Chapter 5

Trend and Seasonality

Consider a time-series situation:

$$y_t = x_t'\beta + \varepsilon_t, \ t = 1, ..., T$$

$$\varepsilon_t \sim \ iid \ N(0, \sigma^2).$$

Note that the disturbance density is assumed Gaussian for simplicity (for example in calculating density forecasts), but we could of course relax that assumption just as we did for cross-section forecasts.

5.1 The Forecasting the Right-Hand-Side Variables (FRV) Problem

In the future period of interest, $T + h$, it must be true that

$$y_{T+h} = x_{T+h}'\beta + \varepsilon_{T+h}.$$

Under quadratic loss the conditional mean forecast is optimal, and we immediately have

$$E(y_{T+h}|x_{T+h}) = x_{T+h}'\beta.$$

Suppose for the moment that we know the regression parameters. Forming the conditional expectation still requires knowing x_{T+h}, which we don't, so it seems that we're stuck.

We call this the "**forecasting-the-right-hand-side-variables (FRV) problem**." It *is* a problem, but it's not nearly as damaging as you might fear.

- We can abandon time series and only work in cross sections, where the FRV problem doesn't exist! But of course that's throwing out the baby with the bathwater and hardly a useful or serious prescription.

- We can move to scenario forecasts. **Time-series scenario forecasts** (also called **stress tests**, or **contingency analyses**), help us answer the "what if" questions that often arise. As with cross-section prediction, there is no FRV problem, and for precisely the same reason. For any given "scenario" x^*, we immediately have

$$E(y_{T+h}|x_{T+h} = x^*) = {x^*}'_{T+h}\beta.$$

  However, notwithstanding the occasional usefulness of scenario analyses, we generally don't want to make forecasts of y conditional upon assumptions about x; rather, we just simply want the best possible forecast of y.

- We can work with models involving lagged rather than current x, that is, models that relate y_t to x_{t-h} rather than relating y_t to x_t. This sounds ad hoc, but it's actually not, and we will have much more to say about it later.

- We can work with models for which we actually *do* know how to forecast x. In some important cases, the FRV problem doesn't arise at all, because the regressors are perfectly deterministic, so we know *exactly* what they'll be at any future time. The trend and seasonality models that we now discuss are leading examples.

5.2 Deterministic Trend

Time series fluctuate over time, and we often mentally allocate those fluctuations to unobserved underlying components, such as trends, seasonals, and cycles. In this section we focus on **trends**.[1] More precisely, in our general unobserved-components model,

$$y_t = T_t + S_t + C_t + \varepsilon_t,$$

we now include only trend and noise,

$$y_t = T_t + \varepsilon_t.$$

Trend is obviously pervasive. It involves slow, long-run, evolution in the variables that we want to model and forecast. In business, finance, and economics, trend is produced by slowly evolving preferences, technologies, institutions, and demographics.

We will study both **deterministic trend**, evolving in a perfectly predictable way, and **stochastic trend**, evolving in an approximately predictable way. We treat the deterministic case here, and we treat the stochastic case later in Chapter 5.3.

5.2.1 Trend Models

Sometimes series increase or decrease like a straight line. That is, sometimes a simple linear function of time,

$$T_t \ = \ \beta_0 \ + \ \beta_1 TIME_t,$$

provides a good description of the trend, in which case we speak of **linear trend**. We construct the variable $TIME$ artificially; it is called a "time

[1]Later we'll define and study **seasonals** and **cycles**. Not all components need be present in all observed series.

trend" or "**time dummy**." Time equals 1 in the first period of the sample, 2 in the second period, and so on. Thus, for a sample of size T, $TIME = (1, 2, 3, ..., T-1, T)$; put differently, $TIME_t = t$. β_0 is the **intercept**; it's the value of the trend at time $t = 0$. β_1 is the **slope**; it's positive if the trend is increasing and negative if the trend is decreasing. The larger the absolute value of β_1, the steeper the trend's slope. In business, finance, and economics, linear trends are typically (but not necessarily) increasing, corresponding to growth.

Sometimes trend appears nonlinear, or curved, as for example when a variable increases at an increasing or decreasing rate. Ultimately, we don't require that trends be linear, only that they be smooth. **Quadratic trend** models can potentially capture nonlinearities. Such trends are simply quadratic, as opposed to linear, functions of time,

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^2.$$

Linear trend emerges as a special (and potentially restrictive) case when $\beta_2 = 0$.[2]

A variety of different nonlinear quadratic trend shapes are possible, depending on the signs and sizes of the coefficients. In particular, if $\beta_1 > 0$ and $\beta_2 > 0$, the trend is monotonically, but nonlinearly, increasing, Conversely, if $\beta_1 < 0$ and $\beta_2 < 0$, the trend is monotonically decreasing. If $\beta_1 < 0$ and $\beta_2 > 0$ the trend is U-shaped, and if $\beta_1 > 0$ and $\beta_2 < 0$ the trend has an inverted U shape. See Figure 5.1. Keep in mind that quadratic trends are used to provide local approximations; one rarely has a U-shaped trend, for example. Instead, all of the data may lie on one or the other side of the "U."

Other types of nonlinear trend are sometimes appropriate. Sometimes, in particular, trend is nonlinear in levels but linear in logarithms. That's

[2] Higher-order polynomial trends are sometimes entertained, but it's important to use low-order polynomials to maintain smoothness.

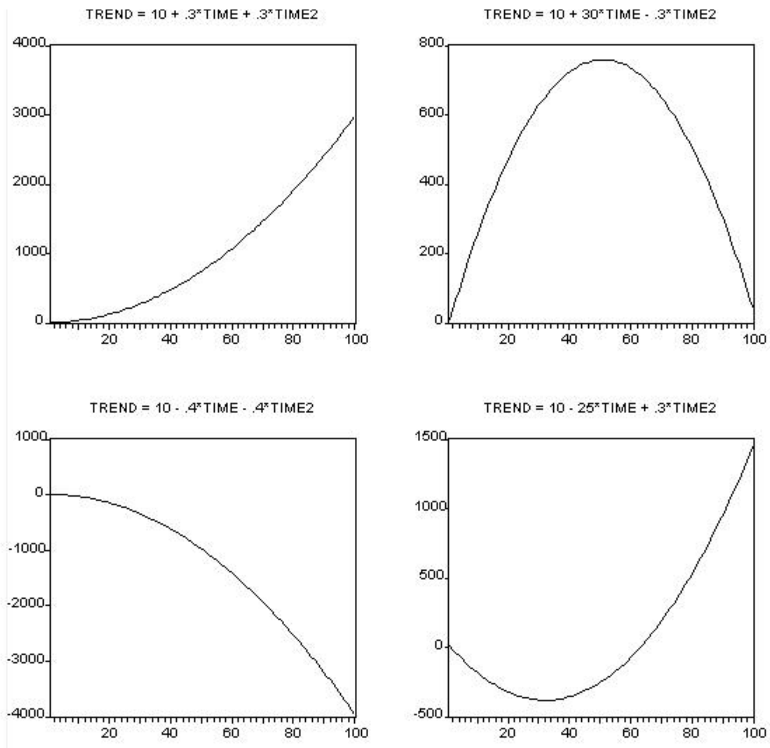


Figure 5.1: Various Shapes of Quadratic Trends

called **exponential trend**, or **log-linear trend**, and is very common in business, finance and economics.[3] It arises because economic variables often display roughly constant growth rates (e.g., three percent per year). If trend is characterized by constant growth at rate β_1, then we can write

$$T_t = \beta_0 e^{\beta_1 TIME_t}. \tag{5.1}$$

The trend is a nonlinear (exponential) function of time in levels, but in logarithms we have

$$\ln(T_t) = \ln(\beta_0) + \beta_1 TIME_t.$$

Thus, $\ln(T_t)$ is a linear function of time. As with quadratic trend, depending on the signs and sizes of the parameter values, exponential trend can achieve a variety of patterns, increasing or decreasing at an increasing or decreasing rate. See Figure 5.2.

[3]Throughout this book, logarithms are *natural* (base e) logarithms.

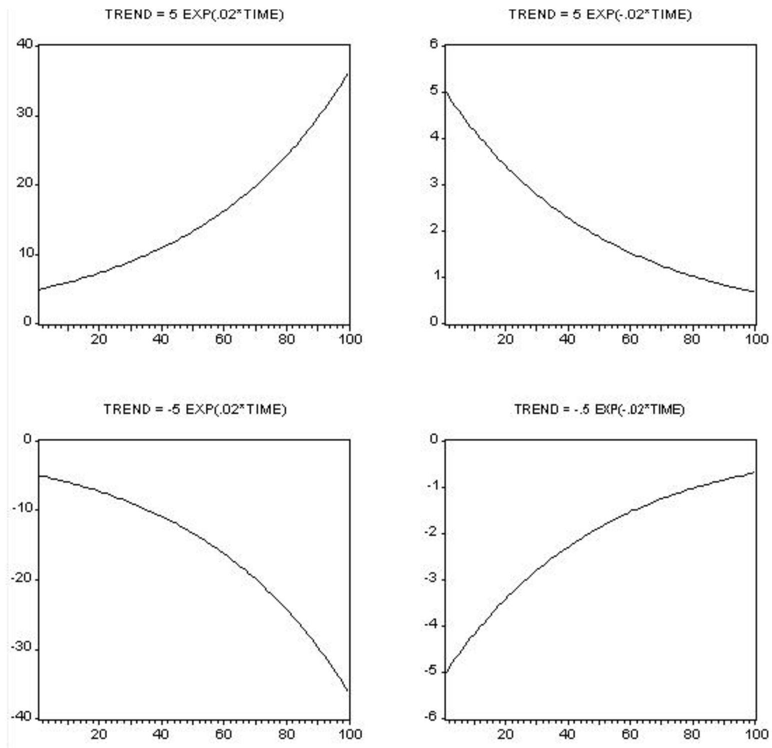


Figure 5.2: Various Shapes of Exponential Trends

It's important to note that, although qualitatively similar trend shapes can be achieved with quadratic and exponential trend, there are subtle differences between them. The nonlinear trends in some series are well approximated by quadratic trend, while the trends in other series are better approximated by exponential trend. Neither is necessarily "better" in general; rather, they're simply different, and which is better in any particular situation is ultimately an empirical matter.

5.2.2 Trend Estimation

Before we can estimate trend models, we need to create and store on the computer variables such as $TIME$ and its square. Fortunately we don't have to type the trend values (1, 2, 3, 4, ...) in by hand; rather, in most software packages a command exists to create $TIME$ automatically, after which we can immediately compute derived variables such as $TIME^2$. Because, for example, $TIME = 1, 2, ..., T$, $TIME^2 = 1, 4, ..., T^2$.

For the most part we fit our various trend models to data on a time series y using **ordinary least-squares regression**. In the linear and quadratic trend cases, the regressions are just simple OLS regressions. In an obvious notation, we run

$$y \to c, TIME$$

and

$$y \to c, TIME, TIME^2,$$

respectively, where c denotes inclusion of an intercept.

The exponential trend, in contrast, is a bit more nuanced. We can estimate it in two ways. First, because the nonlinear exponential trend is nevertheless linear in logs, we can estimate it by regressing $\ln y$ on an intercept and $TIME$,

$$\ln y \to c, TIME.$$

Note that c provides an estimate of $\ln \beta_0$ in equation (5.1) and so must be exponentiated to obtain an estimate of β_0. Similarly the fitted values from this regression are the fitted values of $\ln y$, so they must be exponentiated to get the fitted values of y.

Alternatively, we can proceed directly from the exponential representation and let the computer use numerical algorithms to find[4]

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{argmin} \sum_{t=1}^{T} \left[y_t - \beta_0 e^{\beta_1 TIME_t} \right]^2 .$$

This is called **nonlinear least squares**, or NLS.

NLS can be used to perform least-squares estimation for any model, including linear models, but in the linear case it's more sensible simply to use

[4] "**Argmin**" just means "the argument that minimizes."

OLS. Some intrinsically nonlinear forecasting models can't be estimated using OLS, however, but they can be estimated using nonlinear least squares. We resort to nonlinear least squares in such cases.[5] (We will encounter several in this book.) Even for models like exponential trend, which as we have seen can be transformed to linearity, estimation in levels by NLS is useful, because statistics like AIC and SIC can then be directly compared to those of linear and quadratic trend models.

5.2.3 Forecasting Trends

Suppose we're presently at time T, and we have a sample of historical data, $\{y_1, y_2, ..., y_T\}$ We want to use a trend model to forecast the h-step-ahead value of y. For illustrative purposes, we'll work with a linear trend, but the procedures are identical for quadratic and exponential trends.

First consider point forecasts. The linear trend model, which holds for any time t, is

$$y_t = \beta_0 + \beta_1 TIME_t + \varepsilon_t.$$

In particular, at time $T + h$, the future time of interest,

$$y_{T+h} = \beta_0 + \beta_1 TIME_{T+h} + \varepsilon_{T+h}.$$

Two future values of series appear on the right side of the equation, $TIME_{T+h}$ and ε_{T+h}. If $TIME_{T+h}$ and ε_{T+h} were known at time T, we could immediately crank out the forecast. In fact, $TIME_{T+h}$ *is* known at time T, because the artificially-constructed time variable is perfectly predictable; specifically, $TIME_{T+h} = T + h$. Unfortunately ε_{T+h} is not known at time T, so we replace it with an optimal forecast of ε_{T+h} constructed using information only up

[5] When we estimate by NLS, we use a computer to find the minimum of the sum of squared residual function directly, using numerical methods, by literally trying many (perhaps hundreds or even thousands) of different (β_0, β_1) values until those that minimize the sum of squared residuals are found. This is not only more laborious (and hence slow), but also less numerically reliable, as, for example, one may arrive at a minimum that is local but not global.

to time T.[6] Under the assumption that ε is simply independent zero-mean random noise, the optimal forecast of ε_{T+h} for any future period is 0, yielding the point forecast,[7]

$$y_{T+h,T} = \beta_0 + \beta_1 TIME_{T+h}.$$

The subscript "$T+h, T$" on the forecast reminds us that the forecast is for time $T+h$ and is made at time T. Note that the point forecast formula at which we arrived is not of practical use, because it assumes known values of the trend parameters β_0 and β_1. But it's a simple matter to make it operational – we just replace unknown parameters with their least squares estimates, yielding

$$\hat{y}_{T+h,T} = \hat{\beta}_0 + \hat{\beta}_1 TIME_{T+h}.$$

Now consider density forecasts under normality and ignoring parameter estimation uncertainty We immediately have the density forecast , $N(y_{T+h,T}, \sigma^2)$, where σ is the standard deviation of the disturbance in the trend regression. To make this operational, we use the density forecast $N(\hat{y}_{T+h,T}, \hat{\sigma}^2)$, where $\hat{\sigma}^2$ is the square of the standard error of the regression. Armed with the density forecast, we can construct any desired interval forecast. For example, the 95% interval forecast ignoring parameter estimation uncertainty is $y_{T+h,T} \pm 1.96\sigma$, where σ is the standard deviation of the disturbance in the trend regression. To make this operational, we use $\hat{y}_{T+h,T} \pm 1.96\hat{\sigma}$, where $\hat{\sigma}$ is the standard error of the regression.

We can use the simulation-based methods of Chapter 4 to dispense with the normality assumption and/or account for parameter-estimation uncertainty.

[6] More formally, we say that we're "projecting ε_{T+h} on the time-T information set."

[7] "Independent zero-mean random noise" is just a fancy way of saying that the regression disturbances satisfy the usual assumptions – they are identically and independently distributed.

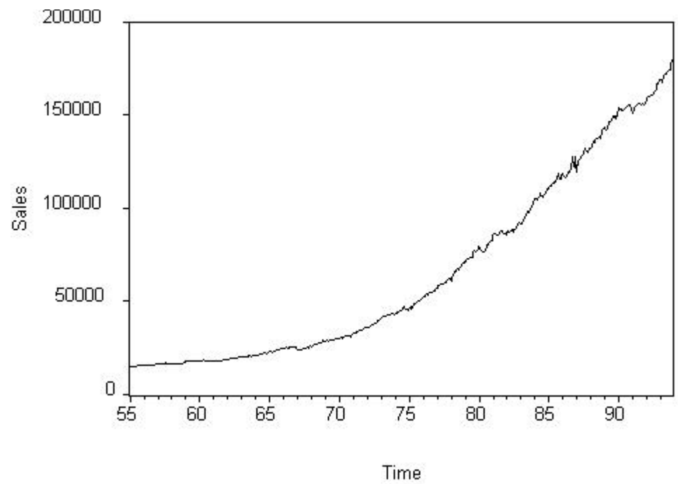


Figure 5.3: Retail Sales

5.2.4 Forecasting Retail Sales

We'll illustrate trend modeling with an application to forecasting U.S. current-dollar retail sales. The data are monthly from 1955.01 through 1994.12 and have been seasonally adjusted.[8] We'll use the period 1955.01-1993.12 to estimate our forecasting models, and we'll use the "holdout sample" 1994.01-1994.12 to examine their out-of-sample forecasting performance.

In Figure 5.3 we provide a time series plot of the retail sales data, which display a clear nonlinear trend and not much else. Cycles are probably present but are not easily visible, because they account for a comparatively minor share of the series' variation.

In Table 5.4a we show the results of fitting a linear trend model by regressing retail sales on a constant and a linear time trend. The trend appears highly significant as judged by the p-value of the t statistic on the time trend, and the regression's R^2 is high. Moreover, the Durbin-Watson statistic indicates that the disturbances are positively serially correlated, so that the disturbance at any time t is positively correlated with the disturbance at time $t-1$. In later chapters we'll show how to model such residual serial

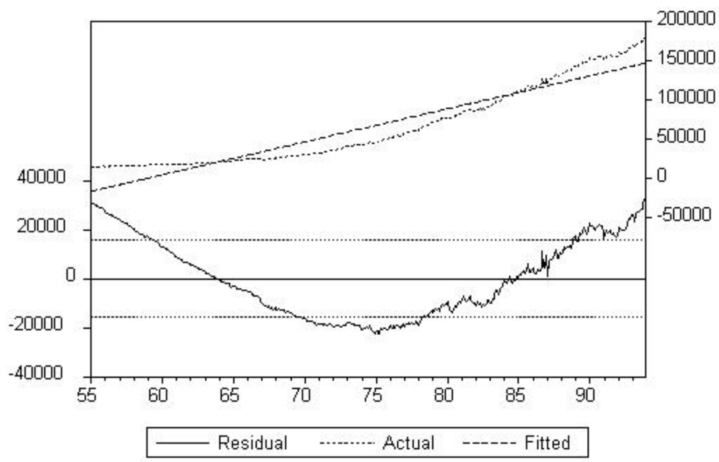
[8] When we say that the data have been "seasonally adjusted," we simply mean that they have been smoothed in a way that eliminates seasonal variation. We'll discuss seasonality in detail in Section 5.3.

Dependent Variable is RTRR
Sample: 1955:01 1993:12
Included observations: 468

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	-16391.25	1469.177	-11.15676	0.0000
TIME	349.7731	5.428670	64.43073	0.0000

R-squared	0.899076	Mean dependent var	65630.56
Adjusted R-squared	0.898859	S.D. dependent var	49889.26
S.E. of regression	15866.12	Akaike info criterion	19.34815
Sum squared resid	1.17E+11	Schwarz criterion	19.36587
Log likelihood	-5189.529	F-statistic	4151.319
Durbin-Watson stat	0.004682	Prob(F-statistic)	0.000000

(a) Retail Sales: Linear Trend Regression



(b) Retail Sales: Linear Trend Residual Plot

Figure 5.4: Retail Sales: Linear Trend

correlation and exploit it for forecasting purposes, but for now we'll ignore it and focus only on the trend.[9]

The residual plot in Figure 5.4b makes clear what's happening. The linear trend is simply inadequate, because the actual trend is nonlinear. That's one key reason why the residuals are so highly serially correlated – first the data are all above the linear trend, then below, and then above. Along with the residuals, we plot plus-or-minus one standard error of the regression, for visual reference.

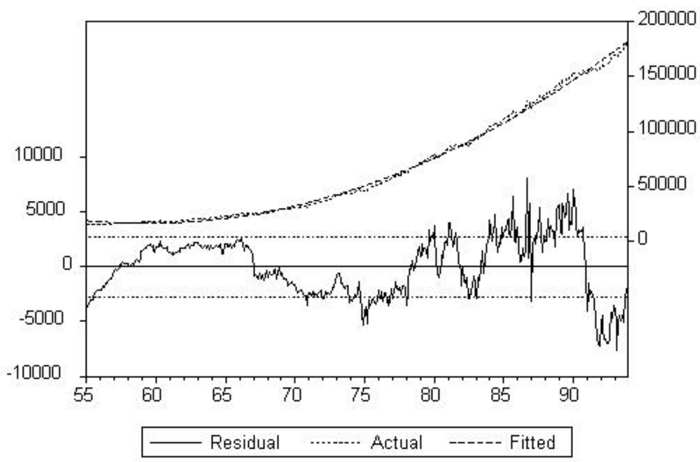
[9]Such residual serial correlation may, however, render the standard errors of estimated coefficients (and the associated t statistics) untrustworthy, and robust standard errors (e.g., Newey-West) can be used. In addition, AIC and SIC remain valid.

Dependent Variable is RTRR
Sample: 1955:01 1993:12
Included observations: 468

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	18708.70	379.9566	49.23905	0.0000
TIME	-98.31130	3.741388	-26.27669	0.0000
TIME2	0.955404	0.007725	123.6754	0.0000

R-squared	0.997022	Mean dependent var	65630.56
Adjusted R-squared	0.997010	S.D. dependent var	49889.26
S.E. of regression	2728.205	Akaike info criterion	15.82919
Sum squared resid	3.46E+09	Schwarz criterion	15.85578
Log likelihood	-4365.093	F-statistic	77848.80
Durbin-Watson stat	0.151089	Prob(F-statistic)	0.000000

(a) Retail Sales: Quadratic Trend Regression



(b) Retail Sales: Quadratic Trend Residual Plot

Figure 5.5: Retail Sales: Quadratic Trend

Table 5.5a presents the results of fitting a quadratic trend model. Both the linear and quadratic terms appear highly significant. R^2 is now almost 1. Figure 5.5b shows the residual plot, which now looks very nice, as the fitted nonlinear trend tracks the evolution of retail sales well. The residuals still display persistent dynamics (indicated as well by the still-low Durbin-Watson statistic) but there's little scope for explaining such dynamics with trend, because they're related to the business cycle, not the growth trend.

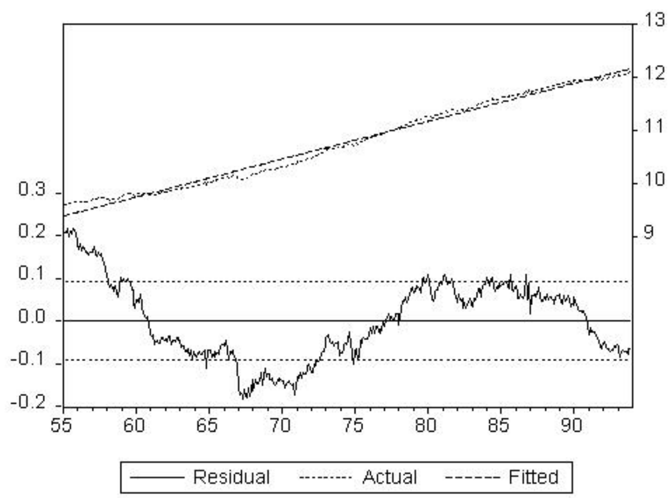
Now let's estimate a different type of nonlinear trend model, the exponential trend. First we'll do it by OLS regression of the log of retail sales on a constant and linear time trend variable. We show the estimation results and

Dependent Variable is LRTRR
Sample: 1955:01 1993:12
Included observations: 468

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	9.389975	0.008508	1103.684	0.0000
TIME	0.005931	3.14E-05	188.6541	0.0000

R-squared	0.987076	Mean dependent var	10.78072
Adjusted R-squared	0.987048	S.D. dependent var	0.807325
S.E. of regression	0.091879	Akaike info criterion	-4.770302
Sum squared resid	3.933853	Schwarz criterion	-4.752573
Log likelihood	454.1874	F-statistic	35590.36
Durbin-Watson stat	0.019949	Prob(F-statistic)	0.000000

(a) Retail Sales: Log Linear Trend Regression



(b) Retail Sales: Log Linear Trend Residual Plot

Figure 5.6: Retail Sales: Log Linear Trend

residual plot in Table 5.6a and Figure 5.6b. As with the quadratic nonlinear trend, the exponential nonlinear trend model seems to fit well, apart from the low Durbin-Watson statistic.

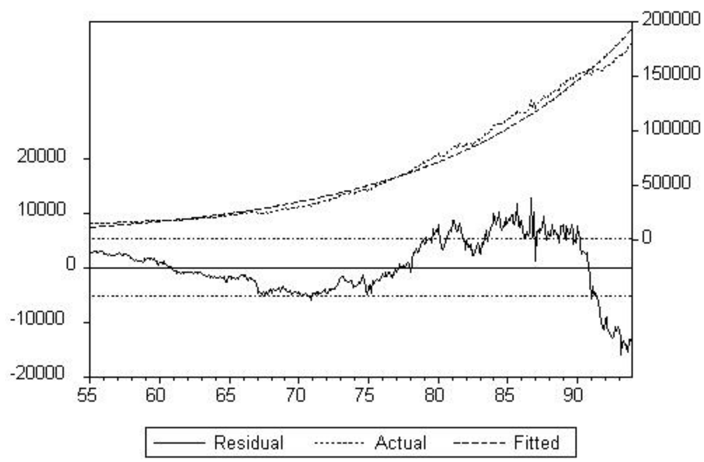
In sharp contrast to the results of fitting a linear trend to retail sales, which were poor, the results of fitting a linear trend to the *log* of retail sales seem much improved. But it's hard to compare the log-linear trend model to the linear and quadratic models because they're in levels, not logs, which renders diagnostic statistics like R^2 and the standard error of the regression incomparable. One way around this problem is to estimate the exponential trend model directly in levels, using nonlinear least squares. In Table 5.7a

Dependent Variable is RTRR
Sample: 1955:01 1993:12
Included observations: 468
Convergence achieved after 1 iterations
RTRR=C(1)*EXP(C(2)*TIME)

	Coefficient	Std. Error	T-Statistic	Prob.
C(1)	11967.80	177.9598	67.25003	0.0000
C(2)	0.005944	3.77E-05	157.7469	0.0000

R-squared	0.988796	Mean dependent var	65630.56
Adjusted R-squared	0.988772	S.D. dependent var	49889.26
S.E. of regression	5286.406	Akaike info criterion	17.15005
Sum squared resid	1.30E+10	Schwarz criterion	17.16778
Log likelihood	-4675.175	F-statistic	41126.02
Durbin-Watson stat	0.040527	Prob(F-statistic)	0.000000

(a) Retail Sales: Exponential Trend Regression - Nonlinear Least Squares



(b) Retail Sales: Exponential Trend Residual Plot

Figure 5.7: Retail Sales: Exponential Trend

and Figure 5.7b we show the nonlinear least squares estimation results and residual plot for the exponential trend model. The diagnostic statistics and residual plot indicate that the exponential trend fits better than the linear but worse than the quadratic.

Thus far we've been informal in our comparison of the linear, quadratic and exponential trend models for retail sales. We've noticed, for example, that the quadratic trend seems to fit the best. The quadratic trend model, however, contains one more parameter than the other two, so it's not surprising that it fits a little better, and there's no guarantee that its better fit on historical data will translate into better out-of-sample forecasting performance. (Recall

	Linear Trend	Quadratic Trend	Exponential Trend
AIC	19.35	15.83	17.15
SIC	19.37	15.86	17.17

Figure 5.8: Model Selection Criteria: Linear, Quadratic, and Exponential Trend Models

the parsimony principle.) To settle upon a final model, we examine the *AIC* or *SIC*, which we summarize in Table 5.8 for the three trend models.[10] Both the *AIC* and *SIC* indicate that nonlinearity is important in the trend, as both rank the linear trend last. Both, moreover, favor the quadratic trend model. So let's use the quadratic trend model.

Figure 5.9 shows the history of retail sales, 1990.01-1993.12, together with out-of-sample point and 95% interval extrapolation forecasts, 1994.01-1994.12. The point forecasts look reasonable. The interval forecasts are computed under the (incorrect) assumption that the deviation of retail sales from trend is random noise, which is why they're of equal width throughout. Nevertheless, they look reasonable.

In Figure 5.10 we show the history of retail sales through 1993, the quadratic trend forecast for 1994, *and* the realization for 1994. The forecast is quite good, as the realization hugs the forecasted trend line quite closely. All of the realizations, moreover, fall inside the 95% forecast interval.

For comparison, we examine the forecasting performance of a simple linear trend model. Figure 5.11 presents the history of retail sales and the out-of-sample point and 95% interval extrapolation forecasts for 1994. The point forecasts look very strange. The huge drop forecasted relative to the historical sample path occurs because the linear trend is far below the sample path by the end of the sample. The confidence intervals are very wide, reflecting the large standard error of the linear trend regression relative to the quadratic trend regression.

[10] It's important that the exponential trend model be estimated in levels, in order to maintain comparability of the exponential trend model *AIC* and *SIC* with those of the other trend models.

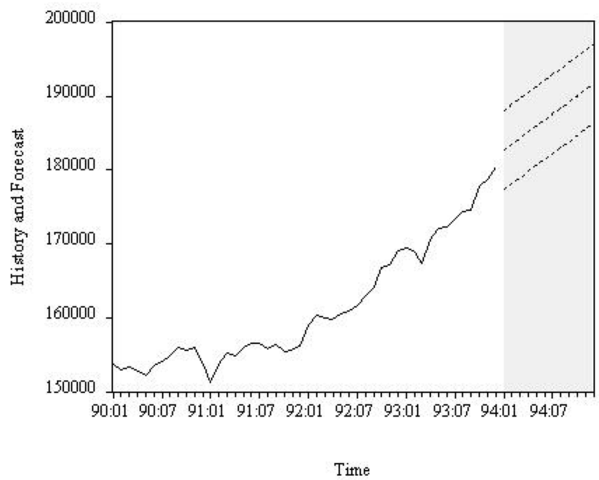


Figure 5.9: Retail Sales: Quadratic Trend Forecast

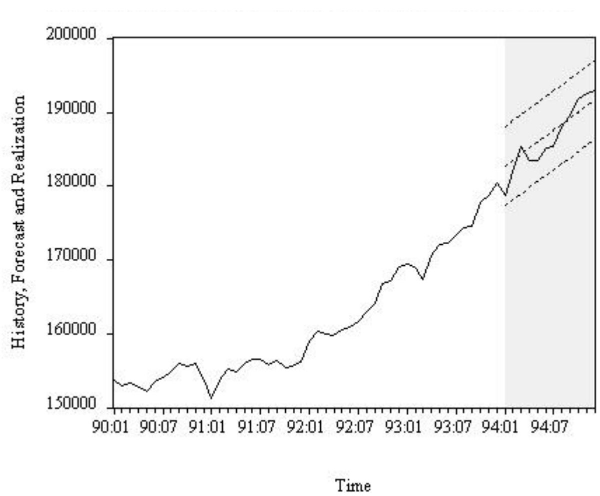


Figure 5.10: Retail Sales: Quadratic Trend Forecast and Realization

Finally, Figure 5.12 shows the history, the linear trend forecast for 1994, and the realization. The forecast is terrible – far below the realization. Even the very wide interval forecasts fail to contain the realizations. The reason for the failure of the linear trend forecast is that the forecasts (point and interval) are computed under the assumption that the linear trend model is actually the true DGP, whereas in fact the linear trend model is a very poor approximation to the trend in retail sales.

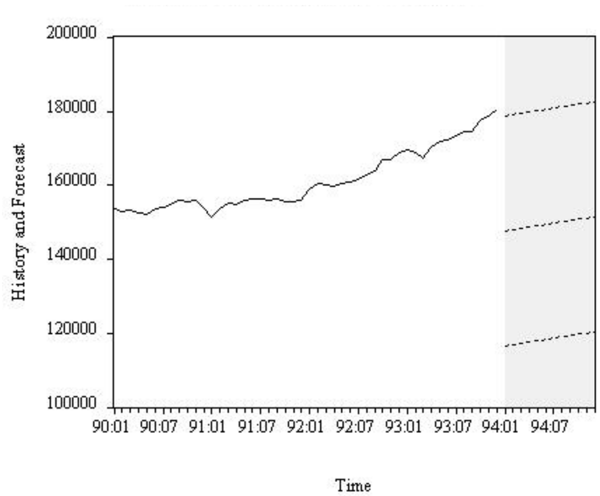


Figure 5.11: Retail Sales: Linear Trend Forecast

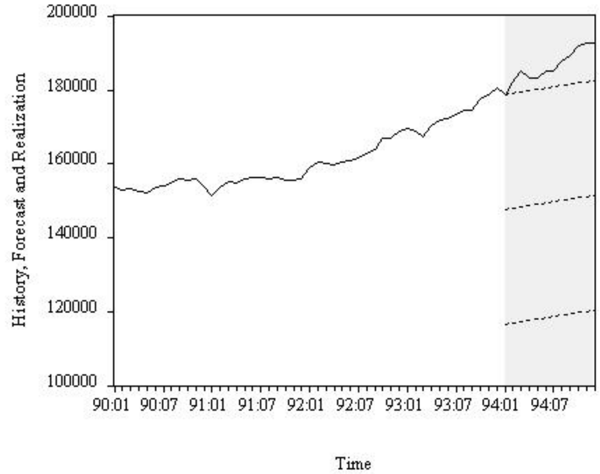


Figure 5.12: Retail Sales: Linear Trend Forecast and Realization

5.3 Deterministic Seasonality

Time series fluctuate over time, and we often mentally allocate those fluctuations to unobserved underlying components, such as trends, seasonals, and cycles. In this section we focus on **seasonals**.[11] More precisely, in our general unobserved-components model,

$$y_t = T_t + S_t + C_t + \varepsilon_t,$$

[11] Later we'll define and study cycles. Not all components need be present in all observed series.

we now include only the seasonal and noise components,

$$y_t = S_t + \varepsilon_t.$$

Seasonality involves patterns that repeat every year.[12] Seasonality is produced by aspects of technologies, preferences and institutions that are linked to the calendar, such as holidays that occur at the same time each year.

You might imagine that, although certain series are seasonal for obvious reasons, seasonality is nevertheless uncommon. On the contrary, and perhaps surprisingly, seasonality is pervasive in business and economics. Any technology that involves the weather, such as production of agricultural commodities, is likely to be seasonal. Preferences may also be linked to the calendar. For example, people want to do more vacation travel in the summer, which tends to increase both the price and quantity of summertime gasoline sales. Finally, social institutions that are linked to the calendar, such as holidays, are responsible for seasonal variation in many series. Purchases of retail goods skyrocket, for example, every Christmas season.

We will introduce both **deterministic seasonality** and **stochastic seasonality**. We treat the deterministic case here, and we treat the stochastic case later in Chapter .

5.3.1 Seasonal Models

A key technique for modeling seasonality is **regression on seasonal dummies**. Let s be the number of seasons in a year. Normally we'd think of four seasons in a year, but that notion is too restrictive for our purposes. Instead, think of s as the number of observations on a series in each year. Thus $s = 4$ if we have quarterly data, $s = 12$ if we have monthly data, $s = 52$ if we have weekly data, and so forth.

[12]Note therefore that seasonality is impossible, and therefore not an issue, in data recorded once per year, or less often than once per year.

Now let's construct **seasonal dummy variables**, which indicate which season we're in. If, for example, there are four seasons ($s = 4$), we create $D_1 = (1, 0, 0, 0, ...)$, $D_2 = (0, 1, 0, 0, ...)$, $D_3 = (0, 0, 1, 0, ...)$ and $D_4 = (0, 0, 0, 1, ...)$. D_1 indicates whether we're in the first quarter (it's 1 in the first quarter and zero otherwise), D_2 indicates whether we're in the second quarter (it's 1 in the second quarter and zero otherwise), and so on. At any given time, we can be in only one of the four quarters, so only one seasonal dummy is nonzero.

The deterministic seasonal component is

$$S_t = \sum_{i=1}^{s} \gamma_i D_{it}.$$

It is an intercept that varies in a deterministic manner over throughout the seasons within each year. Those different intercepts, the γ_i's, are called the **seasonal factors**; they summarize the seasonal pattern over the year.

In the absence of seasonality, the γ_i's are all the same, so we drop all the seasonal dummies and instead include an intercept in the usual way.

Crucially, note that the deterministic seasonal variation is perfectly predictable, just as with our earlier-studied deterministic trend variation.

5.3.2 Seasonal Estimation

Before we can estimate seasonal models we need to create and store on the computer the seasonal dummies D_i, $i = 1, ..., s$. Most software packages have a command to do it instantly.

We fit our seasonal models to data on a time series y using **ordinary least-squares regression**. We simply run

$$y \rightarrow D_1, ..., D_s.$$

We can also blend models to capture trend and seasonality simultaneously.

For example, we capture quadratic trend plus seasonality by running

$$y \rightarrow TIME, TIME^2, D_1, ..., D_s.$$

Note that whenever we include a full set of seasonal dummies, we drop the intercept, to avoid perfect multicollinearity. [13]

5.3.3 Forecasting Seasonals

Consider constructing an h-step-ahead point forecast, $y_{T+h,T}$ at time T. As with the pure trend model, there's no problem of forecasting the right-hand side variables, due to the special (perfectly predictable) nature of seasonal dummies, so point forecasts are easy to generate. The model is

$$y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t,$$

so that at time $T+h$,

$$y_{T+h} = \sum_{i=1}^{s} \gamma_i D_{i,T+h} + \varepsilon_{T+h}.$$

As with the trend model discussed earlier, we project the right side of the equation on what's known at time T (that is, the time-T information set, Ω_T) to obtain the forecast

$$y_{T+h,T} = \sum_{i=1}^{s} \gamma_i D_{i,T+h}.$$

There is no FRV problem, because $D_{i,T+h}$ is known with certainty, for all i and h. As always, we make the point forecast operational by replacing

[13] See also EPC ??.

unknown parameters with estimates,

$$\hat{y}_{T+h,T} = \sum_{i=1}^{s} \hat{\gamma}_i D_{i,T+h}.$$

To form density forecasts we again proceed precisely as in the trend model. If we assume that the regression disturbance is normally distributed, then the density forecast ignoring parameter estimation uncertainty is $N(y_{T+h,T}, \sigma^2)$, where σ is the standard deviation of the regression disturbance. The operational density forecast is then $N(\hat{y}_{T+h,T}, \hat{\sigma}^2)$, and the corresponding 95% interval forecast is $\hat{y}_{T+h,T} \pm 1.96\hat{\sigma}$.

We can use simulation-based methods from Chapter 4 to dispense with the normality assumption or account for parameter-estimation uncertainty.

5.3.4 Forecasting Housing Starts

We'll use the seasonal modeling techniques that we've developed in this chapter to build a forecasting model for housing starts. Housing starts are seasonal because it's usually preferable to start houses in the spring, so that they're completed before winter arrives. We have monthly data on U.S. housing starts; we'll use the 1946.01-1993.12 period for estimation and the 1994.01-1994.11 period for out-of-sample forecasting. We show the entire series in Figure 5.13, and we zoom in on the 1990.01-1994.11 period in Figure 5.14 in order to reveal the seasonal pattern in better detail.

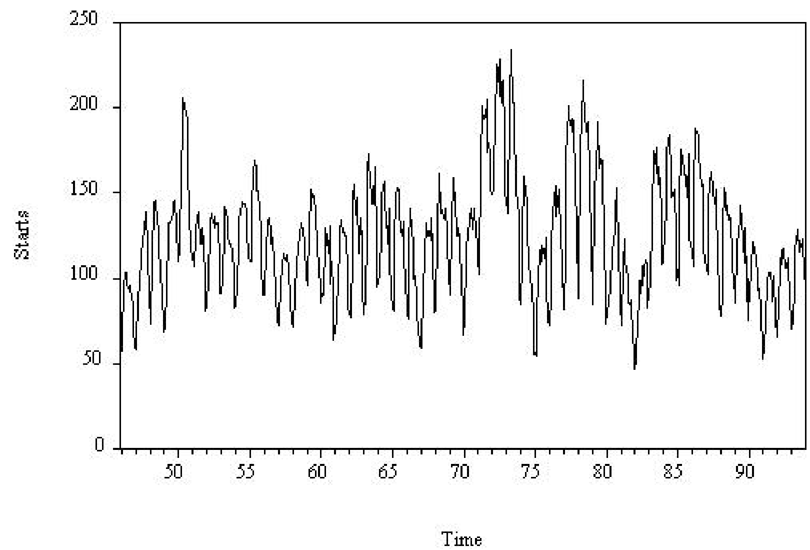


Figure 5.13: Housing Starts, 1946-1994

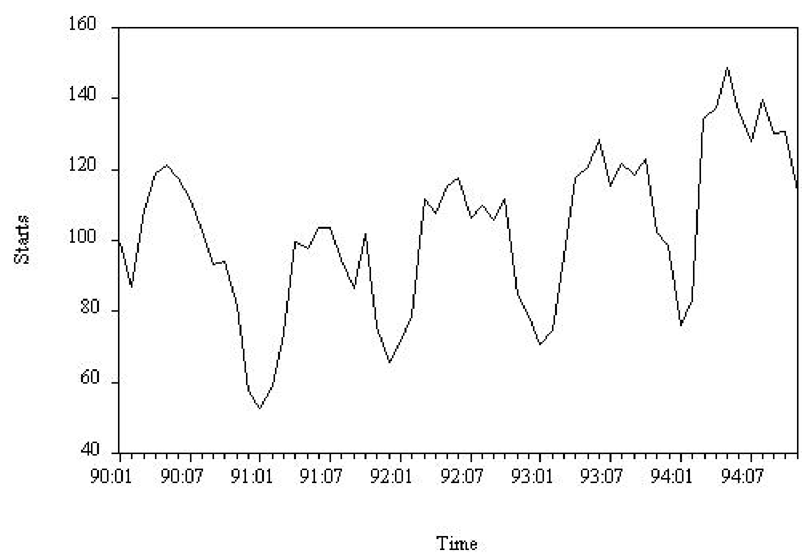


Figure 5.14: Housing Starts, 1946-1994 - Zoom on 1990-1994

The figures reveal that there is no trend, so we'll work with the pure seasonal model,

$$y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t.$$

Table 5.15a shows the estimation results. The twelve seasonal dummies account for more than a third of the variation in housing starts, as $R^2 = .38$. At least some of the remaining variation is cyclical, which the model is not designed to capture. (Note the very low Durbin-Watson statistic.)

The residual plot in Figure 5.15b makes clear the strengths and limitations of the model. First compare the actual and fitted values. The fitted values go through the same seasonal pattern every year – there's nothing in the model other than deterministic seasonal dummies – but that rigid seasonal pattern picks up a lot of the variation in housing starts. It doesn't pick up *all* of the variation, however, as evidenced by the serial correlation that's apparent in the residuals. Note the dips in the residuals, for example, in recessions (e.g., 1990, 1982, 1980, and 1975), and the peaks in booms.

The estimated seasonal factors are just the twelve estimated coefficients on the seasonal dummies; we graph them in Figure 5.16. The seasonal effects are very low in January and February, and then rise quickly and peak in May, after which they decline, at first slowly and then abruptly in November and December.

In Figure 5.17 we see the history of housing starts through 1993, together with the out-of-sample point and 95% interval extrapolation forecasts for the first eleven months of 1994. The forecasts look reasonable, as the model has evidently done a good job of capturing the seasonal pattern. The forecast intervals are quite wide, however, reflecting the fact that the seasonal effects captured by the forecasting model are responsible for only about a third of the variation in the variable being forecast.

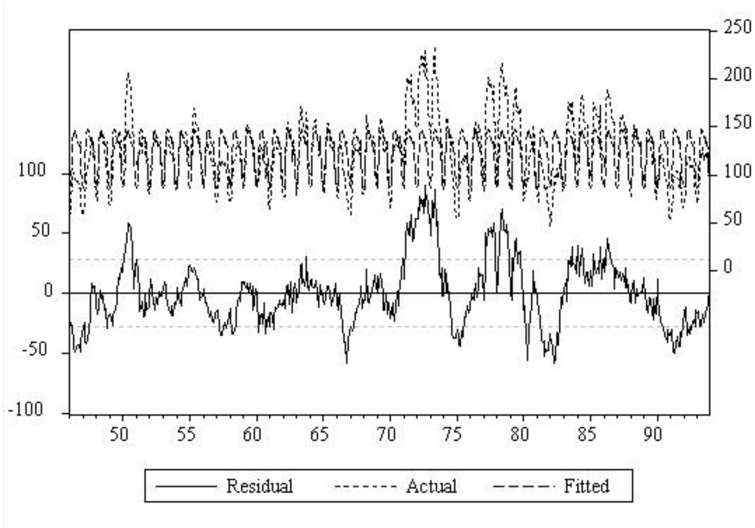
In Figure 5.18, we include the 1994 realization. The forecast appears highly accurate, as the realization and forecast are quite close throughout. Moreover, the realization is everywhere well within the 95% interval.

LS // Dependent Variable is STARTS
Sample: 1946:01 1993:12
Included observations: 576

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D1	86.50417	4.029055	21.47009	0.0000
D2	89.50417	4.029055	22.21468	0.0000
D3	122.8833	4.029055	30.49929	0.0000
D4	142.1687	4.029055	35.28588	0.0000
D5	147.5000	4.029055	36.60908	0.0000
D6	145.9979	4.029055	36.23627	0.0000
D7	139.1125	4.029055	34.52733	0.0000
D8	138.4167	4.029055	34.35462	0.0000
D9	130.5625	4.029055	32.40524	0.0000
D10	134.0917	4.029055	33.28117	0.0000
D11	111.8333	4.029055	27.75671	0.0000
D12	92.15833	4.029055	22.87344	0.0000

R-squared	0.383780	Mean dependent var	123.3944
Adjusted R-squared	0.371762	S.D. dependent var	35.21775
S.E. of regression	27.91411	Akaike info criterion	6.678878
Sum squared resid	439467.5	Schwarz criterion	6.769630
Log likelihood	-2728.825	F-statistic	31.93250
Durbin-Watson stat	0.154140	Prob(F-statistic)	0.000000

(a) Housing Starts: Seasonal Dummy Variables



(b) Housing Starts: Seasonal Dummy Variables, Residual Plot

Figure 5.15: Housing Starts: Seasonal Dummy Model

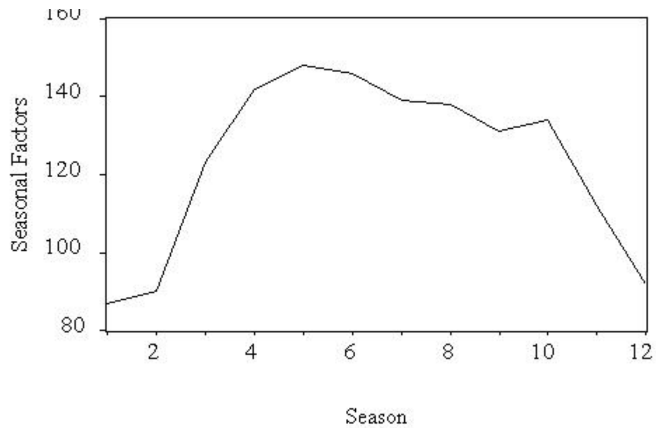


Figure 5.16: Housing Starts: Estimated Seasonal Factors

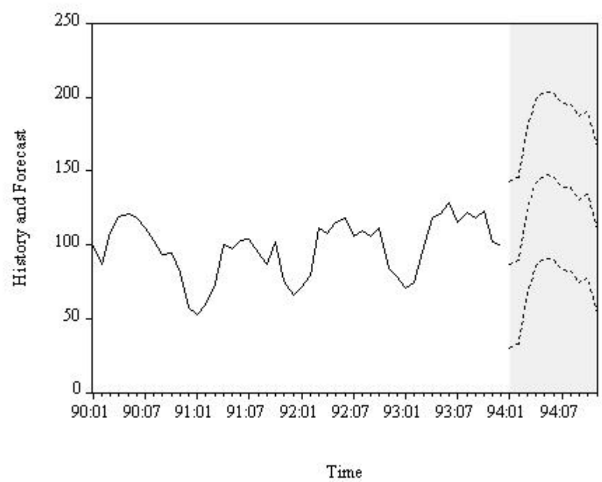


Figure 5.17: Housing Starts: Seasonal Model Forecast

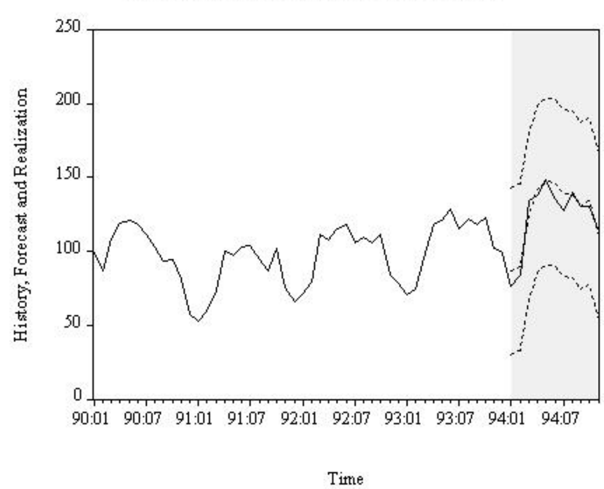


Figure 5.18: Housing Starts: Seasonal Model Forecast and Realization

5.4 Exercises, Problems and Complements

1. Calculating forecasts from trend models.

   You work for the International Monetary Fund in Washington DC, monitoring Singapore's real consumption expenditures. Using a sample of quarterly real consumption data (measured in billions of 2005 Singapore dollars), y_t, $t = 1990.1, ..., 2006.4$, you estimate the linear consumption trend model, $y_t = \beta_0 + \beta_1 TIME_t + \varepsilon_t$, where $\varepsilon_t \sim iidN(0, \sigma^2)$, obtaining the estimates $\hat{\beta}_0 = 0.51, \hat{\beta}_1 = 2.30$, and $\hat{\sigma}^2 = 16$. Based upon your estimated trend model, construct feasible point, interval and density forecasts for 2010.1.

2. Calendar span vs. observation count in trend estimation.

   Suppose it's the last day of the year. You are using a trend model to produce a 1-year-ahead (end-of-year) forecast of a stock (as opposed to flow) variable observed daily. Would you prefer to estimate your forecasting model using the most recent 500 daily observations (and then forecast 365 steps ahead) or 50 annual end-of-year observations (and then forecast 1 step ahead)? Discuss. In particular, if you prefer to use the 50 annual observations, why is that? Isn't 500 a much larger sample size than 50, so shouldn't you prefer to use it?

3. Mechanics of trend estimation and forecasting.

   Obtain from the web an upward-trending monthly series that interests you. Choose your series such that it spans at least ten years, and such that it ends at the end of a year (i.e., in December).

   a. What is the series and why does it interest you? Produce a time series plot of it. Discuss.

   b. Fit linear, quadratic and exponential trend models to your series. Discuss the associated diagnostic statistics and residual plots.

c. Select a trend model using the AIC and using the SIC. Do the selected models agree? If not, which do you prefer?

d. Use your preferred model to forecast each of the twelve months of the next year. Discuss.

e. The *residuals* from your fitted model are effectively a *detrended* version of your original series. Why? Plot them and discuss.

4. Properties of **polynomial trends**.

Consider a tenth-order deterministic polynomial trend:

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^2 + ... + \beta_{10} TIME_t^{10}.$$

a. How many local maxima or minima may such a trend display?

b. Plot the trend for various values of the parameters to reveal some of the different possible trend shapes.

c. Is this an attractive trend model in general? Why or why not?

d. How do you expect this trend to fit in-sample?

e. How do you expect this trend to forecast out-of-sample?

5. Seasonal adjustment.

One way to deal with seasonality in a series is simply to remove it, and then to model and forecast the **seasonally adjusted series**.[14] This strategy is perhaps appropriate in certain situations, such as when interest centers explicitly on forecasting **nonseasonal fluctuations**, as is often the case in macroeconomics. Seasonal adjustment is often inappropriate in business forecasting situations, however, precisely because interest typically centers on forecasting *all* the variation in a series, not just the nonseasonal part. If seasonality is responsible for a large part

[14] Removal of seasonality is called **seasonal adjustment**.

of the variation in a series of interest, the last thing a forecaster wants to do is discard it and pretend it isn't there.

a. Discuss in detail how you'd use dummy variable regression methods to seasonally adjust a series. (Hint: the seasonally adjusted series is closely related to the residual from the seasonal dummy variable regression.)

b. Search the Web (or the library) for information on the latest U.S. Census Bureau seasonal adjustment procedure, and report what you learned.

6. Fourier seasonality.

Thus far we have used seasonal dummies. We can also take a Fourier series approach, the benefits of which are two-fold. First, it produces a smooth seasonal pattern, which accords with the basic intuition that the progression through different seasons is gradual rather than discontinuous. Second, it promotes parsimony, which not only respects the parsimony principle but also enhances numerical stability in estimation.

The Fourier approach may be especially useful with high-frequency data. Consider, for example, seasonal daily data. For a variety of reasons, regression on more than three hundred daily dummies may not be appealing! So instead of using

$$S_t = \sum_{s=1}^{365} \gamma_i D_{it},$$

we can use

$$S_t = \sum_{p=1}^{P} \left(\delta_{c,p} \cos\left(2\pi p \frac{d_t}{365}\right) + \delta_{s,p} \sin\left(2\pi p \frac{d_t}{365}\right)\right),$$

where d_t is a repeating step function that cycles through 1, ..., 365. We can choose P using the usual model selection criteria. (Note that for simplicity we have dropped February 29 in leap years.)

5.5 Notes