i .$$

This looks complicated. But keep in mind that $x_{\cos}(t)$ and $x_{\sin}(t)$ are transformations, just like logs or powers, and are therefore no more difficult to incorporate into the model.

For example, if we fit a model with quarterly dummy variables to the SPSS sales data, we get the following results:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6033.35    1076.83   5.603 1.06e-05 ***
Time          168.37      52.23   3.223 0.003762 **
QuarterQ2   -1024.51    1183.07  -0.866 0.395447
QuarterQ3    4709.55    1186.52   3.969 0.000607 ***
QuarterQ4    2476.75    1192.26   2.077 0.049121 *
---
Residual standard error: 2211 on 23 degrees of freedom
Multiple R-squared: 0.6473,Adjusted R-squared: 0.586
F-statistic: 10.55 on 4 and 23 DF,  p-value: 5.263e-05
```

The fitted values for this regression will look like a wave with a rising trend.

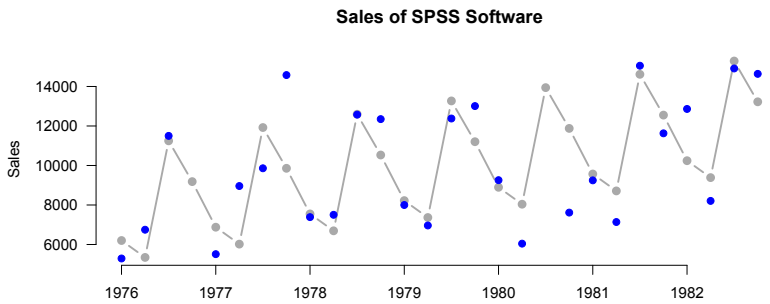

Figure 8.2: Real data (blue dots) plus fitted values (grey line) using quarterly dummy variables.

If we fit the sine/cosine model instead, we get only slight differences in the fitted values:

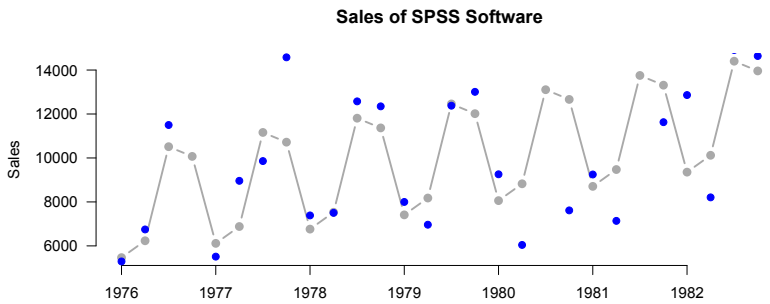

Figure 8.3: Real data (blue dots) plus fitted values (grey line) using sine and cosine to model the seasonality.

The chief difference here is that the sine/cosine terms enforce a bit more regularity from quarter to quarter, since the underlying periodic functions are smoothly varying. As always, there is a tradeoff: less flexibility, but greater simplicity (two parameters versus three). In this case, BIC says that the two models are essentially the same, meaning that there is little in it to choose between them.

Regressing one time series on another

Sometimes we'd like to regress one time series y_t on another time series x_t. For example:

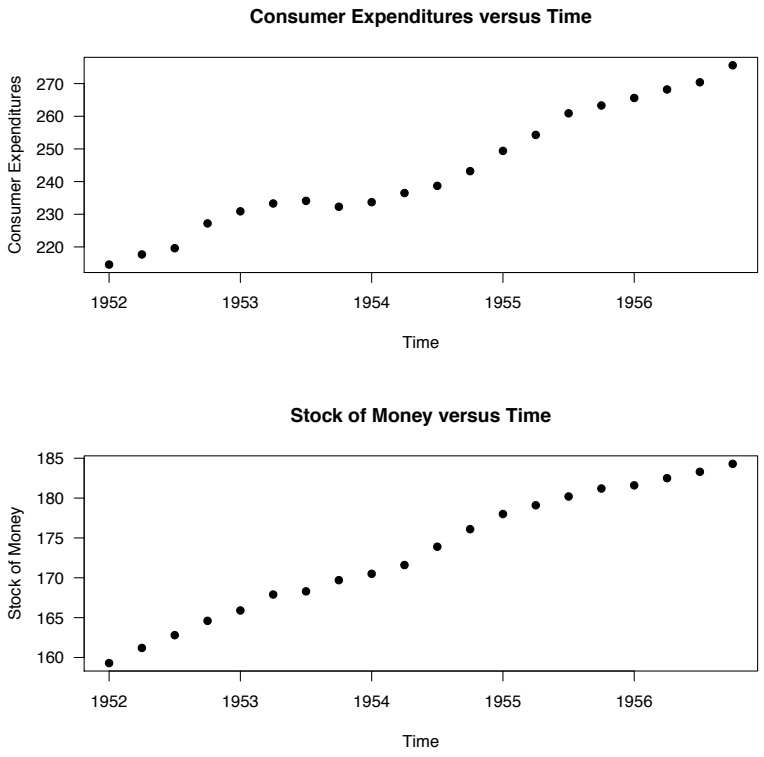

Figure 8.4: Two time series: consumer expenditures and the stock of money in the economy.

Both stock of money and consumer expenditures—observed here on a quarterly basis—are growing over time. (Stock of money is one measure of how much currency is available in the domestic economy.) Economists, moreover, might want to fit a model such as

$$y_t = \alpha + \beta_0 x_t + \epsilon_t \,,$$

where y_t is consumer expenditures, and x_t is the stock of money, in period t. In this model, β_0 would have an interpretation as a multiplier term. Knowing its value would allow us to answer an important macroeconomic question: "How many times per quarter does each dollar in the economy get spent?"

In this case, there is a side benefit to regressing y_t on x_t: we no longer need to explicitly account for the linear trend in y_t.

Instead, that trend can be related to the underlying behavior of the predictor time series x_t.

Lagged predictors

Sometimes y_t may depend upon *past* values of another time series x_t, in addition to (or instead of) the present value. This is because many forces do not make their effects known instantaneously, but rather are characterized by time lags. For example:

- Northern cities tend to spend a lot of money snowplowing their streets the day immediately after a big snow fall (lag: 1 day).
- People don't get colds immediately after being exposed to a cold virus; there is usually a latency period (lag: 1–2 days).
- Mosquito larvae mature about a week after a big rainfall leaves puddles of stagnant water lying around (lag: 1 week, or 7 days).
- Wall Street banks tend to give out exorbitant bonuses the quarter after they show a big profit (lag: 1 quarter).

We might imagine that the dependence graphs for these scenarios look like:

$$x_{t-j} \longrightarrow y_t \,,$$

where j is the lag period—that is, how many time periods it takes for y to "feel the effect of" changes in x.

Of course, there's nothing to stop us from incorporating more than one lag into a regression model. The first little bit of the dependence graph, for example, for a model with both a lag-0 term and a lag-1 term would look like

$$\begin{array}{ccccccccccc} x_1 & & x_2 & & x_3 & & x_4 & & x_5 & & \\ & \searrow & \downarrow & \searrow & \downarrow & \searrow & \downarrow & \searrow & \downarrow & \cdots \\ & & y_2 & & y_3 & & y_4 & & y_5 & & \end{array}$$

The general form of such a model is

$$y_t = \alpha + \sum_{j=1}^{p} \beta_j x_{t-j} + \epsilon_t \,.$$

Here p is the maximum number of lag terms you decide to include in the model.[2] There's nothing to stop you, moreover, from zeroing out some of these terms. For example, there might be a time series where you'd want to include a lag-4 term, but not the terms for lags 1–3.

[2] This choice of p should be guided using the same criteria for model choice in all regression problems—namely, those that allow you to balance the twin virtues of fit and simplicity.

These choices, like most choices in regression, are context dependent. For example, in the data on consumer expenditures versus stock of money, it turns out that the lag-1-only model, where

$$y_t = \alpha + \beta_1 x_{t-1} + \epsilon_t \,,$$

works a little bit better. It yields $R^2 = 0.982$, versus $R^2 = 0.947$ for the contemporaneous (that is, lag-0-only) model. The lag-2-only model, by contrast, yields $R^2 = 0.953$. (Since all of these models have the same number of free parameters in them, it is sensible to compare them using R^2.)

In a model with lagged predictors, your matrix of observations would look something like this:

Response		Lagged predictors			
y_t	x_t	x_{t-1}	x_{t-2}	$\cdots$	x_{t-p}
y_1	x_1	—	—	$\ldots$	—
y_2	x_2	x_1	—	$\ldots$	—
y_3	x_3	x_2	x_1	$\ldots$	—
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
y_{p+1}	x_{p+1}	x_p	x_{p-1}	$\ldots$	x_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_T	x_T	x_{T-1}	x_{T-2}	$\ldots$	x_{T-p}

The entries marked "—" indicate we don't have a predictor at that distant a lag because we've not yet observed the time series long enough. The entry at y_{p+1} is where the regression must begin, since that is the first point for which we have a complete set of predictors.

Notice how we call the intercept α so that we can maintain a consistent naming scheme for the coefficients associated with the lag terms: β_1 pairs with x_{t-1}, the lag-1 term; β_2 pairs with x_{t-2}, the lag-2 term; and so forth. This means, according to this naming scheme, that β_0 must go with the lag-0 term, x_t. If that's the case, then we must use a different—inevitably, Greek—letter for the intercept. Hence α. In the end, of course, it doesn't matter what naming scheme you use, as long as you are clear and consistent!

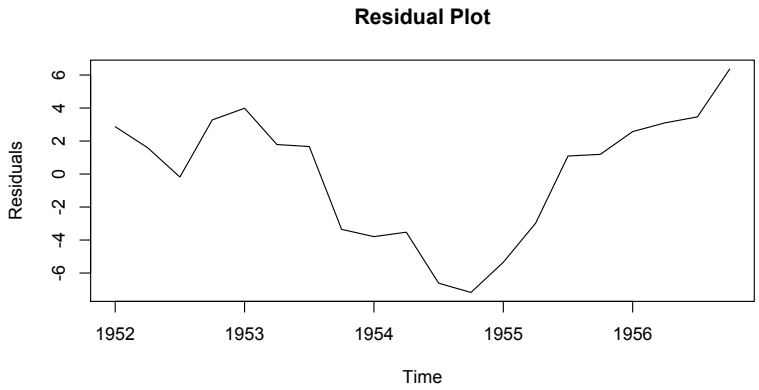

Figure 8.5: Residuals after regressing consumer expenditures upon stock of money.

Autoregressive processes

Let's return to the lag-0 model for consumer expenditures regressed upon the stock of money in the economy. When we fit this model and plot the quarterly residuals over time, we don't see a pretty picture (Figure 8.5). There is a clear pattern in the residuals, with at least two possible explanations:

1. y_t depends nonlinearly upon x_t. This might be the conclusion you'd jump to based on your past experience with regression and residual plots.
2. the residuals ϵ_t are not independent, but are instead correlated with each other.

Given the temporal nature of the data, this second possibility seems likely. For one thing, it just makes sense that successive quarters would be more similar than distantly separated quarters, even after accounting for the similarities that can be explained by the stock of money. Moreover, let's recall the interpretation of a residual: as the cumulative effect of all the predictors we've left out of the model. In this case, the other predictors would be other time series—and if all of these unaccounted-for predictors are correlated in time, then their cumulative effects might be correlated in time as well.

The alternative to independent residuals is to incorporate the assumption *autocorrelation*, where $\text{cor}(\epsilon_t, \epsilon_{t-1}) \neq 0$. We can formally test for autocorrelation by fitting a regression model to ϵ_t,

using past residuals as predictors. For example:

$$\epsilon_t = \phi\epsilon_{t-1} + \gamma_t \quad \text{where} \quad \gamma_t \sim \mathrm{N}(0, \tau^2)\,.$$

This is called an AR(1) model: "AR" for autoregressive, and "1" because that's how many lags (i.e. past residuals) one looks back to forecast the residual at time t. This is just a simple, one-variable linear regression, where last period's residual is used to predict this period's.

If the residuals from the original regression model are independent, then $\phi = 0$. For the regression of consumer expenditures versus stock of money, this hypothesis is roundly rejected by the data:

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.1430     0.5286   0.270     0.79
ResidualLag1   0.8783     0.1472   5.967 1.53e-05 ***
---
Residual standard error: 2.294 on 17 degrees of freedom
Multiple R-squared: 0.6768, Adjusted R-squared: 0.6578
F-statistic:  35.6 on 1 and 17 DF,  p-value: 1.531e-05
```

The lag-1 term appears to be significantly different from zero, meaning the presence of autocorrelation in the residuals is strongly supported. This calls the original regression into question, since it was based on the assumption of independent residuals.

AR(1) models

We'll soon return to the question of how to fit the original expenditure versus stock-of-money model with autocorrelated residuals. For now, let's think about the AR(1) model as a general description of a mean-reverting time series—that is, a time series whose highs and lows are temporary, and that fluctuates around its long-term mean.

Autoregressive models are best for situations where yesterday's result is strongly predictive of today's. For example, below we see the daily trading volumes for the year 2007 of the United States Oil Fund (ticker: USO), an exchange-traded fund that is constructed to mirror the movements in the price of West Texas Intermediate crude oil. (Recall that "volume" is just a measure of how many shares of an asset were exchanged on a given day, and is a mea-

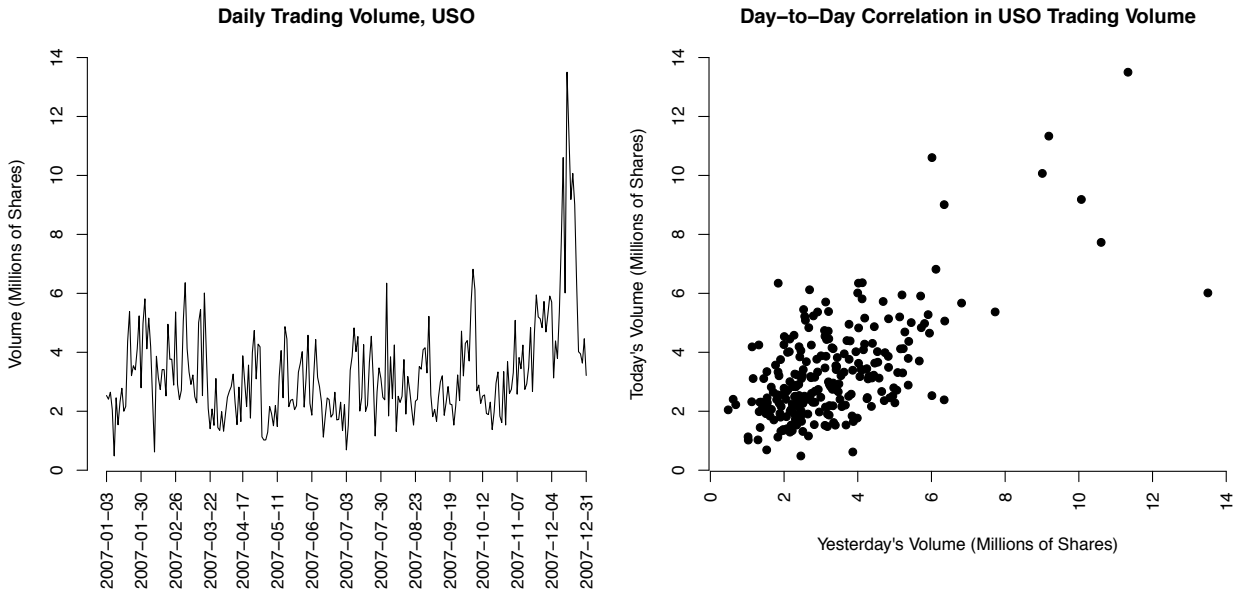

Figure 8.6: Daily trading volume of the United States Oil Fund (USO), 2007

sure both of liquidity and of the market's interest in trading a particular asset.)

In looking at the scatter plot on the right, it is obvious that the trading volume on one day is correlated with the trading volume on the previous day—in this case, $r = 0.65$. That means if you know today's volume, then tomorrow's volume is at least partially predictable. A simple AR(1) model is probably an inadequate description of any stock's trading volume, since there are probably some other covariates we could introduce into the model. But the presence of autocorrelation is undeniable.

Trading volumes aren't the only things in financial markets that show autocorrelation. Volatility, or the size of a stock's daily swings (regardless of direction), is often taken as a measure of uncertainty in the market, and also shows this property of autocorrelation (Figure 8.7).

Here we've taken the difference between USO's daily high and its daily low as a measure of volatility. The daily high is the highest price anybody paid on a given day for 1 share of USO, and the daily low is the lowest. When these numbers are far apart, it means that the asset's price must seen undergone considerable movement that day. Judging by the plots above, it appears that today's volatility is at least weakly predictive ($r = 0.25$) of tomor-

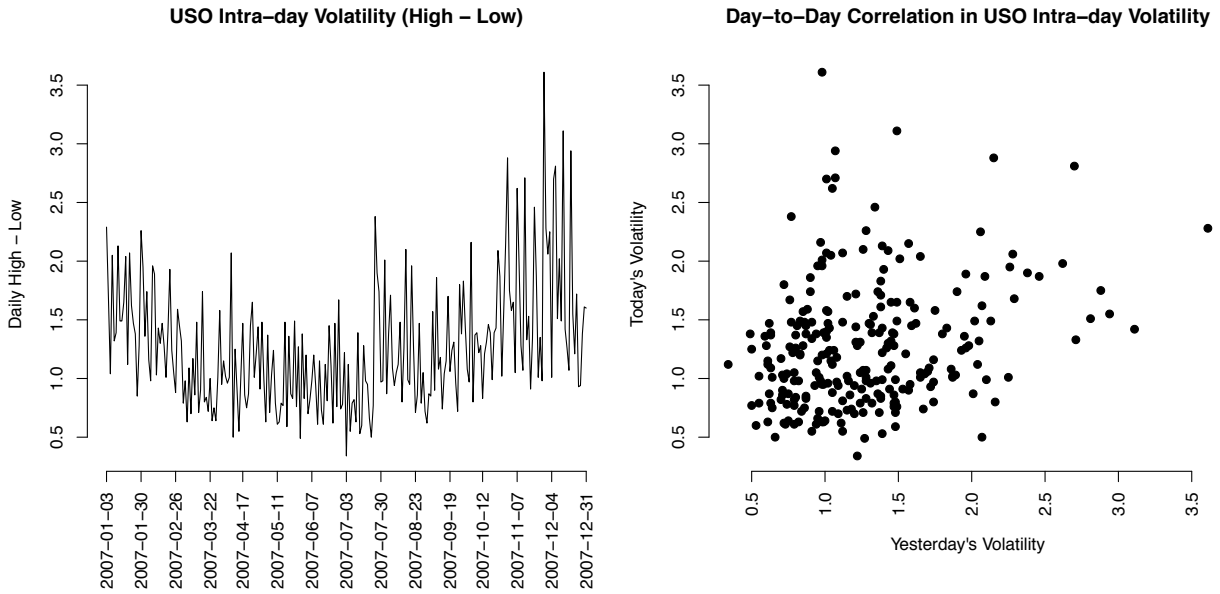

Figure 8.7: Intra-day volatility of USO in 2007.

row's volatility. Again, we wouldn't want to take the AR(1) model too seriously here, but there is no doubt that the volatilities on successive days are statistically dependent.

Recall the equation of an AR(1) model:

$$y_t = \phi y_{t-1} + \epsilon_t \quad \text{where} \quad \epsilon_t \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)\,.$$

The behavior of y_t depends heavily upon ϕ, which tells us the degree of correlation between observations separated by one lag:

$$\begin{aligned} \mathrm{cor}(y_t,\, y_{t-1}) &= \mathrm{cor}(\phi y_{t-1} + \epsilon_t,\, y_{t-1}) \\ &= \mathrm{cor}(\phi y_{t-1},\, y_{t-1}) \\ &= \phi \cdot \mathrm{cor}(y_{t-1},\, y_{t-1}) \\ &= \phi\,. \end{aligned}$$

Hence the closer ϕ is to -1 or 1, the more strongly correlated successive observations will be. The "noise" terms ϵ_t are often called *shocks* or *innovations*.

Suppose, for example, that there were no noise in the system, meaning that $\sigma^2 = 0$ and therefore $\epsilon_t = 0$ for all times t. Figure 8.8 shows the way in which subsequent values of y_t decay to zero at a rate determined by ϕ.

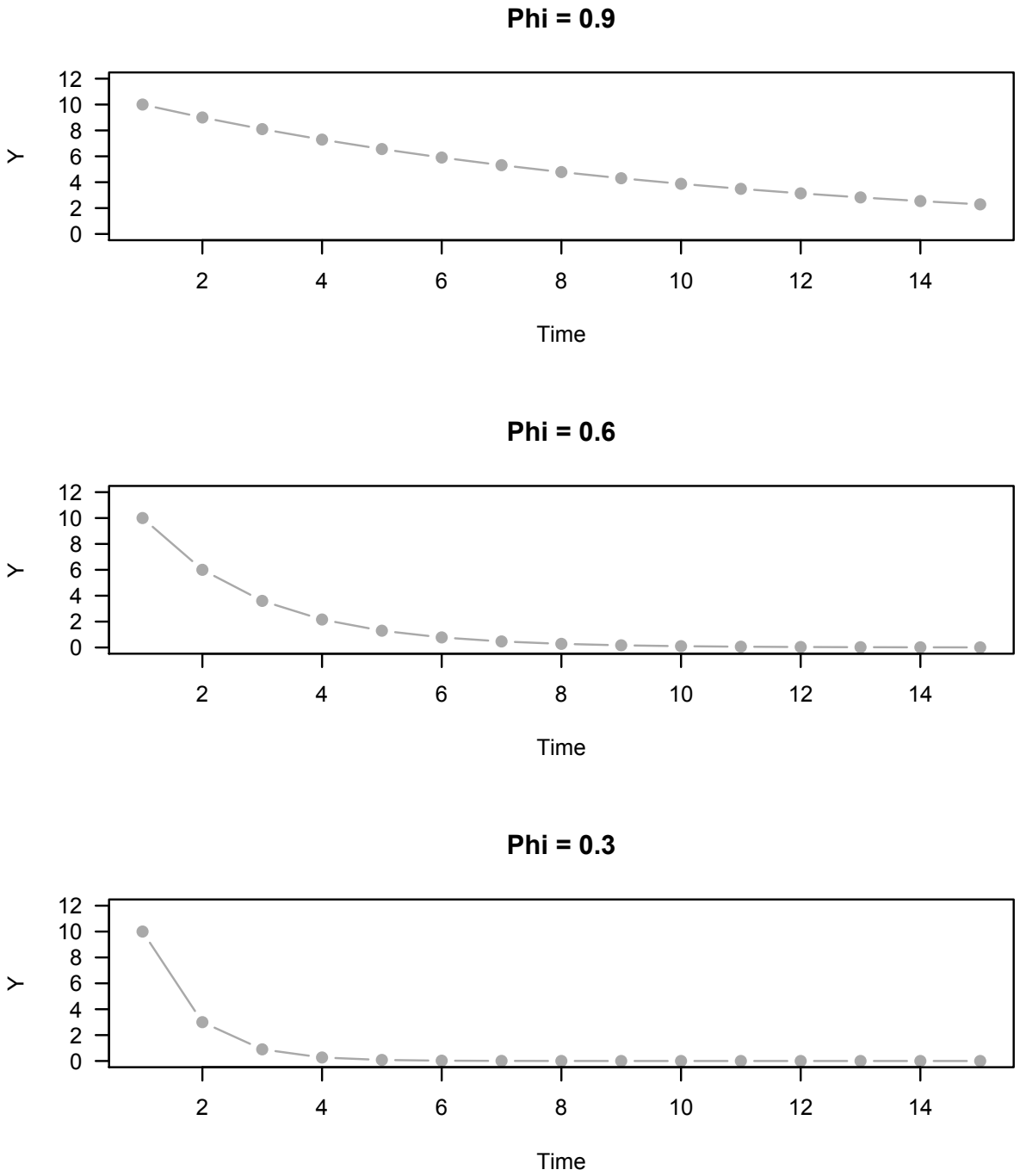

Figure 8.8: The rate at which different AR(1) time series decay to zero in the absence of random shocks ϵ_t.

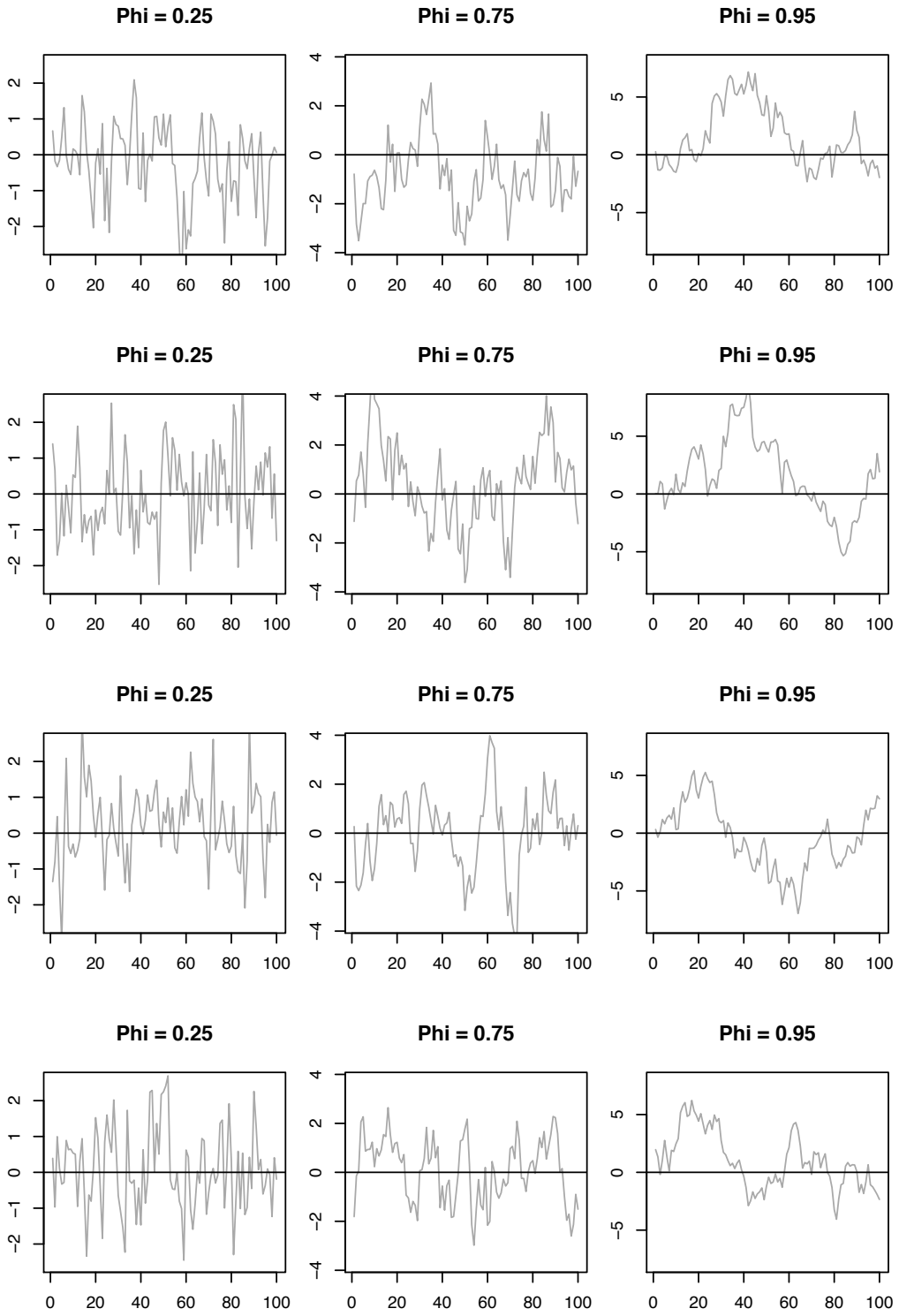

Figure 8.9: Examples of AR(1) processes with $\sigma^2 = 1$ and three different values of ϕ.

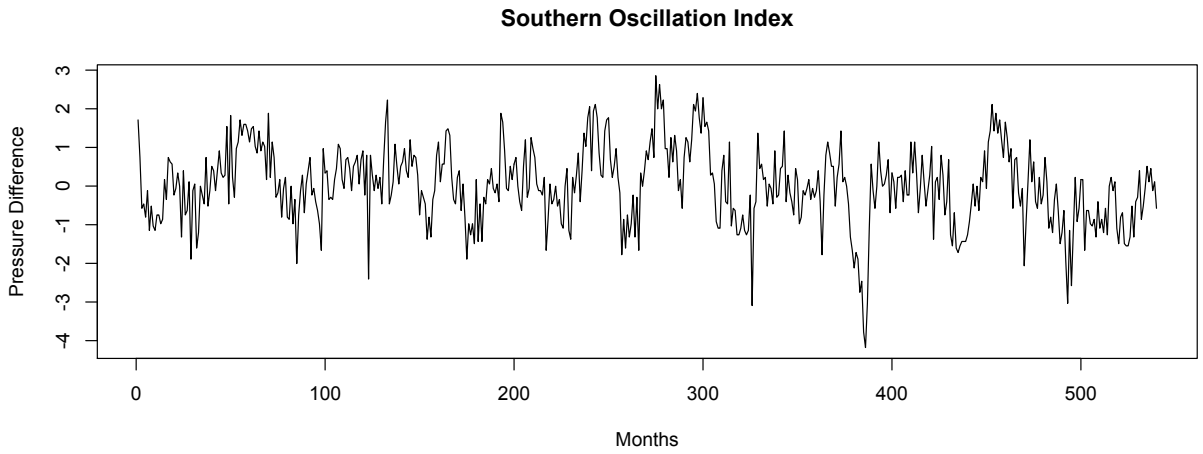

Figure 8.10: The Souther Oscillation Index during the 1990's.

Things get more interesting when we add noise to y_t at each stage. Now the time series can never decay all the way to zero, because it is constantly being "refreshed" by new random shocks ϵ_t. But the effect of these random shocks on future values of y_t decays quite rapidly with time—more rapidly if ϕ is close to zero, less rapidly if ϕ is close to 1. The daily trading volumes of USO clearly have a higher value of ϕ than the intra-day volatilities.

Figure 8.9, for example, shows 12 different AR(1) time series: four each for three different values of ϕ. When $\phi = 0.25$, y tends to bounce around its mean of zero quite freely; there is very little "stickiness" from y_t to y_{t+1}. When $\phi = 0.95$, on the other hand, the time series is quite sticky; it takes a long time for the time series to revert to its mean when it ends up far away from zero.

AR(p) models

An AR(1) model is the analogue of the single-lag predictor model, where $y_t = \beta_1 x_{t-1} + \epsilon_t$; just substitute y_{t-1} for x_{t-1} and you get the AR(1) instead.

Moreover, just as we could incorporate more lags of x_t for predicting y_t, we can also incorporate more lags of y_t itself. This leads to the so-called AR(p) model: where p denotes how many lags we go back to forecast the current value of y:

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \epsilon_t \,.$$

In words: an AR(p) model means that, in order to predict the value of y today, we use the value from yesterday, *and* from the day before, *and* from the day before, all the way until we get to p days back from the present. Then we stop. The AR(1) model is a special case of the AR(p) model, where $p = 1$. (In the case of weekly or yearly data, substitute "week" or "year" for "day" in this description.)

To fit an AR(p) model, simply copy your y series p times, and move it "over and down" in to generate your data matrix. Then run the regression as normal:

Response	Lagged values of y_t				
y_t	y_{t-1}	y_{t-2}	y_{t-3}	$\cdots$	y_{t-p}
y_1	—	—	—	$\cdots$	—
y_2	y_1	—	—	$\cdots$	—
y_3	y_2	y_1	—	$\cdots$	—
y_4	y_3	y_2	y_1	$\cdots$	—
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
y_{p+1}	y_p	y_{p-1}	y_{p-2}	$\cdots$	y_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_T	y_{T-1}	y_{T-2}	y_{T-3}	$\cdots$	y_{T-p}

As before, we must start the regression at y_{p+1}, since this is the first time point for which we have a complete set of predictors.

Whenever you see data that seems to behave "quasi-periodically," an AR(p) model might be a good bet. Take, for example, the plot in Figure 8.10 of the Southern Oscillation Index, or SOI, which is one quantitative measure of the cyclical climactic phenomenon known as El Niño. The SOI measures fluctuations in the difference between sea-level air pressures in Tahiti and Darwin, Australia. Above, we see the SOI measured on a weekly basis for every week during the decade spanning 1990 to 1999.

The index doesn't repeat, but it does "rhyme."[3] There are episodic highs and lows that come at fairly—but not perfectly—regular intervals. We could also fit an AR(1) model to this data, but this would fail to capture the quasi-periodicity, and turns out to have a worse value of BIC than either AR(2), AR(3), or AR(4).

For example, if we fit an AR(4) model, we get the following output:

```
Coefficients:
```

[3] Karl Marx: "History repeats itself; first as tragedy, second as farce." Mark Twain: "History doesn't repeat itself, but it does rhyme."

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -0.0007144  0.0318282  -0.022   0.9821
SoiLag4         0.0712519  0.0430507   1.655   0.0985 .
SoiLag3         0.0855428  0.0470633   1.818   0.0697 .
SoiLag2         0.1903357  0.0470159   4.048 5.93e-05 ***
SoiLag1         0.4351820  0.0432936  10.052  < 2e-16 ***
---
Residual standard error: 0.7369 on 531 degrees of freedom
Multiple R-squared: 0.4611, Adjusted R-squared: 0.457
```