Lecture Notes in Macroeconomic and Financial Forecasting (BSc course at UNISG)

Paul Söderlind[1]

26 January 2006

[1]University of St. Gallen. *Address:* s/bf-HSG, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. *E-mail:* Paul.Soderlind@unisg.ch. I thank Michael Fischer for comments and help. Document name: MFForecastAll.TeX.

Contents

1 Elementary Statistics

More advanced material is denoted by a star (*). It is not required reading.

1.1 Mean, Standard Deviation, Covariance and Correlation

The mean and variance of a series are estimated as

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t \text{ and } \widehat{\text{Var}}(x) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2 . \tag{1.1}$$

(Sometimes the variance has $T-1$ in the denominator instead. The difference is typically small.) The standard deviation (here denoted $\text{Std}(x_t)$), the square root of the variance, is the most common measure of volatility.

The mean and standard deviation are often estimated on rolling data windows (for instance, a "Bollinger band" is ± 2 standard deviations from a moving data window around a moving average—sometimes used in analysis of financial prices.)

The covariance of two variables (here x and y) is typically estimated as

$$\widehat{\text{Cov}}(x_t, z_t) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(z_t - \bar{z}) . \tag{1.2}$$

The correlation of two variables is then estimated as

$$\widehat{\text{Corr}}(x_t, z_t) = \frac{\widehat{\text{Cov}}(x_t, z_t)}{\widehat{\text{Std}}(x_t)\,\widehat{\text{Std}}(z_t)}, \tag{1.3}$$

where $\widehat{\text{Std}}(x_t)$ is an estimated standard deviation. A correlation must be between -1 and 1 (try to show it). Note that covariance and correlation measure the degree of *linear* relation only. This is illustrated in Figure 1.1.

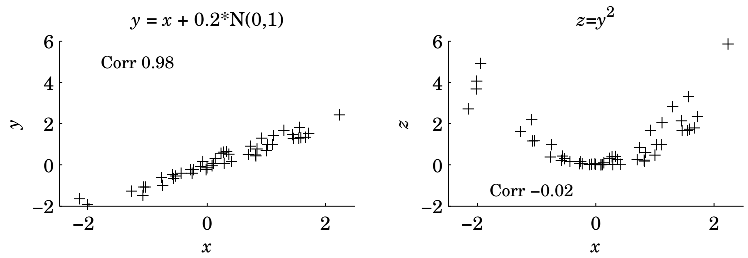


Figure 1.1: Example of correlations on an artificial sample

1.2 Least Squares

1.2.1 Simple Regression: Constant and One Regressor

The simplest regression model is

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \text{ where } \mathrm{E}\,\varepsilon_t = 0 \text{ and } \text{Cov}(x_t, \varepsilon_t) = 0. \tag{1.4}$$

Note the two very important assumptions: (*i*) the mean of the residual, ε_t, is zero; and (*ii*) the residual is not correlated with the regressor, x_t. If the regressor summarizes all the useful information we have in order to describe y_t, then these assumptions imply that we have no way of making a more intelligent guess of ε_t (even after having observed x_t) than that it will be zero.

Suppose you do not know β_0 or β_1, and that you have a sample of data at your hand: y_t and x_t for $t = 1, ..., T$. The LS estimator of β_0 and β_1 minimizes the loss function

$$(y_1 - b_0 - b_1 x_1)^2 + (y_2 - b_0 - b_1 x_2)^2 + \ldots = \sum_{t=1}^{T}(y_t - b_0 - b_1 x_t)^2 \tag{1.5}$$

by choosing b_0 and b_1 to make the loss function value as small as possible. The objective is thus to pick values of b_0 and b_1 in order to make the model fit the data as close as possible—where close is taken to be a small variance of the unexplained part (the residual), $y_t - b_0 - b_1 x_t$. See *Figure 1.2* for an example.

The solution to this minimization problem is fairly simple (involves just some mul-

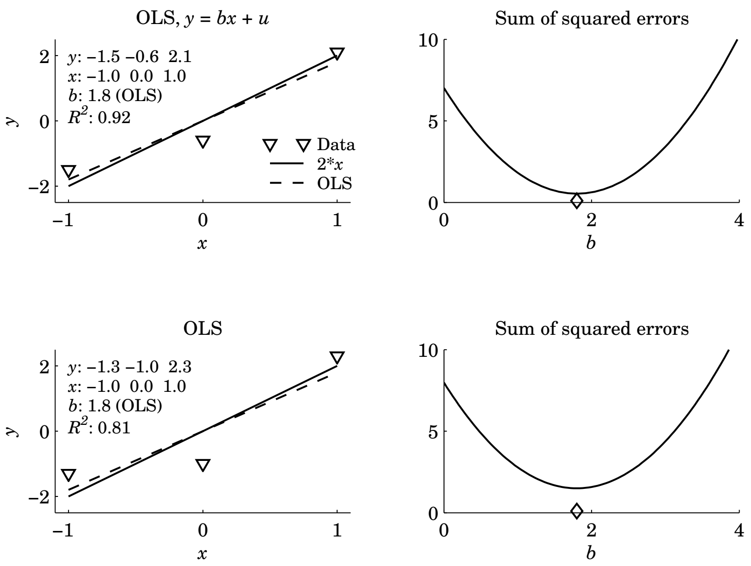


Figure 1.2: Example of OLS estimation

tiplications and summations), which makes it quick to calculate (even with an old computer). This is one of the reasons for why LS is such a popular estimation method (there are certainly many alternatives, but they typically involve more difficult computations). Another reason for using LS is that it produces the most precise estimates in many cases (especially when the residuals are normally distributed and the sample is large).

The estimates of the coefficients (denoted $\hat{\beta}_0$ and $\hat{\beta}_1$) will differ from the true values because we are not able to observe an undisturbed relation between y_t and x_t. Instead, the data provides a blurred picture because of the residuals in (1.4). The estimate is therefor only a (hopefully) smart guess of the true values. With some luck, the residuals are fairly stable (not volatile) or the sample is long so we can effectively average them out. In this case, the estimate will be precise. However, we are not always that lucky. (See Section 1.2.2 for more details.)

By plugging in the estimates in (1.4) we get

$$y_t = \underbrace{\hat{\beta}_0 + \hat{\beta}_1 x_t}_{\hat{y}_t} + \hat{\varepsilon}_t, \tag{1.6}$$

where $\hat{\varepsilon}_t$ are the *fitted* residuals. They differ from the true residuals in (1.4) since the estimated coefficients are not perfect, but LS will generate fitted residuals that have two important features: zero mean (if the regression includes a constant) and zero covariance with every regressor. This mimics the assumptions about the true residuals in (1.4). The systematic part of (1.6), $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$, is the fitted value of the regressor. This can be thought of as a "forecast" of y_t based on the information about x_t, using the estimated coefficients. The volatility of the fitted residuals (forecast error) will be an important indicator of the quality of this forecast.

1.2.2 Simple Regression: The Formulas and Why Coefficients are Uncertain*

Remark 1 *(First order condition for minimizing a differentiable function). We want to find the value of b in the interval $b_{low} \leq b \leq b_{high}$, which makes the value of the differentiable function $f(b)$ as small as possible. The answer is b_{low}, b_{high}, or the value of b where $df(b)/db = 0$.*

The first order conditions for minimum are that the partial derivatives of the loss function (1.5) with respect to b_0 and b_1 should be zero. To illustrate this, consider the simplest case where there is no constant—this makes sense only if both y_t and x_t have zero means (perhaps because the means have been subtracted before running the regression). The LS estimator picks a value of b_1 to minimize

$$L = (y_1 - b_1 x_1)^2 + (y_2 - b_1 x_2)^2 + \ldots = \sum_{t=1}^{T} (y_t - b_1 x_t)^2 \tag{1.7}$$

which must be where the derivative with respect to b_1 is zero

$$\frac{dL}{db_1} = -2(y_1 - b_1 x_1) x_1 - 2(y_2 - b_1 x_2) x_2 - \ldots = -2 \sum_{t=1}^{T} (y_t - b_1 x_t) x_t = 0. \tag{1.8}$$

The value of b_1 that solves this equation is the LS estimator, which we denote $\hat{\beta}_1$. This notation is meant to show that this is the LS *estimator* of the true, but unknown, parameter

β_1 in (1.4). Multiply (1.8) by $-1/(2T)$ and rearrange as

$$\frac{1}{T}\sum_{t=1}^{T} y_t x_t = \hat{\beta}_1 \frac{1}{T}\sum_{t=1}^{T} x_t x_t \text{ or}$$
$$\hat{\beta}_1 = \frac{\frac{1}{T}\sum_{t=1}^{T} y_t x_t}{\frac{1}{T}\sum_{t=1}^{T} x_t x_t}. \tag{1.9}$$

In this case, the coefficient estimator is the sample covariance (recall: means are zero) of y_t and x_t, divided by the sample variance of the regressor x_t (this statement is actually true even if the means are not zero and a constant is included on the right hand side—just more tedious to show it).

With more than one regressor, we get a first order condition similar to (1.8) for each of the regressors.

Note that the estimated coefficients are random variables since they depend on which particular sample that has been "drawn." This means that we cannot be sure that the estimated coefficients are equal to the true coefficients (β_0 and β_1 in (1.4)). We can calculate an estimate of this uncertainty in the form of variances and covariances of $\hat{\beta}_0$ and $\hat{\beta}_1$. These can be used for testing hypotheses about the coefficients, for instance, that $\beta_1 = 0$, and also for generating confidence intervals for forecasts (see below).

To see where the uncertainty comes from consider the simple case in (1.9). Use (1.4) to substitute for y_t (recall $\beta_0 = 0$)

$$\hat{\beta}_1 = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t (\beta_1 x_t + \varepsilon_t)}{\frac{1}{T}\sum_{t=1}^{T} x_t x_t}$$
$$= \beta_1 + \frac{\frac{1}{T}\sum_{t=1}^{T} x_t \varepsilon_t}{\frac{1}{T}\sum_{t=1}^{T} x_t x_t}, \tag{1.10}$$

so the OLS estimate, $\hat{\beta}_1$, equals the true value, β_1, plus the sample covariance of x_t and ε_t divided by the sample variance of x_t. One of the basic assumptions in (1.4) is that the covariance of the regressor and the residual is zero. This should hold in a very large sample (or else OLS cannot be used to estimate β_1), but in a small sample it may be slightly different from zero. Since ε_t is a random variable, $\hat{\beta}_1$ is too. Only as the sample gets very large can we be (almost) sure that the second term in (1.10) vanishes.

Alternatively, if the residual ε_t is very small (you have an almost perfect model), then

the second term in (1.10) is likely to be very small so the estimated value, $\hat{\beta}_1$, will be very close to the true value, β_1.

1.2.3 Least Squares: Goodness of Fit

The quality of a regression model is often measured in terms of its ability to explain the movements of the dependent variable.

Let $\hat{y}_t$ be the fitted (predicted) value of y_t. For instance, with (1.4) it would be $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$. If a constant is included in the regression (or the means of y and x are zero), then a measure of the *goodness of fit* of the model is given by

$$R^2 = \text{Corr}\left(y_t, \hat{y}_t\right)^2. \tag{1.11}$$

This is the squared correlation of the actual and predicted value of y_t.[1]

To get a bit more intuition for what R^2 represents, suppose (just to simplify) that the estimated coefficients equal the true coefficients, so $\hat{y}_t = \beta_0 + \beta_1 x_t$. In this case (1.11) is

$$R^2 = \text{Corr}\left(\beta_0 + \beta_1 x_t + \varepsilon_t, \beta_0 + \beta_1 x_t\right)^2. \tag{1.12}$$

Clearly, if the model is perfect so the residual is always zero ($\varepsilon_t = 0$), then $R^2 = 1$. On contrast, when the regression equation is useless, that is, when there are no movements in the systematic part ($\beta_1 = 0$), then $R^2 = 0$.

1.2.4 Least Squares: Forecasting

Suppose the regression equation has been estimated on the sample $1, 2..., T$. We now want to use the estimated model to make forecasts for $T+1$, $T+2$, etc. The hope is, of course, that the same model holds for the future as for the past.

Consider the simple regression (1.4), and suppose we know x_{T+1} and want to make a prediction of y_{T+1}. The expected value of the residual, ε_{T+1}, is zero, so our forecast is

$$\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 x_{T+1}, \tag{1.13}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the OLS estimates obtained from the sample $1, 2..., T$.

[1] It can be shown that the standard definition, $R^2 = 1 - \text{Var(residual)}/\text{Var(dependent variable)}$, is the same as (1.11).

We want to understand how uncertain this forecast is. The forecast error will turn out to be

$$y_{T+1} - \hat{y}_{T+1} = (\beta_0 + \beta_1 x_{T+1} + \varepsilon_{T+1}) - (\hat{\beta}_0 + \hat{\beta}_1 x_{T+1})$$
$$= \varepsilon_{T+1} + (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) x_{T+1}. \quad (1.14)$$

Although we do not know the components of this expression at the time we make the forecast, we understand the structure and can use that knowledge to make an assessment of the forecast uncertainty. If we are willing to assume that the model is the same in the future as on the sample we have estimated it on, then we can estimate the variance of the forecast error $y_{T+1} - \hat{y}_{T+1}$.

In the standard case, we pretend that we know the coefficients, even though they have been estimated. In practice, this means that we disregard the terms in (1.14) that involves the difference between the true and estimated coefficients. Then we can measure the uncertainty of the forecast as the variance of the fitted residuals $\hat{\varepsilon}_{t+1}$ (used as a proxy for the true residuals)

$$\text{Var}\left(\hat{\varepsilon}_t\right) = \hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2, \quad (1.15)$$

since $\hat{\varepsilon}_t$ has a zero mean (this is guaranteed in OLS if the regression contains a constant). This variance is estimated on the historical sample and, provided the model still holds, is an indicator of the uncertainty of forecasts also outside the sample. The larger $\hat{\sigma}^2$ is, the more of y_t depends on things that we cannot predict.

We can produce "confidence intervals" of the forecast. Typically we assume that the forecast errors are normally distributed with zero mean and the variance in (1.15). In this case, we can write

$$y_{T+1} = \hat{y}_{T+1} + \varepsilon_{T+1}. \quad (1.16)$$

The uncertainty of y_{T+1}, conditional on what we know when we make the point forecast $\hat{y}_{T+1}$ is due to the error term, which has an expected value of zero. Suppose ε_{T+1} is normally distributed, $\varepsilon_{T+1} \sim N(0, \sigma^2)$. In that case, the distribution of y_{T+1}, conditional on what we know when we make the forecast, is also normal

$$y_{T+1} \sim N(\hat{y}_{T+1}, \sigma^2). \quad (1.17)$$

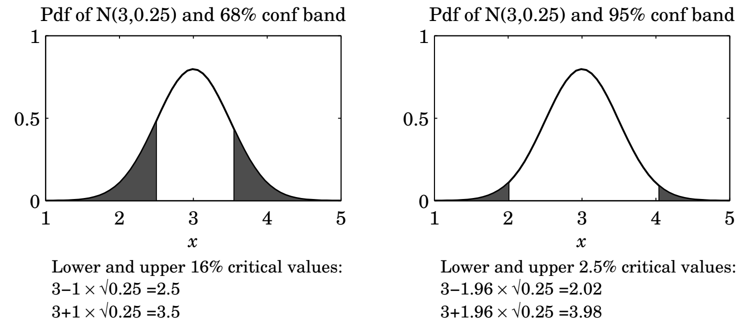


Figure 1.3: Creating a confidence band based on a normal distribution

We can therefore construct confidence intervals. For instance,

$$\hat{y}_{T+1} \pm 1.96\sigma \text{ gives a 95\% confidence interval of } y_{T+1}. \quad (1.18)$$

Similarly, $\hat{y}_{T+1} \pm 1.65\sigma$ gives a 90% confidence interval and $\hat{y}_{T+1} \pm \sigma$ gives a 68% confidence interval.

See *Figure 1.3* for an example.

Example 2 *Suppose* $\hat{y}_{T+1} = 3$*, and the variance is 0.25, then we say that there is a 68% probability that* y_{T+1} *is between* $3 - \sqrt{0.25}$ *and* $3 + \sqrt{0.25}$ *(2.5 and 3.5), and a 95% probability that it is between* $3 - 1.96\sqrt{0.25}$ *and* $3 + 1.96\sqrt{0.25}$ *(approximately, 2 and 4).*

The motivation for using a normal distribution to construct the confidence band is mostly pragmatic: many alternative distributions are well approximated by a normal distribution, especially when the error term (residual) is a combination of many different factors. (More formally, the averages of most variables tend to become normally distributed ash shown by the "central limit theorem.") However, there are situations where the symmetric bell-shape of the normal distribution is an unrealistic case, so other distributions need to be used for constructing the confidence band.

Remark 3 **(Taking estimation error into account.) In the more complicated case, we take into account the uncertainty of the estimated coefficients in our assessment of the* *forecast error variance. Consider the prediction error in (1.14), but note two things. First, the residual for the forecast period, ε_{T+1}, cannot be correlated with the past—and therefore not with the estimated coefficients (which where estimated on a sample of past data). Second, x_{T+1} is known when we make the forecast, so it should be treated as a constant. The result is then*

$$\mathrm{Var}\left(y_{T+1}-\hat{y}_{T+1}\right)=\mathrm{Var}\left(\varepsilon_{T+1}\right)+\mathrm{Var}\left(\beta_0-\hat{\beta}_0\right)+x_{T+1}^2\,\mathrm{Var}\left(\beta_1-\hat{\beta}_1\right)$$
$$+2x_{T+1}\,\mathrm{Cov}\left(\beta_0-\hat{\beta}_0,\beta_1-\hat{\beta}_1\right).$$

The term $\mathrm{Var}(\varepsilon_{T+1})$ *is given by (1.15). The true coefficients, β_0 and β_1 are constants. The last three terms can then be calculated with the help of the output from the OLS estimation.*

1.2.5 Least Squares: Outliers

Since the loss function in (1.5) is quadratic, a few outliers can easily have a very large influence on the estimated coefficients. For instance, suppose the true model is $y_t = 0.75x_t + \varepsilon_t$, and that the residual is very large for some time period s. If the regression coefficient happened to be 0.75 (the true value, actually), the loss function value would be large due to the ε_s^2 term. The loss function value will probably be lower if the coefficient is changed to pick up the y_s observation—even if this means that the errors for the other observations become larger (the sum of the square of many small errors can very well be less than the square of a single large error).

There is of course nothing sacred about the quadratic loss function. Instead of (1.5) one could, for instance, use a loss function in terms of the absolute value of the error $\Sigma_{t=1}^{T}|y_t-\beta_0-\beta_1 x_t|$. This would produce the Least Absolute Deviation (LAD) estimator. It is typically less sensitive to outliers. This is illustrated in Figure 1.4. However, LS is by far the most popular choice. There are two main reasons: LS is very easy to compute and it is fairly straightforward to construct standard errors and confidence intervals for the estimator. (From an econometric point of view you may want to add that LS coincides with maximum likelihood when the errors are normally distributed.)

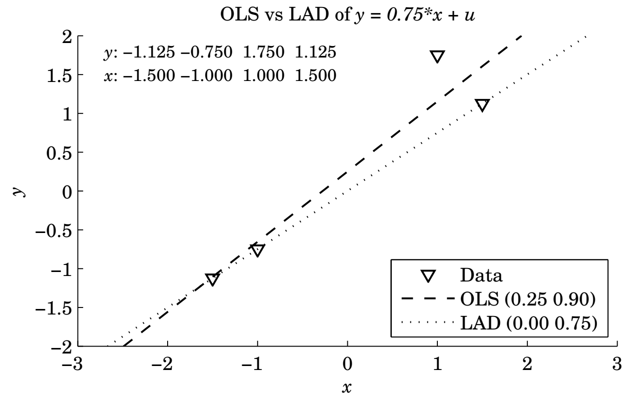


Figure 1.4: Data and regression line from OLS and LAD

1.3 Presenting Economic Data

Further reading: Diebold (2001) 3

This section contains some personal recommendations for how to present and report data in a professional manner. Some of the recommendations are quite obvious, others are a matter of (my personal) taste—take them with a grain of salt. (By reading these lecture notes you will readily see that am not (yet) able to live by my own commands.)

1.3.1 Figures (Plots)

See Figures 1.5–1.7 for a few reasonably good examples, and Figure 1.8 for a bad example. Here are some short comments on them.

- Figure 1.5 is a time series plot, which shows the development over time. The first subfigure shows how to compare the volatility of two series, and the second subfigure how to illustrate their correlation. This is achieved by changing the scales. Notice the importance of using different types of lines (solid, dotted, dashed,...) for different series.

- Figure 1.6 is a scatter plot. It shows no information about the development over time—only how the two variables are related. By changing the scale, we can either highlight the relative volatility or the finer details of the comovements.

- Figure 1.7 shows histograms, which is a simple way to illustrate the distribution of a variable. Also here, the trade off is between comparing the volatility of two series or showing finer details.

- Figure 1.8 is just a mess. Both subfigures should use curves instead, since this gives a much clearer picture of the development over time.

A few more remarks:

- Use clear and concise titles and/or captions. Don't forget to use labels on the x and y axes (unless the unit is obvious, like years). It is a matter of taste (or company policy...) if you place the caption above or below the figure.

- Avoid clutter. A figure with too many series (or other information) will easily become impossible to understand (except for the creator, possibly).

- Be careful with colours: use only colours that have different brightness. There are at least two reasons: quite a few people are colour blind, and you can perhaps not be sure that your document will be printed by a flashy new colour printer.

- If you want to compare several figures, keep the scales (of the axes) the same.

- Number figures consequtively: Figure 1, Figure 2,...

- In a text, place the figure close to where it is discussed. In the text, mention all the key features (results) of the figure—don't assume readers will find out themselves. Refer to the figure as Figure i, where i is the number.

- Avoid your own abbreviations/symbols in the figure, if possible. That is, even if your text uses y to denote real gross deomestic product, try to aviod using y in the figure. (Don't expect the readers to remember your abbreviations.) Depending on your audience, it might be okey to use well known abbreviations, for instance, GDP, CPI, or USD.

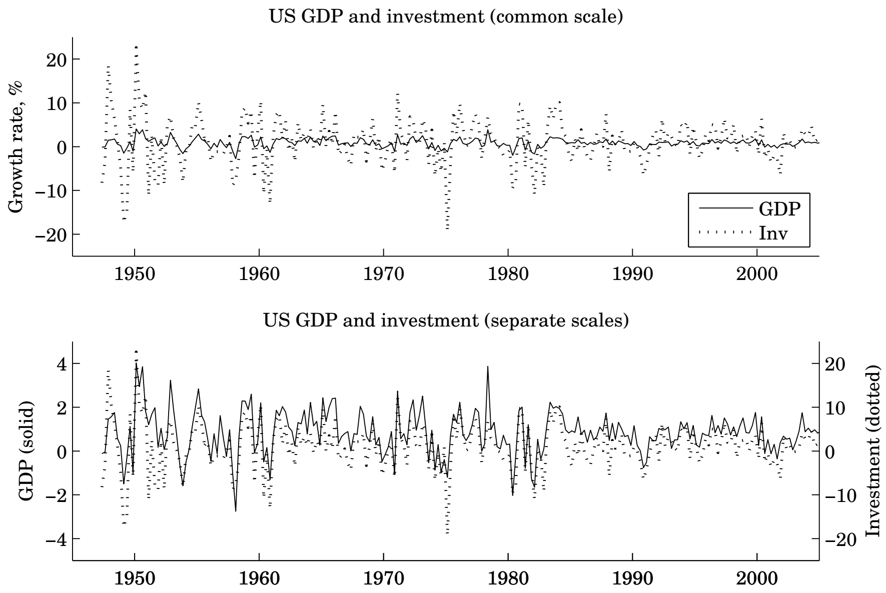


Figure 1.5: Examples of time series plots

- Remember who your audience is. For instance, if it is a kindergarten class, then you are welcome to use a pie chart with five bright colours—or even some sort of animation. Otherwise, a table with five numbers might look more professional.

1.3.2 Tables

Most of the rules for figures apply to tables too. To this, I would like to add: don't use a ridiculous number of digits after the decimal point. For instance, GDP growth should probably be reported as 2.1%, whereas 2.13% look less professional (since every one in the business know that there is no chance of measuring GDP growth with that kind of precision). As another example, the R^2 of a regression should probably be reported as 0.81 rather than 0.812, since no one cares about the third digit anayway.

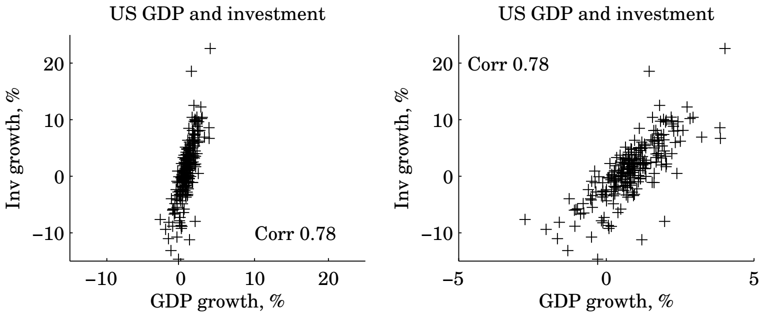


Figure 1.6: Examples of scatter plots

See *Table 1.1* for an example.

	1 quarter	2 quarters	4 quarters
GDP	−3.1	0.9	0.7
Private consumption	1.3	1.8	0.2
Government consumption	−0.2	−1.4	−2.6
Investments (business, fixed)	−21.7	−1.5	−0.6
Investments (residential)	−17.4	−14.5	−18.7
Exports	−38.3	−7.9	−5.3
Imports	−17.5	−4.9	−8.0
Money stock (M1)	−8.1	−6.4	−4.0
CPI	0.0	0.0	−0.0

Table 1.1: Mean errors in preliminary data on US growth rates, in basis points (%/100), 1965Q4–. Data 6 quarters after are used as proxies of the 'final' data.

2 Trends and Seasons

Main reference: Diebold (2001) 4–5; Evans (2003) 4–6; Newbold (1995) 17; or Pindyck and Rubinfeld (1998) 15

Further reading: Gujarati (1995) 22; The Economist (2000) 2 and 4

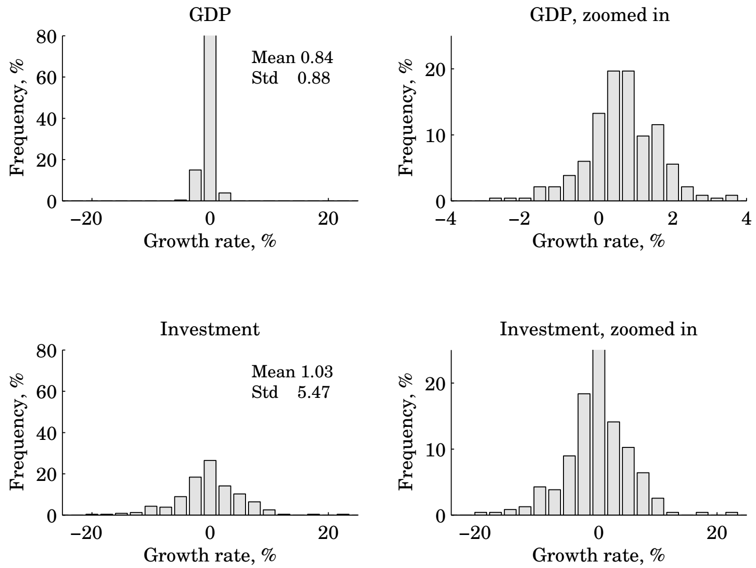


Figure 1.7: Examples of histogram plots

2.1 Trends, Cycles, Seasons, and the Rest

An economic time series (here denoted y_t) is often decomposed as

$$y_t = \text{trend} + \text{“cycle”} + \text{season} + \text{irregular}. \tag{2.1}$$

The reason for the decomposition is that we have very different understanding of and interest in, say, the decade-to-decade changes compared to the quarter-to-quarter changes. The exact definition of the various components will therefore depend on which series we are analyzing—and for what purpose. In most macroeconomic analyses a "trend" spans at least a decade, a (business) cycle lasts a few years, and the season is monthly or quarterly. See *Figure 2.1* for an example which shows both a clear trend, cycle, and a seasonal pattern. In contrast, "technical analysis" of the stock market would define a trend as the overall movements over a week or month.

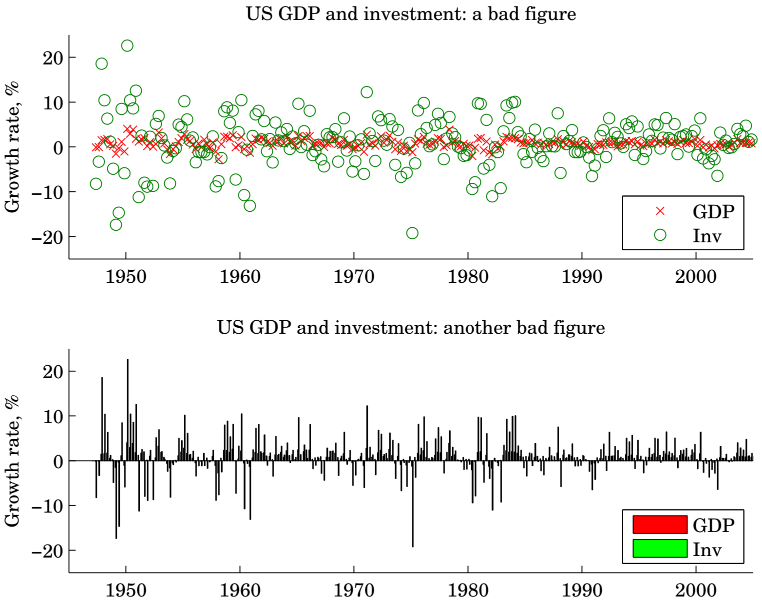


Figure 1.8: Examples of ugly time series plots

It is a common practice to split up the series into its components—and then analyze them separately. Figure 2.1 illustrates that simple transformations highlight the different components.

Sometimes we choose to completely suppress some of the components. For instance, in macro economic forecasting we typically work with seasonally adjusted data—and disregard the seasonal component. In development economics, the focus is instead on understanding the trend. In other cases, different forecasting methods are used for the different components and then the components are put together to form a forecast of the original series.

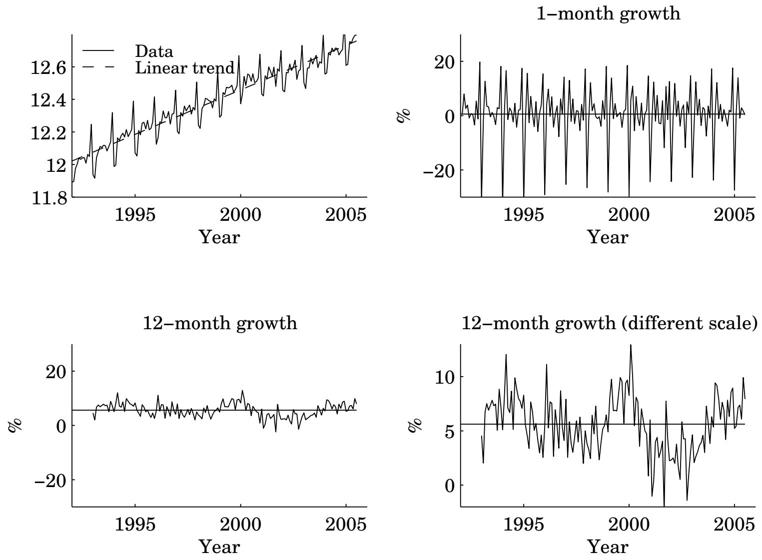


Figure 2.1: Seasonal pattern in US retail sales, current USD

2.2 Trends

This section discusses different ways to extract a trend from a time series.

Let $\tilde{y}_t$ denote the trend component of a series y_t. Consider the following trend models

$$
\begin{aligned}
\text{linear}: \ \tilde{y}_t &= a + bt, \\
\text{quadratic}: \ \tilde{y}_t &= a + bt + ct^2, \\
\text{Exponential}: \ \tilde{y}_t &= ae^{bt} \\
\text{Moving average smoothing}: \ \tilde{y}_t &= \theta_0 y_t + \theta_1 y_{t-1} + \ldots + \theta_q y_{t-q}, \ \textstyle\sum_{s=0}^{q} \theta_s = 1 \qquad (2.2) \\
\text{Logistic}: \ \tilde{y}_t &= \frac{M\tilde{y}_0}{\tilde{y}_0 + (M - \tilde{y}_0)e^{-kMt}}, \ k > 0.
\end{aligned}
$$

See *Figures 2.2–2.4* for examples of some of these.

The linear and quadratic trends can be generated by using the fitted values from an

OLS regression of y_t on a constant and a time variable (for instance, 1971, 1972, ...). Sometimes these models have different slopes before and after a special date (a "segmented" linear trend). This is typically the case for macro variables like GDP where trend growth was much higher during the 1950s and 1960s than during the 1970s and 1980s.

The exponential model can also be estimated with OLS on the log of y_t, since

$$\ln \tilde{y}_t = \ln a e^{bt} = \ln a + bt. \tag{2.3}$$

Note that the exponential model implies that the *growth rate* of $\tilde{y}_t$ is b. To see that note that the change (over a very short time interval)

$$\frac{d\tilde{y}_t}{dt} = abe^{bt}, \text{ so the relative change (growth rate) is } \frac{d\tilde{y}_t}{dt}\frac{1}{\tilde{y}_t} = b. \tag{2.4}$$

The moving average (MA) smooths a series by taking a weighted average over current and past observations. The coefficients are seldom estimated, but rather imposed a priori. Two special cases are popular. In the *equally-weighted moving average* all the θ coefficients in (2.2) are equal (and therefore equal to $1/(1+q)$ to make them sum to unity). In the *exponential moving average* (also called exponential smoothing) all available observations are used on the right hand side, but the weights are higher for recent observations

$$\tilde{y}_t = (1-\lambda)(y_t + \lambda y_{t-1} + \lambda^2 y_{t-2} + \ldots), \text{ where } 0 < \lambda < 1. \tag{2.5}$$

Since $0 < \lambda < 1$, the weights are declining. This trend can equivalently be calculated by the convenient recursive formula

$$\tilde{y}_t = \lambda \tilde{y}_{t-1} + (1-\lambda) y_t, \tag{2.6}$$

which just needs a starting value for the first trend value (for instance, $\tilde{y}_1 = y_1$). See Figure 2.3 for an example.

The logistic trend is often used for things that are assumed to converge to some level M as $t \to \infty$. It has been used for population trends and for ratios that cannot trend outside a certain range like [0, 1]. It is the solution to the differential equation $d\tilde{y}_t/dt = k\tilde{y}_t(M - \tilde{y}_t)$. It converges from above if $\tilde{y}_0 > M$ and from below if $\tilde{y}_0 < M$. Estimating

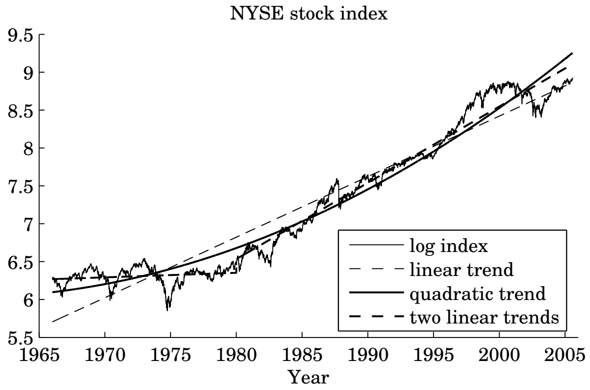


Figure 2.2: NYSE index (composite)

the parameters of the logistic trend requires a nonlinear estimation technique. See Figure 2.4 for an example.

Remark 4 **The Hodrick-Prescott filter. Another popular trend model (especially for macro data) is to use a Hodrick-Prescott (HP) filter (also called a Whittaker-Henderson filter or a cubic spline). It calculates the trend components, $\tilde{y}_1$, $\tilde{y}_2$, ..., $\tilde{y}_T$ by minimizing the loss function*

$$\sum_{t=1}^{T} (y_t - \tilde{y}_t)^2 + \lambda \sum_{t=3}^{T} \left[(\tilde{y}_t - \tilde{y}_{t-1}) - (\tilde{y}_{t-1} - \tilde{y}_{t-2})\right]^2.$$

The first term punishes (squared) deviations of the trend from the actual series; the second punishes (squared) acceleration (change of change) of the trend level. The result is thus a trade-off between tracking the original series and smoothness of the trend level: $\lambda = \infty$ gives a linear trend, while $\lambda = 0$ gives a trend that equals the original series. $\lambda = 1600$ is a common value for quarterly macro data. The minimization problem gives (approximately) a symmetric two-sided moving average

$$\tilde{y}_t = \cdots + \theta_1 y_{t-1} + \theta_0 y_t + \theta_1 y_{t+1} + \cdots$$

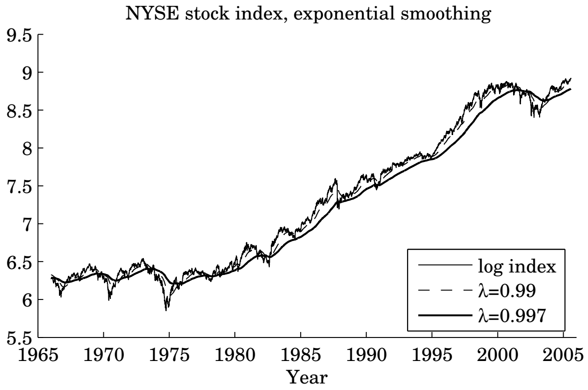


Figure 2.3: NYSE index (composite)

Note that any two-sided moving average needs future values of y in order to calculate the trend level in t. In "real-time" applications this is typically handled by making forecasts of future levels and using them in the calculation of today's trend level.

2.3 Seasonality

This section discusses how seasonality in data can be handled. Most macroeconomic data series have fairly regular seasonal patterns; financial series do not. See Figure 2.1 for an example.

The typical macro seasonality for European countries is: low in Q1, high in Q2, low in Q3, and high in Q4. The calendar and vacation habits must take most of the blame. The first quarter is shorter than the other quarters, and countries on the northern hemisphere typically take vacation in July or August.

It should be noticed that the number of working days in a quarter changes from year to year—mostly because of how traditional holidays interact with the calendar. The most important case is that Easter (which essentially follows the Jewish calendar) is sometimes in Q1 and sometimes in Q2.

There are also some important regional differences. For instance, the southern hemi-

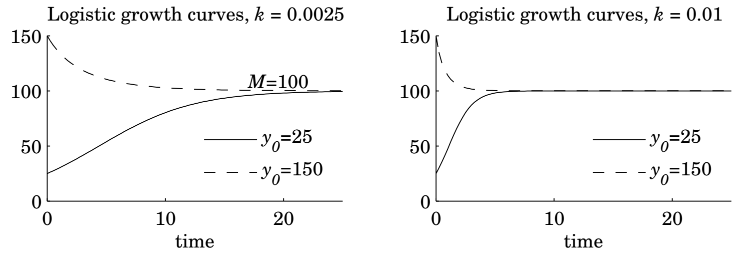


Figure 2.4: Logistic trend curves

sphere typically have vacations in Jan/Feb. Another difference is that countries in northern Europe typically have vacation in July, while southern Europe opts for August.

In most cases, macroeconomists choose to work with seasonally adjusted data: this makes it easier to see the business cycle movements. It is typically also believed that the seasonal factors have only small effects on financial markets and the inflation pressure (although there may be seasonal movements in the price index).

There are several ways of getting rid of the season. The most obvious is to get a *seasonally adjusted series* from the statistical agency. In fact, it can be argued that seasonally adjusted series are of better quality than the raw series. This may sound strange, since the seasonally adjusted series is based on the raw series, but the fact is that statistical agencies spend more time on controlling the quality of the seasonally adjusted series (since it is the one that most users care about).

If you need to do the seasonal adjustment yourself, then the following methods are useful:

- Run the data series through a filter like X11 (essentially a two-sided moving average). To do that, you typically have to take a stand on whether the seasonal factor is additive or multiplicative.

- "Difference away" the season by calculating the growth rate since the same month/quarter last year. See Figure 2.1 for an example.

- Do your own seasonal adjustment. Here is a simple routine:

- Construct a set of dummy variables for each season, for instance, $Q_{1t}, \ldots, Q_{4t}$ for quarterly data. Note that $Q_{qt} = 1$ if period t is in season q, and $Q_{qt} = 0$ otherwise.
- Run the least squares regression $y_t = a_1 Q_{1t} + a_2 Q_{2t} + a_3 Q_{3t} + a_4 Q_{4t} + bt + u_t$ and take $\hat{b}t + \hat{u}_t$ as your seasonally adjusted series. If you want multiplicative season effects, then you run this regression on the logarithm of y_t instead.

Remark 5 **Here is an alternative, slightly more complicated, routine for seasonal adjustment.*
(a) Construct a trend component as a long centered moving average spanning the seasonal pattern, y_t^* For monthly data this would be $y_t^* = (y_{t+6} + \ldots + y_{t+1} + y_t + y_{t-1} + \ldots + y_{t-5})/12$. This new series will contain the trend plus cyclical component and could be used as a very crude seasonally adjusted series (too smooth since also the irregular component is averaged out too).
(b) Define $z_t = y_t - y_t^*$ or $z_t = y_t / y_t^*$ which should contain the seasonal and irregular components.
(c) Average the observations of z_t corresponding to the same season and call these averages s_1, ...
(d) Define the seasonally adjusted series as the original series minus the seasonal average, $y_t - s_q$ (where period t is in season q), or as the original series divided by the seasonal average, y_t / s_q.

3 Forecasting

3.1 Evaluating Forecast Performance

Further reading: Diebold (2001) 11; Stekler (1991); Diebold and Mariano (1995)

To do a solid evaluation of the forecast performance (of some forecaster/forecast method/forecast institute), we need a sample (history) of the forecasts and the resulting forecast errors. The reason is that the forecasting performance for a single period is likely to be dominated by luck, so we can only expect to find systamtic patterns by looking at results for several periods.

Let e_t be the forecast error in period t

$$e_t = y_t - \hat{y}_t, \tag{3.1}$$

where $\hat{y}_t$ is the forecast and y_t the actual outcome. (Warning: some autors prefer to work with $\hat{y}_t - y_t$ as the forecast error instead.)

Most statistical forecasting methods are based on the idea of minimizing the sum of squared forecast errors, $\Sigma_{t=1}^T e_t^2$. For instance, the least squares (LS) method picks the regression coefficient in

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{3.2}$$

to minimize the sum of squared residuals, $\Sigma_{t=1}^T \varepsilon_t^2$. This will, among other things, give a zero mean of the fitted residuals and also a zero correlation between the fitted residual and the regressor.

Evaluation of a forecast often involve extending these ideas to the forecast method, irrespective of whether a LS regression has been used or not. In practice, this means studying if *(i)* the forecast error, e_t, has a zero mean; *(ii)* the forecast error is uncorrelated to the variables (information) used in constructing the forecast; and *(iii)* to compare the sum (or mean) of squared forecasting errors of different forecast approaches. A non-zero mean of the errors clearly indicates a bias, and a non-zero correlation suggests that the information has not been used efficiently (a forecast error should not be predictable...)

Remark 6 *(Autocorrelation of forecast errors*) Suppose we make one-step-ahead forecasts, so we are forming a forecast of y_{t+1} based on what we know in period t. Let $e_{t+1} = y_{t+1} - \mathrm{E}_t\, y_{t+1}$, where $\mathrm{E}_t\, y_{t+1}$ just denotes our forecast. If the forecast error is unforecastable, then the forecast errors cannot be autocorrelated, for instance, $\mathrm{Corr}(e_{t+1}, e_t) = 0$. For two-step-ahead forecasts, the situation is a bit different. Let $e_{t+2,t} = y_{t+2} - \mathrm{E}_t\, y_{t+2}$ be the error of forecasting y_{t+2} using the information in period t (notice: a two-step difference). If this forecast error is unforecastable using the information in period t, then the previously mentioned $e_{t+2,t}$ and $e_{t,t-2} = y_t - \mathrm{E}_{t-2}\, y_t$ must be uncorrelated—since the latter is known when the forecast $\mathrm{E}_t\, y_{t+2}$ is formed (assuming this forecast is efficient). However, there is nothing hat guarantees that $e_{t+2,t}$ and $e_{t+1,t-1} = y_{t+1} - \mathrm{E}_{t-1}\, y_{t+1}$ are uncorrected—since the latter contains new information compared to what was known when the forecast $\mathrm{E}_t\, y_{t+2}$ was formed. This generalizes to the following: an efficient h-step-ahead forecast error must have a zero correlation with* *the forecast error $h-1$ (and more) periods earlier.*

The comparison of forecast approaches/methods is not always a comparison of actual forecasts. Quite often, it is a comparison of a forecast method (or forecasting institute) with some kind of naive forecast like a “no change” or a random walk. The idea of such a comparison is to study if the resources employed in creating the forecast really bring value added compared to a very simple (and inexpensive) forecast.

It is sometimes argued that forecasting methods should not be ranked according to the sum (or mean) squared errors since this gives too much weight to a single large error. Ultimately, the ranking should be done based on the true benefits/costs of forecast errors—which may differ between organizations. For instance, a forecasting agency has a reputation (and eventually customers) to loose, while an investor has more immediate pecuniary losses. Unless the relation between the forecast error and the losses are immediately understood, the ranking of two forecast methods is typically done based on a number of different criteria. The following are often used:

- mean error, $\Sigma_{t=1}^{T} e_t / T$,
- mean squared error, $\Sigma_{t=1}^{T} e_t^2 / T$,
- mean absolute error, $\Sigma_{t=1}^{T} |e_t| / T$,
- fraction of times that the absolute error of method a smaller than that of method b,
- fraction of times that method a predicts the direction of change better than method b,
- profitability of a trading rule based on the forecast (for financial data),
- results from a regression of the outcomes on two forecasts ($\hat{y}_t^a$ and $\hat{y}_t^b$)

$$y_t = \omega \hat{y}_t^a + \gamma \hat{y}_t^b + \text{ residual}, \tag{3.3}$$

  where $\omega = 1$ and $\gamma = 0$ indicates that forecast a contains all the information in b and more.
- A pseudo R^2 defined as $\text{Corr}(y_t, \hat{y}_t)^2$, where y_t is the actual value and $\hat{y}_t$ is the forecast.

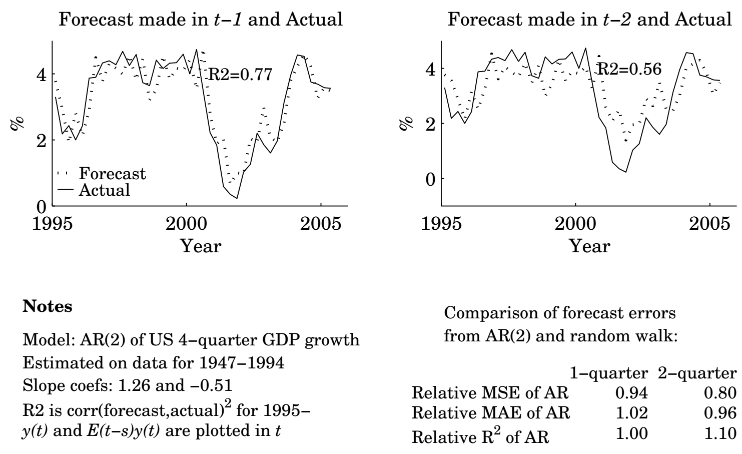


Figure 3.1: Forecasting with an AR(2)

See *Figure 3.1* for an example.

As an example, Leitch and Tanner (1991) analyze the profits from selling 3-month T-bill futures when the forecasted interest rate is above futures rate (forecasted bill price is below futures price). The profit from this strategy is (not surprisingly) strongly related to measures of correct direction of change (see above), but (perhaps more surprisingly) not very strongly related to mean squared error, or absolute errors.

Example 7 *We want to compare the performance of the two forecast methods a and b. We have the following forecast errors* $(e_1^a, e_2^a, e_3^a) = (-1, -1, 2)$ *and* $(e_1^b, e_2^b, e_3^b) = (-1.9, 0, 1.9)$. *Both have zero means, so there is (in this very short sample) no constant bias. The mean squared errors are*

$$MSE^a = [(-1)^2 + (-1)^2 + 2^2]/3 = 2$$
$$MSE^b = [(-1.9)^2 + 0^2 + 1.9^2]/3 \approx 2.41,$$

so forecast a is better according to the mean squared errors criterion. The mean absolute *errors are*

$$MAE^a = [|-1| + |-1| + |2|]/3 \approx 1.33$$
$$MAE^b = [|-1.9| + |0| + |1.9|]/3 \approx 1.27,$$

so forecast b is better according to the mean absolute errors criterion. The reason for the difference between these criteria is that forecast b has fewer but larger errors—and the quadratic loss function punishes large errors very heavily. Counting the number of times the absolute error (or the squared error) is smaller, we see that a is better one time (first period), and b is better two times.

3.2 Combining Forecasts from Different Forecasters/Models

Further reading: Diebold (2001) 11; Evans (2003) 8; Batchelor and Dua (1995)

There is plenty of evidence that taking averages of different forecasts typically reduces the forecast error variance. The intuition is that all forecasts are noisy signals of the actual value, and by taking an average the noise becomes less important.

3.3 Forecast Uncertainty and Disagreement

It is fairly straightforward to gauge the forecast uncertainty of an econometric model by looking at, for instance, the variance of the errors of the one-step ahead forecasts. We can, of course, do the same on a time series of historical judgemental forecasts. Unfortunately, this approach only gives an average number which typically says very little about how uncertainty has changed over time.

Some surveys of forecasters ask for probabilities which can be used to assess the forecast uncertainty in "real-time." Another popular measure of uncertainty is the disagreement between forecasters—typically measured as the variance of the point forecasts or summarized by giving the minimum and maximum among a set of forecasts made at the same point in time (see, for instance, The Economist). Some research (see, for instance, Giordani and Söderlind (2003)) finds that these different measures of uncertainty typically are highly correlated.

See *Figure 3.2* for an example of survey data.

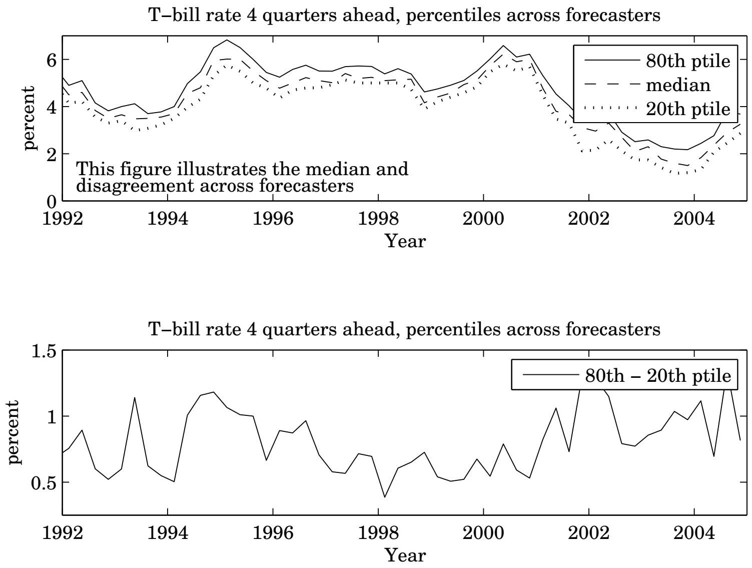


Figure 3.2: Forecasts of US T-bill rates 4 quarters ahead, Survey of Professional Forecasters

3.4 Words of Wisdom: Forecasting in Practice

Main reference: Makridakis, Wheelwright, and Hyndman (1998) 11 and 12.1

Evaluations of the post-sample forecasting performance of macroeconomic data using different types of models have revealed some (reasonably) clear patterns:

- Simple models are often as good as more complex models, especially when relatively much of the variability in the data is unforecastable.
- Different forecasting horizons require different models. (This is probably just a restatement of the first point.)
- Averaging forecasts over methods/forecasters often produces better forecasts.

Some often observed features of judgemental forecasts:

- overoptimism,
- understating of uncertainty,
- recency bias (too influenced by recent events).

4 Time Series Analysis

Main reference: Diebold (2001) 6–8; Evans (2003) 7; Newbold (1995) 17 or Pindyck and Rubinfeld (1998) 13.5, 16.1-2, and 17.2
Further reading: Makridakis, Wheelwright, and Hyndman (1998) (additional useful reading, broad overview)

We now focus on the "cycle" of the series. In practice this means that this section disregards constants (including seasonal effects) and trends—you can always subtract them before applying the methods in this section—and then add them back later.

Time series analysis has proved to be a fairly efficient way of producing forecasts. Its main drawback is that it is typically not conducive to structural or economic analysis of the forecast. Still, small VAR systems (see below) have been found to forecast as well as large structural macroeconometric models (see Makridakis, Wheelwright, and Hyndman (1998) 11 for a discussion).

As an example, consider forecasting the inflation rate. Macroeconomics would tell us that current inflation depends on at least three things: recent inflation, current expectations about future inflation, and the current business cycle conditions. This means that we need to forecast both future inflation expectations and future business cycle conditions in order to forecast future inflation—which is hard (costly). A simple time series model is easy to estimate and forecasts can be produced on the fly. Of course, a time series model has forecasting power only if future inflation is related to current values of inflation and other series that we include in the model. This will happen if there is lots of inertia in price setting (for instance, today's price decision depends on what the competitors did last period) or in the business cycle conditions (for instance, investment decisions take time to implement). This is likely to be the case for inflation, but certainly not for all variables—time series models are not particularly good at forecasting exchange rate changes or equity returns (no one is, it seems). In any case, even if a time series model is good at forecasting inflation, it will probably not explain the economics of the forecast.

4.1 Autocorrelations

Autocorrelations measure how a the current value of a series is linearly related to earlier (or later) values. The *pth autocovariance* of x is estimated by

$$\widehat{\text{Cov}}(x_t, x_{t-p}) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})(x_{t-p} - \bar{x}), \tag{4.1}$$

where we use the same estimated (using all data) mean in both places. Similarly, the *pth autocorrelation* is estimated as

$$\widehat{\text{Corr}}(x_t, x_{t-p}) = \frac{\widehat{\text{Cov}}(x_t, x_{t-p})}{\widehat{\text{Std}}(x_t)^2}. \tag{4.2}$$

Compared with a traditional estimate of a correlation (1.3) we here impose that the standard deviation of x_t and x_{t-p} are the same (which typically does not make much of a difference).

4.2 AR(1)

In this section we study the *first-order autoregressive* process, AR(1), in some detail in order to understand the basic concepts of autoregressive processes.

An AR(1) is

$$y_t = a y_{t-1} + \varepsilon_t, \tag{4.3}$$

where ε_t is identically and independently distributed (iid) and also uncorrelated with y_{t-1}. If $-1 < a < 1$, then the effect of a shock eventually dies out: y_t is stationary. Since there is no constant in (4.3), so we have implicitely assumed that y_t has a zero mean, that is, is a demeaned variable (an original variable minus its mean, for instance $y_t = z_t - \bar{z}_t$).

The AR(1) model can be estimated with OLS (since ε_t and y_{t-1} are uncorrelated) and the usual tools for testing significance of coefficients and estimating the variance of the residual all apply.

The basic properties of an AR(1) process are (provided $|a| < 1$)

$$\text{Var}(y_t) = \text{Var}(\varepsilon_t)/(1 - a^2) \tag{4.4}$$

$$\text{Corr}(y_t, y_{t-s}) = a^s, \tag{4.5}$$

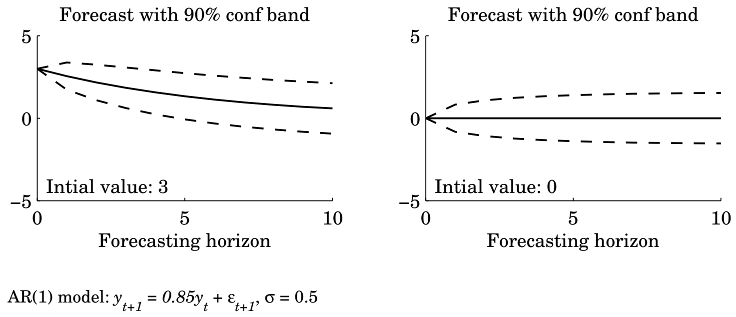


Figure 4.1: Forecasting an AR(1) process

so the variance and autocorrelation are increasing in a.

If $a = 1$ in (4.3), then we get a *random walk*. It is clear from the previous analysis that a random walk is non-stationary—that is, the effect of a shock never dies out. This implies that the variance is infinite and that the standard tools for testing coefficients etc. are invalid. The solution is to study changes in y instead: $y_t - y_{t-1}$. In general, processes with the property that the effect of a shock never dies out are called non-stationary or unit root or integrated processes. Try to avoid them.

4.2.1 Forecasting with an AR(1)

Suppose we have estimated an AR(1). To simplify the exposition, we assume that we actually know a and $\text{Var}(\varepsilon_t)$, which might be a reasonable approximation if they were estimated on long sample. (See Section 1.2.4 for a full treatment where the parameter uncertainty is incorporated in the analysis.)

We want to *forecast y_{t+1} using information available in t*. From (4.3) we get

$$y_{t+1} = ay_t + \varepsilon_{t+1}. \tag{4.6}$$

Since the best guess of ε_{t+1} is that it is zero, the best forecast and the associated forecast

error are

$$\text{E}_t\; y_{t+1} = ay_t, \text{ and} \tag{4.7}$$

$$y_{t+1} - \text{E}_t\; y_{t+1} = \varepsilon_{t+1} \text{ with variance } \sigma^2. \tag{4.8}$$

Recall that y_t is a demeaned variable, so to forecast the original variable we just have to add its mean to the forecast of y_{t+1}.

We may also want to forecast y_{t+2} using the information in t. To do that note that (4.3) gives

$$\begin{aligned} y_{t+2} &= ay_{t+1} + \varepsilon_{t+2} \\ &= a\underbrace{(ay_t + \varepsilon_{t+1})}_{y_{t+1}} + \varepsilon_{t+2} \\ &= a^2 y_t + a\varepsilon_{t+1} + \varepsilon_{t+2}. \end{aligned} \tag{4.9}$$

Since the $\text{E}_t\; \varepsilon_{t+1}$ and $\text{E}_t\; \varepsilon_{t+2}$ are both zero, we get that

$$\text{E}_t\; y_{t+2} = a^2 y_t, \text{ and} \tag{4.10}$$

$$y_{t+2} - \text{E}_t\; y_{t+2} = a\varepsilon_{t+1} + \varepsilon_{t+2} \text{ with variance } a^2\sigma^2 + \sigma^2. \tag{4.11}$$

The two-periods ahead forecast is clearly more uncertain: more shocks can hit the system. This is a typical pattern—and then extends to longer forecasting horizons.

If the shocks ε_t, are normally distributed, then we can calculate 90% confidence intervals around the point forecasts in (4.7) and (4.10) as

$$\text{90\% confidence band of } \text{E}_t\; y_{t+1} : ay_t \pm 1.65 \times \sigma \tag{4.12}$$

$$\text{90\% confidence band of } \text{E}_t\; y_{t+2} : a^2 y_t \pm 1.65 \times \sqrt{a^2\sigma^2 + \sigma^2}. \tag{4.13}$$

(Recall that 90% of the probability mass is within the interval -1.65 to 1.65 in the N(0,1) distribution). To get 95% confidence bands, replace 1.65 by 1.96. Figure 4.1 gives an example. As before, recall that y_t is demeaned so to get the confidence band for the original series, add its mean to the forecasts and the confidence band boundaries.

Remark 8 *(White noise as special case of AR(1).) When $a = 0$ in (4.3), then the AR(1) collapses to a white noise process. The forecast is then a constant (zero) for all forecasting horizons, and the forecast error variance is also the same for all horizons.*

Example 9 *If $y_t = 3$, $a = 0.85$ and $\sigma = 0.5$, then (4.12)–(4.13) become*

$$90\% \text{ confidence band of } \mathrm{E}_t\, y_{t+1} : 0.85 \times 3 \pm 1.65 \times 0.5 \approx [1.7, 3.4]$$
$$90\% \text{ confidence band of } \mathrm{E}_t\, y_{t+2} : 0.85^2 \times 3 \pm 1.65 \times \sqrt{0.85^2 \times 0.5^2 + 0.5^2} \approx [1.1, 3.2].$$

If the original series has a mean of 7 (say), then add 7 to each of these numbers to get the confidence bands [8.7, 10.4] *for the one-period horizon and* [8.1, 10.2] *for the two-period horizon.*

4.3 AR(p)

The *pth-order autoregressive* process, AR(p), is a straightforward extension of the AR(1)

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + ... + a_p y_{t-p} + \varepsilon_t. \tag{4.14}$$

All the previous calculations can be made on this process as well—it is just a bit messier. This process can also be estimated with OLS since ε_t are uncorrelated with lags of y_t.

4.3.1 Forecasting with an AR(2)*

As an example, consider making a forecast of y_{t+1} based on the information in t by using an AR(2)

$$y_{t+1} = a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}. \tag{4.15}$$

This immediately gives the one-period point forecast

$$\mathrm{E}_t\, y_{t+1} = a_1 y_t + a_2 y_{t-1}. \tag{4.16}$$

We can use (4.15) to write y_{t+2} as

$$\begin{aligned} y_{t+2} &= a_1 y_{t+1} + a_2 y_t + \varepsilon_{t+2} \\ &= a_1 \underbrace{(a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1})}_{y_{t+1}} + a_2 y_t + \varepsilon_{t+2} \\ &= (a_1^2 + a_2) y_t + a_1 a_2 y_{t-1} + a_1 \varepsilon_{t+1} + \varepsilon_{t+2}. \end{aligned} \tag{4.17}$$

The two-period forecast is therefore

$$\mathrm{E}_t\, y_{t+2} = (a_1^2 + a_2) y_t + a_1 a_2 y_{t-1}. \tag{4.18}$$

The expressions for the forecasts and forecast error variances quickly get somewhat messy—and even more so with an AR of higher order than two. There is a simple, and approximately correct, shortcut that can be taken. Note that both the one-period and two-period forecasts are linear functions of y_t and y_{t-1}. We could therefore estimate the following two equations with OLS

$$y_{t+1} = a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1} \tag{4.19}$$
$$y_{t+2} = b_1 y_t + b_2 y_{t-1} + v_{t+2}. \tag{4.20}$$

Clearly, (4.19) is the same as (4.15) and the estimated coefficients can therefore be used to make one-period forecasts, and the variance of ε_{t+1} is a good estimator of the variance of the one-period forecast error. The coefficients in (4.20) will be very similar to what we get by combining the a_1 and a_2 coefficients as in (4.17): b_1 will be similar to $a_1^2 + a_2$ and b_2 to $a_1 a_2$ (in an infinite sample they should be identical). Equation (4.20) can therefore be used to make two-period forecasts, and the variance of v_{t+2} can be taken to be the forecast error variance for this forecast. This approach extends to longer forecasting horizons and to AR models of higher order.

Figure 3.1 gives an empirical example.

4.4 ARMA(p,q)*

Autoregressive-moving average models add a moving average structure to an AR model. For instance, an ARMA(2,1) could be

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1},$$

where ε_t is iid. This type of model is much harder to estimate than the autoregressive model (LS cannot be used). The appropriate specification of the model (number of lags of y_t and ε_t) is often unknown. The Box-Jenkins methodology is a set of guidelines for arriving at the correct specification by starting with some model, study the autocorrelation structure of the fitted residuals and then changing the model.

Most ARMA models can be well approximated by an AR model—provided we add some extra lags. Since AR models are so simple to estimate, this approximation approach is often used.

4.5 VAR(p)

The vector autoregression is a multivariate version of an AR(1) process: we can think of y_t and ε_t in (4.14) as vectors and the a_i as matrices.

We start by considering the (fairly simple) VAR(1) is (in matrix form)

$$\begin{bmatrix} x_{t+1} \\ z_{t+1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_t \\ z_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{xt+1} \\ \varepsilon_{zt+1} \end{bmatrix}, \tag{4.21}$$

or equivalently

$$x_{t+1} = a_{11}x_t + a_{12}z_t + \varepsilon_{xt+1}, \text{ and} \tag{4.22}$$

$$z_{t+1} = a_{21}x_t + a_{22}z_t + \varepsilon_{zt+1}. \tag{4.23}$$

Both (4.22) and (4.23) are regression equations, which can be estimated with OLS (since ε_{xt+1} and ε_{zt+1} are uncorrelated with x_t and z_t). It is then straightforward, but a bit messy, to construct forecasts.

As for the AR(p) model, a practical way to get around the problem with messy calculations is to estimate a separate model for each forecasting horizon. In a large sample, the difference between the two ways is trivial. For instance, suppose the correct model is the VAR(1) in (4.21) and that we want to forecast x one and two periods ahead. Clearly, forecasts for any horizon must be functions of x_t and z_t so the regression equations should be of the form

$$x_{t+1} = \delta_1 x_t + \delta_2 z_t + u_{t+1}, \text{ and} \tag{4.24}$$

$$x_{t+2} = \gamma_1 x_t + \gamma_2 z_t + w_{t+s}, \tag{4.25}$$

and similarly for forecasts of z. With estimated coefficients (OLS can be used), it is straightforward to calculate forecasts and forecast error variances.

In a more general VAR(p) model we need to include p lags of both x and z in the regression equations ($p = 1$ in (4.24) and (4.25)).

4.5.1 *Granger Causality

If z_t can help predict future x, over and above what lags of x itself can, then z is said to *Granger cause* x. This is a statistical notion of causality, and may not necessarily have much to do with economic causality (Christmas cards may Granger cause Christmas). In (4.24) z does Granger cause x if $\delta_2 \neq 0$, which can be tested with an F-test. More generally, there may be more lags of both x and z in the equation, so we need to test if all coefficients on different lags of z are zero.

Bibliography

Batchelor, R., and P. Dua, 1995, "Forecaster Diversity and the Benefits of Combining Forecasts," *Management Science*, 41, 68–75.

Diebold, F. X., 2001, *Elements of Forecasting*, South-Western, 2nd edn.

Diebold, F. X., and R. S. Mariano, 1995, "Comparing Predcitve Accuracy," *Journal of Business and Economic Statistics*, 13, 253–265.

Evans, M. K., 2003, *Practical Business Forecasting*, Blackwell Publishing.

Giordani, P., and P. Söderlind, 2003, "Inflation Forecast Uncertainty," *European Economic Review*, 47, 1037–1059.

Gujarati, D. N., 1995, *Basic Econometrics*, McGraw-Hill, New York, 3rd edn.

Leitch, G., and J. E. Tanner, 1991, "Economic Forecast Evaluation: Profit versus the Conventional Error Measures," *American Economic Review*, 81, 580–590.

Makridakis, S., S. C. Wheelwright, and R. J. Hyndman, 1998, *Forecasting: Methods and Applications*, Wiley, New York, 3rd edn.

Newbold, P., 1995, *Statistics for Business and Economics*, Prentice-Hall, 4th edn.

Pindyck, R. S., and D. L. Rubinfeld, 1998, *Econometric Models and Economic Forecasts*, Irwin McGraw-Hill, Boston, Massachusetts, 4ed edn.

Stekler, H. O., 1991, "Macroeconomic Forecast Evaluation Techniques," *International Journal of Forecasting*, 7, 375–384.

The Economist, 2000, *Guide to Economic Indicators*, Profile Books Ltd, London, 4th edn.

5 Overview of Macroeconomic Forecasting

5.1 The Forecasting Process

The forecasting processes can be broadly divided into two categories: top-down forecasting and bottom-up forecasting. A top-down forecasts begins with an overall assessment of the aggregate economy—which is later used for making forecasts of the GDP components and other more detailed aspects. A bottom-up forecast instead starts with forecasts at a low level of aggregation (the production of special steel or the production of saw mills) and then works upwards by summing up. It is probably fair to say that most forecasting institutes have moved away from the bottom-up forecasting of earlier times and are now practicing a more mixed approach where there is a number of rounds between the aggregate and disaggregate level. The aim of the forecasting process is to first get a clear picture of the current economic situation—and then (and only then) make projections.

Some sectors of the economy can be treated as "exogenous" (predetermined) to the rest of the economy—and will therefore be forecasted without much input from the other sectors. They are therefore the natural starting points in the forecasting process (since they surely affect the other components). The international economy is such a "sector." It is hardly affected by business cycle conditions of even medium-sized European countries (the very largest countries are different) and it influences the domestic economy via (at least) the demand for exports, import prices, and (more and more) the overall economic sentiment. For short run forecasts, some other sectors are to a large extent predetermined because of time lags. For instance, the major bulk of this month's construction work was surely begun in earlier months. The same goes for other very large projects (for instance, ship building, manufacturing of generators). The spending of government and local authorities also have large predetermined components—at least in countries with tight budget discipline.

The inputs to the forecasting process are the most recent (and often preliminary) data on the national accounts, production, exports, etc. Many other indicators are also used—for at least three reasons. First, the preliminary data are known to be somewhat erratic and large revisions (later) are common. Second, data on the national accounts and production are produced only with a lag (and is still preliminary). Third, many other economic time series have been shown to have better predictive power—they are often called "leading indicators" (see below).

Macroeconomic forecasting is still rooted in the Keynesian model: the demand side of GDP is given more attention than the supply side. There are at least three (related) reasons for this: it is commonly believed that output is essentially demand determined in the short run (supply considerations are allowed to play a role in the inflation process, though), economic theories for the demand side are perhaps better understood than the theories for the supply side, and the data on the demand side components is typically of better quality than for the supply side.

Virtually no forecasting institute relies on an econometric model to produce their main forecast. Instead, the forecasts are products of a large number of committee meetings. One possible scenario is as follows. There could be a committee for analyzing investment, another for private consumption. They reach their preliminary conclusions by using all sorts of inputs: the latest data releases, consumer and producer surveys, other leading indicators, econometric models, and a large dose of judgmental reasoning. In the next step, these preliminary committee reports are submitted to a higher level, checked for consistency (do the various forecasts of the various GDP growth add up to the assessment of the aggregate economy) and modified, and then sent back to the committees for a second round.

The outputs of the forecasting process are typically the following: *(i)* a detailed description of the current economic situation; *(ii)* a fairly detailed forecast for the near future; *(iii)* an outlook beyond that with a discussion of alternative scenarios. The discussion and analysis is often as important as the numbers, since it through the analysis that the forecasters might be able to convince its readers.

The most recent developments in macroeconomic forecasting are to produce confidence bands around the point forecast and to show different scenarios ("what if there instead is a world wide depression?" or "what if the government stimulates the economy by spending more/taxing less?").

5.2 Forecasting Institutes

Macroeconomic forecasts are produced by many organizations and companies—and for quite different reasons. Government and central banks want to have early warnings of business cycle movements in order to implement counter cyclical policies. The business sector needs forecasts for production and investment planning, and financial investors for asset allocation decisions.

The most sophisticated forecasts are typically produced by central banks and national forecasting agencies. They have large and well educated staffs and employ a range of different forecasting models (often including some large-scale macroeconometric models). All other macroeconomic forecasts are typically just marginal adjustments of these first-tier forecasts.

International organizations (in particular OECD) also produce quality forecasts. However, the relative importance of these forecasts has decreased over the last decades—as more and more countries have set up advanced forecasting agencies of their own (see above).

Finance ministries typically also use large forecasting resources, but it should be kept in mind that their forecasts are subject to political approval.

Commercial banks produce macroeconomic forecasts for at least three reasons: as PR (it is worth a lot to have the bank's chief economist being interviewed on TV), as a service to the large clients of the bank (which are often supplied with the forecasts before the public—and regularly updated), and to provide a background for the bank's own trading.

Trade unions and employers' associations often produce their own forecasts with the aim of influencing politics and to provide background material for wage negotiations.

Most forecasting institutes produce a few (two to four) forecasts per year. The number (and their location in time) depends mainly on the releases of new economic statistics and/or the forecast schedules of the competitors and major institutes.

6 Business Cycle Facts

6.1 Key Features of Business Cycle Movements

The business cycle is identified from general economic conditions, which effectively means the movements in production (GDP), unemployment, or capacity utilization. The two key features of a business cycles are that most sectors of the economy move in the same direction and that the movements are persistent. Sectors that produce durable goods (for instance, consumer durables, investment goods for business, and construction) are more heavily affected than sectors that produce non-durables (for instance, food, and energy) and services.

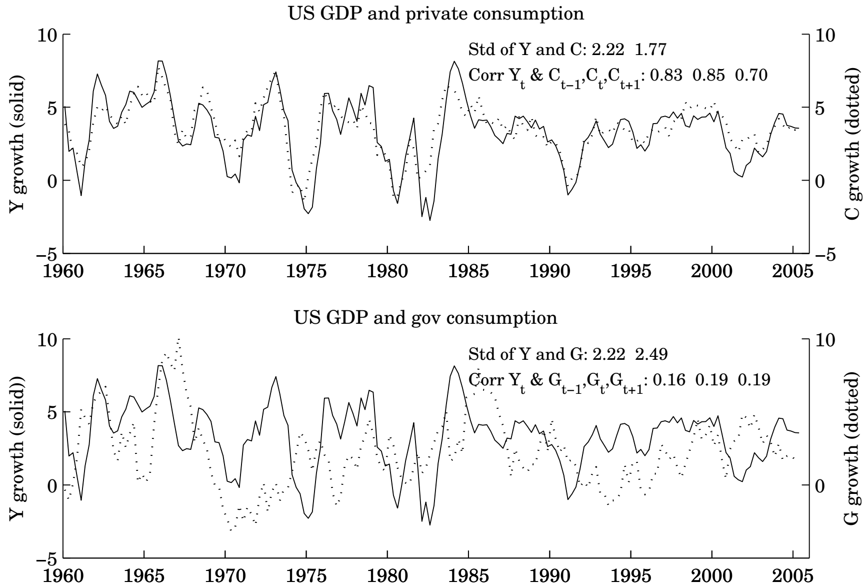


Figure 6.1: US GDP and its components, 4-quarter growth, %

GDP is defined as the total value added in the economy (from the supply side) or as

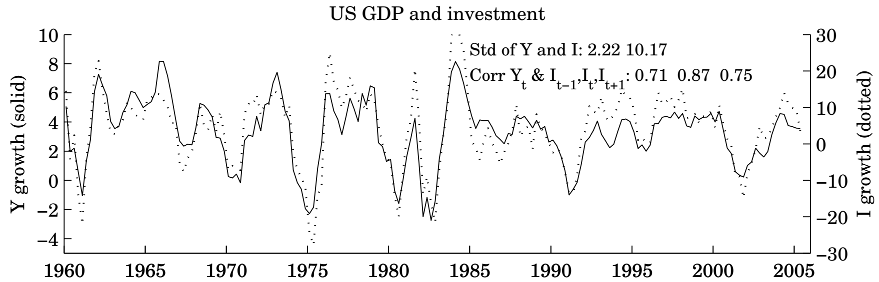


Figure 6.2: US GDP and its components, 4-quarter growth, %

the sum of the following demand components (from the demand side)

$$\text{GDP } (Y) = \text{Private consumption } (C) + \text{Investment } (I) + \text{Government purchases } (G) + \text{Exports } (X) - \text{Imports } (M). \quad (6.1)$$

For most developed economies, private consumption accounts for 66% of GDP (although only 50% in some northern European countries), investment is around 15%–20%, and government purchases the rest. Exports and imports typically balance over the longer run—and both account for 20%–50% of GDP (with the higher number for small European countries).

Figures 6.1–6.3 and *Table 6.1*show the US GDP components. They illustrate some typical (across time and countries) features of business cycles. We see the following.

1. Volatility: GDP, private and government consumption are less volatile than investment and foreign trade.
2. Correlations: GDP, consumption, and imports are typically highly correlated. Government consumption and exports are often less procyclical, although this can be different for smaller countries (with relatively more exports).
3. Timing: the GDP components are mostly coincident with GDP.

For forecasting purposes, it is often useful to split up some of the GDP components further.

	Std	Corr(Y_t, z_{t-1})	Corr(Y_t, z_t)	Corr(Y_t, z_{t+1})
GDP (Y)	2.22	0.86	1.00	0.86
Private consumption	1.77	0.83	0.85	0.70
Government consumption	2.49	0.16	0.19	0.19
Investments	10.17	0.71	0.87	0.75
Exports	6.26	0.12	0.31	0.42
Imports	7.10	0.59	0.71	0.70
Private cons (durables)	6.37	0.74	0.76	0.58
Residential inv	13.18	0.71	0.60	0.36
Wages	2.29	−0.27	−0.21	−0.16
Productivity	1.69	0.58	0.55	0.25
Unemployment	1.45	−0.14	−0.30	−0.42

Table 6.1: US GDP components and business cycle indicators. Standard deviations and the correlation of GDP (Y_t) with the other variables (z, in $t-1, t, t+1$)

Private consumption is split up into consumer durables (very cyclical, large import content), consumer non-durables (less cyclical, smaller import content), and services (stable, small import content).

Investment is split up into residential investment (depends on the households' economy), business construction (cyclical, long lags), machinery (very cyclical), inventory investment (very cyclical).

Government purchases includes government consumption (and in the US also government investments), but not transfer payments (social or unemployment benefits). The split is often done in terms of central versus local government.

Figures 6.4–6.5 show some other business cycle indicators. We see the following.

1. Private consumption of durables and residential investment are much more volatile than GDP (and also aggregate private consumption), and are somewhat leading GDP.

2. Productivity (output per hour) is about as volatile as GDP and somewhat leading.

3. The unemployment rate is not very volatile, is countercyclical and is lagging GDP.

The growth of GDP between two periods ($(Y_t - Y_{t-1})/Y_{t-1}$) is shown in terms of its (demand conponents). By taking differences of (6.1) between two periods (t and $t-1$)

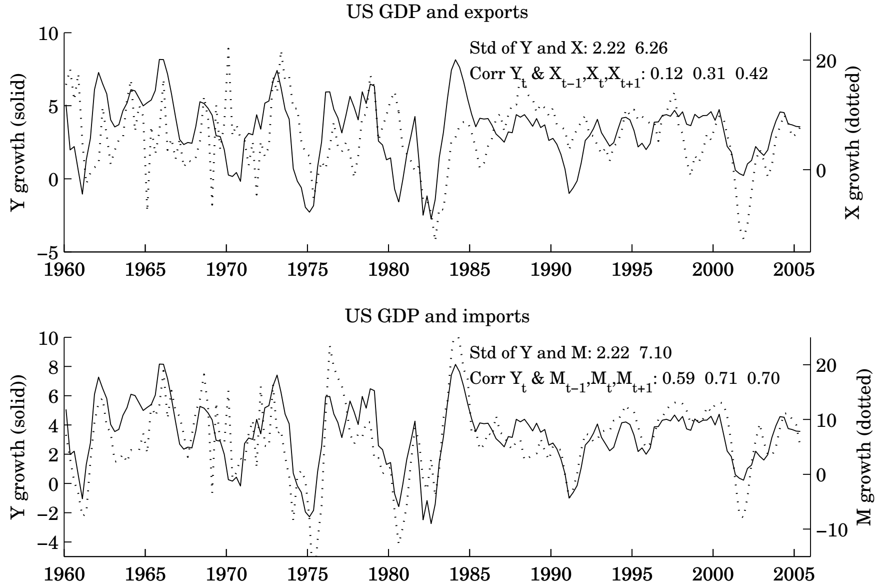


Figure 6.3: US GDP and its components, 4-quarter growth, %

and dividing by the inital value we get the "contributions to GDP growth"

$$
\begin{aligned}
\frac{Y_t - Y_{t-1}}{Y_{t-1}} &= \frac{C_t - C_{t-1}}{Y_{t-1}} + \frac{I_t - I_{t-1}}{Y_{t-1}} + \frac{G_t - G_{t-1}}{Y_{t-1}} + \frac{X_t - X_{t-1}}{Y_{t-1}} - \frac{M_t - M_{t-1}}{Y_{t-1}} \\
&= \frac{C_t - C_{t-1}}{C_{t-1}} \underbrace{\frac{C_{t-1}}{Y_{t-1}}}_{w_C} + \frac{I_t - I_{t-1}}{I_{t-1}} \underbrace{\frac{I_{t-1}}{Y_{t-1}}}_{w_I} + \frac{G_t - G_{t-1}}{G_{t-1}} \underbrace{\frac{G_{t-1}}{Y_{t-1}}}_{w_G} \\
&\quad + \frac{X_t - X_{t-1}}{X_{t-1}} \underbrace{\frac{X_{t-1}}{Y_{t-1}}}_{w_X} - \frac{M_t - M_{t-1}}{M_{t-1}} \underbrace{\frac{M_{t-1}}{Y_{t-1}}}_{w_M}. \qquad (6.2)
\end{aligned}
$$

The second equation shows that GDP growth can be decomposed a sum of the respective growth rates times the shares of GDP. (These shares are fairly constant over time.)

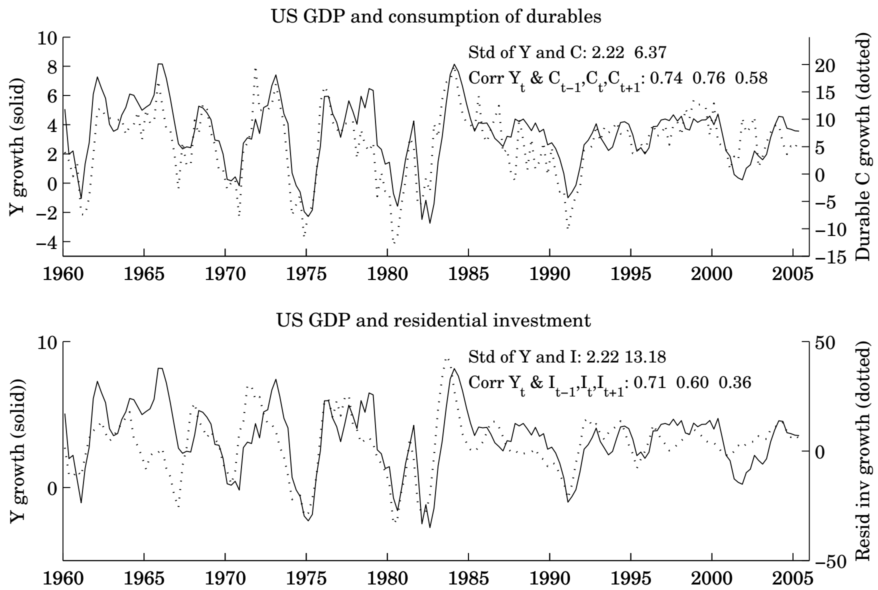


Figure 6.4: US GDP and business cycle indicators, 4-quarter growth, %

6.2 Defining “Recessions”

Main reference: NBER’s Business Cycle Dating Committee (2003) and Abel and Bernanke (1995) 9.1

The National Bureau of Economic Research (NBER) dates “recessions” by a complicated procedure. This chronology has become fairly influential in press and the policy debate, so this section will summarize it. Of course, there are other ways to define recessions—for instance in terms of unemployment or capacity utilization.

According to NBER, a *recession* is the period between a *peak* and a *trough* (low point) in economic conditions—see Figure 6.7. The (typically long) period between two recessions is an *expansion* (or boom) and the peaks and troughs themselves are *turning points*. To qualify as a recession, the downturn should last more than a few months (so a

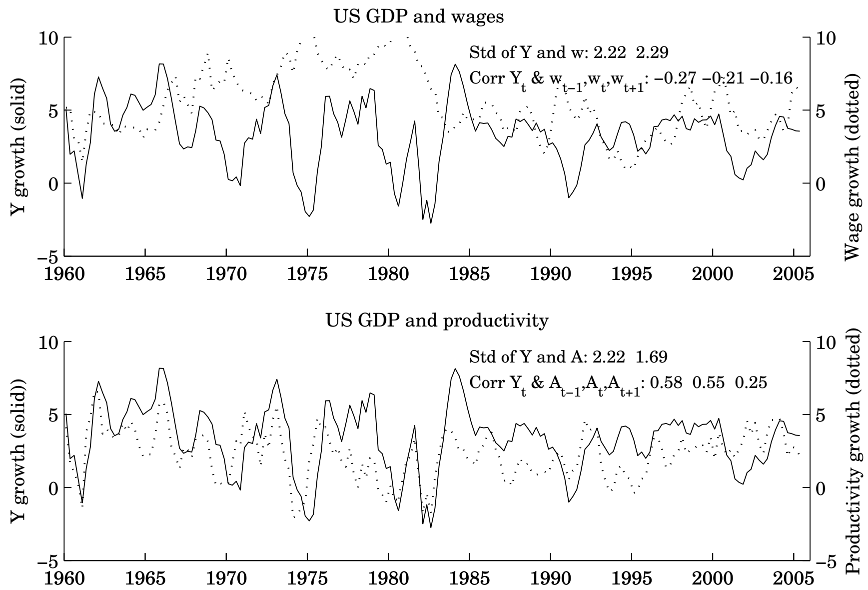


Figure 6.5: US GDP (4-quarter growth, %) and business cycle indicators

temporary dip in production does not count).

The dating is based on a number of different economic variables, but GDP is certainly the most important. A short cut version of the dating procedure says that two consecutive quarters of negative GDP growth is a recession. In practice, the quarter of the turning point is (more or less) governed by the quarterly GDP series (GDP comes only as quarterly and annual data). To narrow down the dating to a month other series are studied. In the latest statement by the dating committee (NBER’s Business Cycle Dating Committee (2003)), emphasis is put on real income (less transfers), employment, industrial production, and real sales of the manufacturing and the wholesale-retail sectors. The dating is committee’s decision comes a long time after the peak, so it is mostly of historical interest. For instance, the November 2001 trough was declared only in July 2003.

See Figure 6.8 for an illustration.

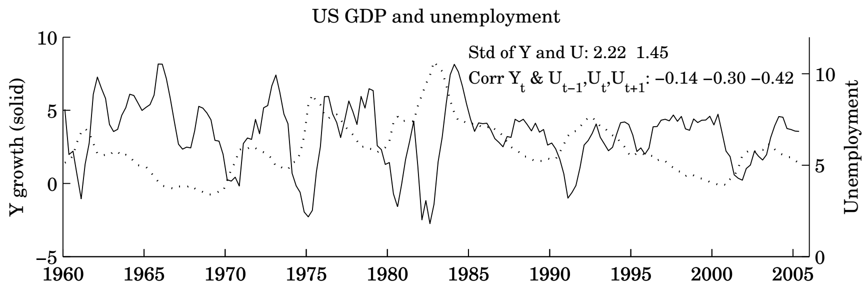


Figure 6.6: US GDP (4-quarter growth, %) and business cycle indicators

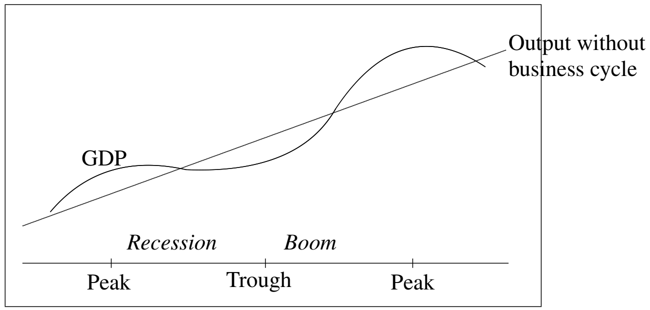


Figure 6.7: Business cycle chronology

7 Data Quality, Survey Data and Indicators

Main reference: The Economist (2000) 4

7.1 Poor and Slow Data: Data Revisions

Macroeconomic data is often both slow and of poor quality. For instance, preliminary national accounts data is typically available only after a quarter (at best), and the subsequent revisions (over the next 2 years or so) can be sizeable. There are much less, if any, revisions in data on some price indices (like CPI) and financial prices.

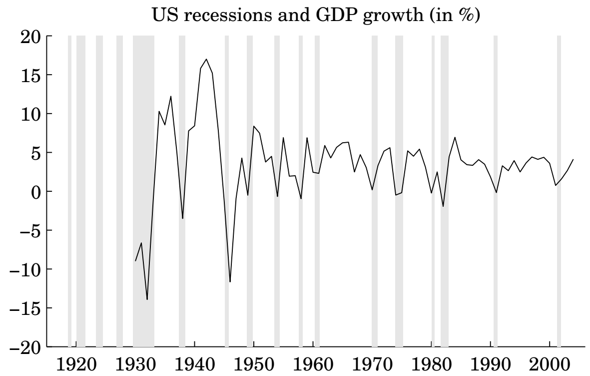


Figure 6.8: US recessions and GDP growth, %

See *Tables 7.1–7.2* and *Figures 7.1–7.2* for an illustration.

The poor quality and considerable lag in publication of important economic statistics influence the forecasting practice in several ways. Second, a whole range of *indicators* are used in the forecasting process. Some of these indicators are official statistics, but of the sort that comes quickly and with relatively good quality (for instance, price data). Some other indicators are based on specially designed surveys, for instance, of the capacity utilization of the manufacturing sector.

7.2 Survey Data

Many forecasting institute collect survey data on such different things like consumer confidence, economists' inflation expectation, and purchase managers' business outlook.

7.2.1 Consumer Surveys

Private consumption is around 2/3 of GDP in most countries. In particular "consumption" of durable goods (in practice purchases, but counted as consumption in the national accounts) is very sensitive to consumers' expectation about future economic conditions.

Consumer surveys are used to estimate the strength of these expectations. The results

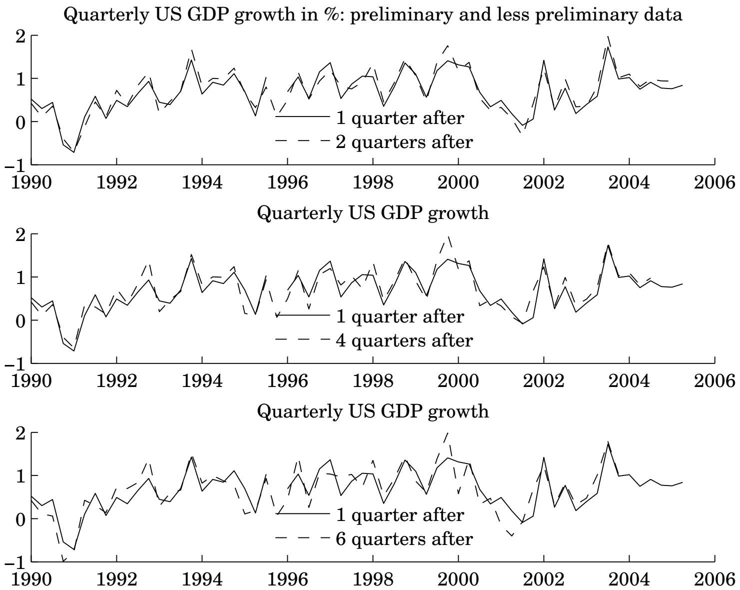


Figure 7.1: Revisions of US GDP, 1-quarter growth, %

from these surveys are often reported as an index based on several qualitative questions. A good survey is characterized by simple and concrete questions.

Surveys of Consumers, University of Michigan (2003) is one of the better known surveys of US consumer confidence (as well as consumer expectations of inflation and other variables). Their index is based on questions about the respondent's own economic situation as well as the country's situation. The answers are qualitative, for instance, either better, same, worse, or don't know (for instance, a year from now compared to today).[1]

[1]The five questions in the index are:

- "We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?" (better, same, worse, don't know)
- "Now looking ahead—do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?" (better, same, worse, don't know)

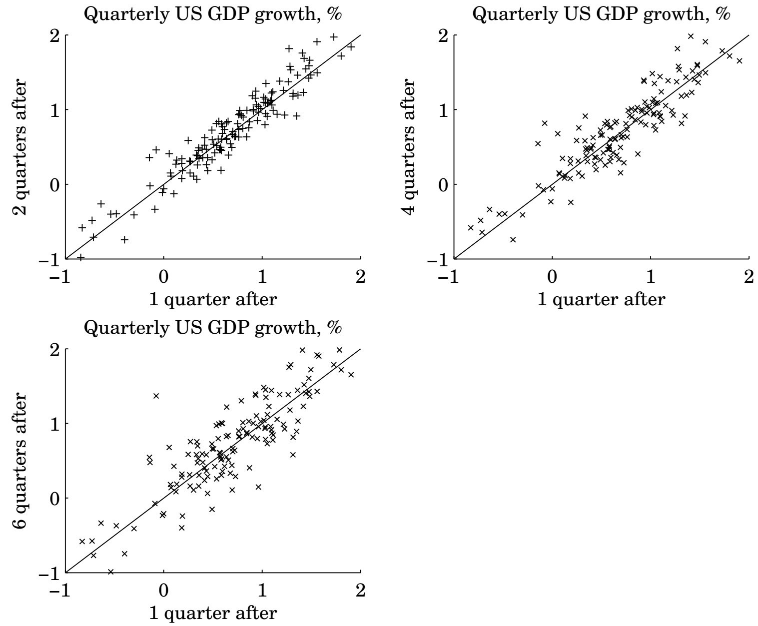


Figure 7.2: Revisions of US GDP, 1-quarter growth, %

The index should therefore be interpreted as indicating expectation over the medium-term horizon. To construct the index, the "balance" of each question is calculated: the

- "Now turning to business conditions in the country as a whole—do you think that during the next twelve months we'll have good times financially, or bad times, or what?" (good times, good with qualifications, pro-con, bad with qualifications, bad times, don't know)
- "Looking ahead, which would you say is more likely—that in the country as a whole we'll have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression,or what?" (open question, but each point must be ranked as good or bad)
- "About the big things people buy for their homes—such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items?" (good, pro-con, bad, don't know)

	1 quarter	2 quarters	4 quarters
GDP	−3.1	0.9	0.7
Private consumption	1.3	1.8	0.2
Government consumption	−0.2	−1.4	−2.6
Investments (business, fixed)	−21.7	−1.5	−0.6
Investments (residential)	−17.4	−14.5	−18.7
Exports	−38.3	−7.9	−5.3
Imports	−17.5	−4.9	−8.0
Money stock (M1)	−8.1	−6.4	−4.0
CPI	0.0	0.0	−0.0

Table 7.1: Mean errors in preliminary data on US growth rates, in basis points (%/100), 1965Q4–. Data 6 quarters after are used as proxies of the 'final' data.

fraction of respondent's who gives positive answers minus the fraction who gives negative answers. Then, the balances for the different questions are summed up (and possibly scaled by some constants) to form a composite index. The resulting series is shown in *Figure 7.3*.

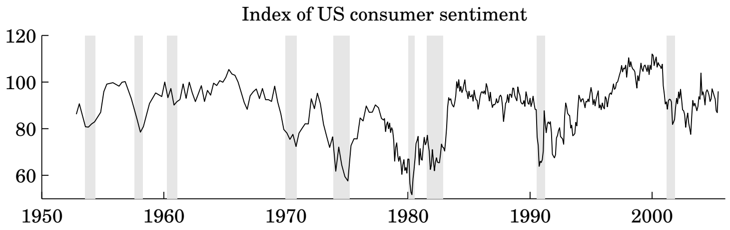


Figure 7.3: US consumer sentiments, University of Michigan survey

7.2.2 Producer Surveys

It is also common to survey producers to ask for current and expected future business conditions. The surveys are typically sent to people close to the production or purchase/sale of the firms—general management tends to give (surprisingly) bad answers. The answers to several questions are often combined into an index.

	1 quarter	2 quarters	4 quarters
GDP	0.87	0.92	0.96
Private consumption	0.87	0.92	0.95
Government consumption	0.75	0.86	0.93
Investments (business, fixed)	0.78	0.89	0.93
Investments (residential)	0.88	0.89	0.95
Exports	0.74	0.93	0.97
Imports	0.63	0.79	0.88
Money stock (M1)	0.94	0.95	0.98
CPI	1.00	1.00	1.00

Table 7.2: Pseudo R^2 of preliminary data on US growth rates, 1965Q4–. Data 6 quarters after are used as proxies of the 'final' data.

As an example, the US Purchase Managers' Index (PMI) from the Institute for Supply Management is closely watched by many forecasters. Purchase managers are asked about the development of new orders, production, employment, supplier deliveries, and inventories. The answers are qualitative: better, same, worse (compared to the pervious month). This means that the index should be calculated as indicated short-run changes in the business cycle.

The index is calculated as the fraction of managers that replies "better" plus 1/2 times the fraction that replies "same." This means that an index value of 50 can be interpreted as a neutral (same) situation, while values above (below) 50% are positive (negative). *Figure 7.4* shows the series.

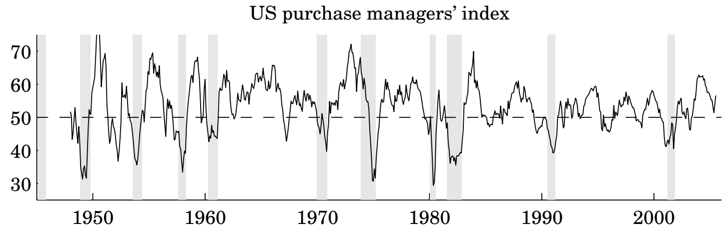


Figure 7.4: US purchase managers' index, Institute for supply management

7.3 Leading and Lagging Indicators

Although business cycles may be due to all sorts of reasons (for instance, oil price shocks, expansionary fiscal policy, technological breakthroughs), there are some fairly consistent patterns in the timing of different series. This is used extensively in forecasting and economic analysis in general. Leading indicators are clearly useful for forecasting, while lagging indicators are useful for confirming the preliminary GDP data (which is always erratic).

For many countries the following pattern seems to hold.

1. Leading indicators:
   - (a) Early leading indicators (more than 3 quarters): interest rates, share prices, business and consumer confidence, housing starts
   - (b) Intermediate leading indicators (2 quarters): consumer credit, car sales, manufacturing orders
   - (c) Late leading indicators (a quarter): retail sales
2. Lagging indicators:
   - (a) Early: capacity utilization, vacancies
   - (b) Intermediate: wages, inflation
   - (c) Late: investment, order backlogs, stocks

Indicators that close to coincident with GDP (either late leading or early lagging) are sometimes referred to as coincident indicators.

Since there are so many leading indicators, it is often convenient to summarize them in an index. For instance, OECD's Composite *index of leading indicators* (see OECD (2003)) for the US is a weighted average of the following series: dwellings started, net new orders for durable goods, share price index, consumer sentiment indicator, weekly hours of work in manufacturing, purchasing managers index, spread of interest rates. For other countries, other (but similar) variables are typically used. In practice, the variables are chosen by statistical techniques from a larger set of "typical" leading indicators.

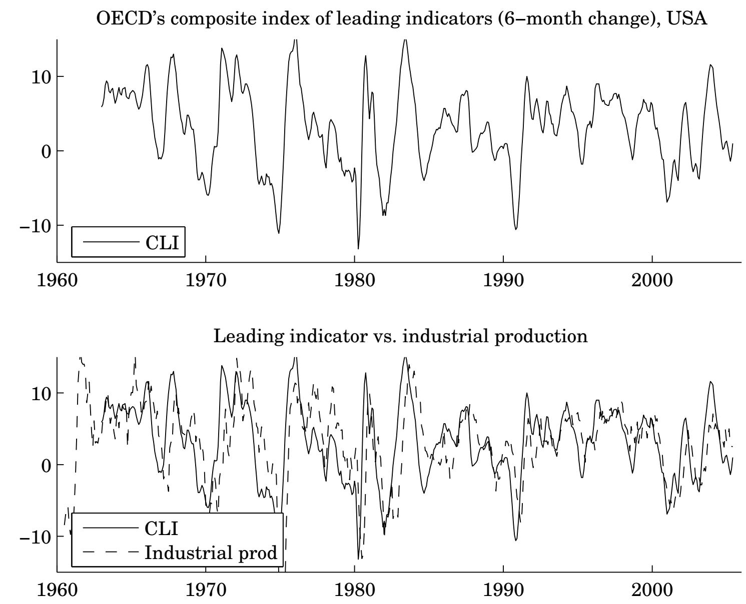


Figure 7.5: OECD's index of leading indicators

The index series is of considerable interest itself, but it is often transformed into some kind of deviation from trend. This makes it easier to see the business cycle movements. For instance, OECD transforms the index by using the ratio between the index series and the moving average of the last 12 months.[2] See *Figure 7.5* for an example.

[2]"The OECD CLI is designed to provide early signals of turning points (peaks and troughs) between expansions and slowdowns of economic activity. The OECD uses the six-month rate of change of the CLI as its preferred pointer to possible turning points. The six-month rate of change of the CLI is calculated by using the ratio between the figure for a given month m and the average of the figures from $m-12$ to $m-1$. Thus, the six-month rate of change is less volatile and provides earlier and clearer signals for future turning points than the CLI itself. In practice, peaks in GDP have been found about nine months (on average) after the signals of peaks had been detected in the six-month rate of change." (OECD (2003))

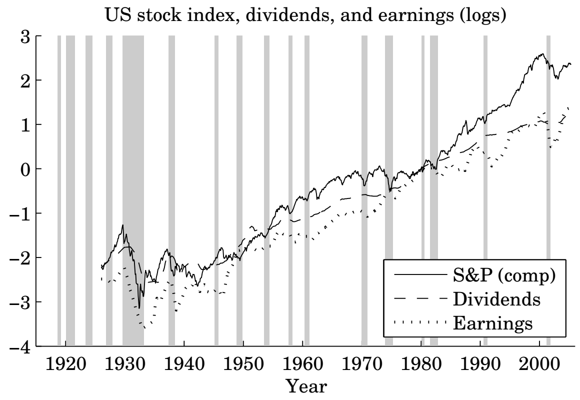


Figure 8.1: US stock index (S&P), dividends, earnings, and NBER recessions (shaded)

8 Using Financial Data in Macroeconomic Forecasting

8.1 Financial Data as Leading Indicators of the Business Cycle

Further reading: Harvey (1989) and Harvey (1993)

Since asset prices are inherently forward looking, they are often used as leading indicators. The idea is that the asset market summarizes a lot of information about future profitability—and that this information indeed has predictive value.

The fundamental stock price is the present value of all expected future dividends. The discounting is made by a risk-adjusted "interest rate," which typically is higher than the riskfree rate: it corresponds to the expected return on the stock. If this discount rate is a constant R, then the fundamental price is

$$P_t = \frac{\mathrm{E}_t\, D_{t+1}}{1+R} + \frac{\mathrm{E}_t\, D_{t+2}}{(1+R)^2} + \frac{\mathrm{E}_t\, D_{t+3}}{(1+R)^3} + \dots. \tag{8.1}$$

The expectations are formed using whatever information we have in t. (If we assume that dividends will grow at a constant rate, then we get the Gordon model and (8.1) can

be simplified further.)[1] This equation suggests that a price increase is driven by upwards revisions of the expected future dividends (earnings). Since earnings are strongly linked to the business cycle, this equation suggests that price changes may signal higher expected GDP. If investors are rational, this could be used as an indicator of a boom.

See *Figure 8.1* for an example.

The main drawbacks are, of course, that price changes can also be driven by movements in discount rates and by non-rational expectations.[2]

In any case, broad stock market indices are often used as leading indicators of the general economic situations. Industry indices are also watched: some industries are known to be procyclical and even leading the general business cycle (building companies, manufacturers of investment goods and cars).

8.2 Nominal Interest Rates as Forecasters of Future Inflation

Reference: Bodie, Kane, and Marcus (2002) 5.1

It is clear that nominal interest rates and inflation are strongly related. See *Figure 8.2* for an example.

The *Fisher equation* says that the nominal interest rate includes compensation both for inflation expectations, $\mathrm{E}_t\, \pi_{t+1}$, the real interest rate, r_t, and possibly a risk premium, ψ_t,

$$i_t \approx \mathrm{E}_t\, \pi_{t+1} + r_t + \psi_t. \tag{8.2}$$

(The approximate equality is an equality if we use continuously compounded rates.)

It is fairly common to use this equation for *eliciting inflation expectations* from nominal interest rates by making simple assumptions about the real interest rate and the risk premium. For instance, if the real interest rate and the risk premium are constant, then changes in interest rates can be interpreted as changes in inflation expectations.

Example 10 *Suppose the nominal interest rate is* $i = 0.07$*, the real interest rate is* $r = 0.03$*, and the nominal bond has no risk premium (*$\psi = 0$*), then the expected inflation is*

[1] To derive (8.1), notice that the pricce can be written as the discounted value of next period's price plus dividends, $P_t = (\mathrm{E}_t\, D_{t+1} + \mathrm{E}_t\, P_{t+1})/(1+R)$. Substitute for P_{t+1} by using $P_{t+1} = (\mathrm{E}_{t+1}\, D_{t+2} + \mathrm{E}_{t+1}\, P_{t+2})/(1+R)$. Repeat this and use the law of iterated expectations, for instance, $\mathrm{E}_t\, \mathrm{E}_{t+1}\, D_{t+2} = \mathrm{E}_t\, D_{t+2}$.

[2] This provides a potential explanation to the following quote: "The stock market has predicted nine out of the last five recessions!" (Paul Samuelson, Newsweek, 19 Sep 1966, p. 92).

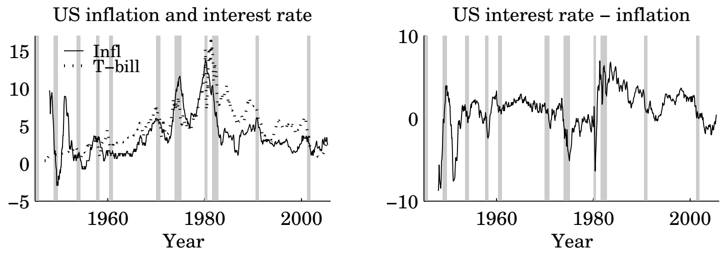


Figure 8.2: US inflation and 3-month interest rate

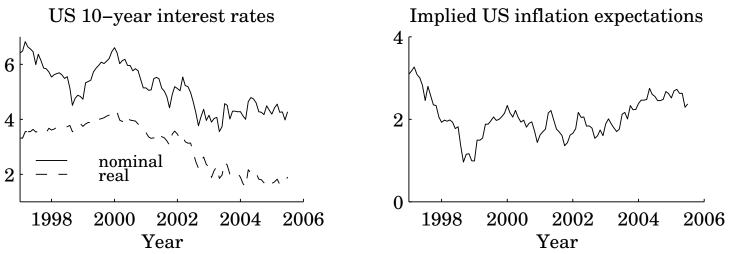


Figure 8.3: US nominal and real interest rates.

$\mathrm{E}_t\, \pi_{t+1} = 0.04$.

A more modern approach is to use the real interest rates from *index-linked bonds*, that is, bonds which give automatic compensation for actual inflation. This gives a number for the real interest rate ($\pm$ some risk premia), which can be used in (8.2) together with nominal interest rates to extract the inflation expectations of the market. See *Figure 8.3* for an example.

Empirical results typically indicate that there are non-trivial movements in the real interest rate and/or risk premia—especially for short forecasting horizons. This holds also when inflation expectations, as measured by surveys, are used as the dependent variable. Inflation expectations seems to vary by less than the interest rate. It is therefore not straightforward to extract inflation expectations from nominal interest rates.

Figure 8.4: Timing convention of forward contract

8.3 Forward Prices as Forecasters of Future Spot Prices

Reference: Bodie, Kane, and Marcus (2002) 22–23

A forward contract stipulates a price to be paid at a future delivery. (A futures contract is very similar.) Quite often, the forward price, F_t, is quoted as a forward premium, that is, it is divided by the current spot price. See Figure 8.4 for an illustration.

The forward price can be thought of as the sum of the expected future spot price, $\mathrm{E}_t\, P_{t+1}$, and a risk premium, ψ_t,

$$F_t \approx \mathrm{E}_t\, P_{t+1} + \psi_t. \tag{8.3}$$

(The approximate equality is actually an equality if we use log prices.)

It is common to use the forward (or futures) price as an indicator of the (market's) expected spot price in the future, that is, to assume a zero (or possibly constant) risk premium. (On the currency market, this is assuming "uncovered interest rate parity.")

Such "market expectations" are often interesting in themselves, but could also be used as forecasts, especially when other forecast methods are bad. The logic is simple: participants on the financial markets have every incentive to make the best possible guess of the future price, so their "forecasts" should contain useful information.

Unfortunately, there are two problems with using forward (or futures) prices as forecasters. First, the forward price contains virtually the same information at the current spot price. The *forward-spot parity*, which is a no-arbitrage relation, says that the forward price equals the current spot price times an interest rate factor. The intuition is simple: the forward contract is just like buying the asset on credit.

Second, the forecasting record of forward prices is mixed. There are three possible reasons: *(i)* there could be sizeable movements in risk premia; *(ii)* the market expectations could be systematically wrong; and *(iii)* even the best forecasting model can be pretty bad.

To illustrate the forecasting power, *Figures 8.5* and *8.6* show some GBP/USD exchange rate data. Let S_t denote the exchange rate—here the number of GBP per USD (the price of USD, in terms of GBP).[3]

They include, among other things the mean squared errors (MSE) of three different forecasting methods where the forecasted depreciation, $\hat{S}_{t+1}/S_t$, is

$$\begin{aligned}\hat{S}_{t+1}/S_t &= \hat{a} + \hat{b}(F_t/S_t) \text{ where } \hat{a} \text{ and } \hat{b} \text{ are LS estimates,}\\ \hat{S}_{t+1}/S_t &= 1 \text{ (assuming } S_t \text{ is a random walk)}\\ \hat{S}_{t+1}/S_t &= F_t/S_t. \end{aligned} \tag{8.4}$$

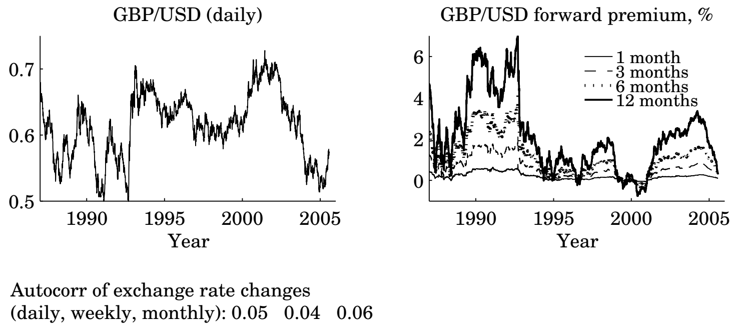


Figure 8.5: GBP/USD exchange rate and interest rate

8.4 Long Interest Rates as Forecasters of Future Short Interest Rates

Reference: Bodie, Kane, and Marcus (2002) 15

Bonds often have coupons—and similar bonds can have different coupons. This makes it difficult to directly analyze bond prices. Instead, prices of different coupon bonds are combined to calculate implied *spot interest rates* (on zero coupon bonds) for different maturities. Sometimes, the implied "STRIPped" zero-coupon bonds are traded on bond markets. In other cases, it is just a calculation.

[3]Let i_t be the domestic (effective) interest rate, i^* the foreign interest rate, and S_t the exchange rate (price of one unit of foreign currency, in terms of domestic currency). A straightforward no-arbitrage argument (covered interest rate parity) then shows that the forward price of the foreign currency must be $F_t = S_t(1+i_t)/(1+i_t^*)$.

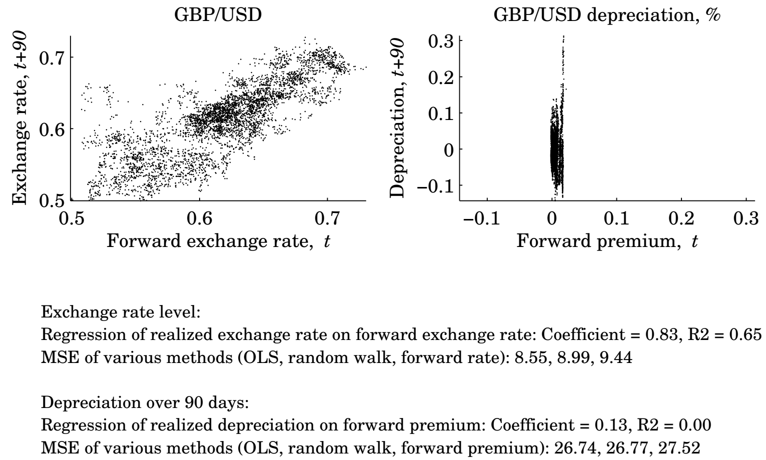


Figure 8.6: GBP/USD forward and realized exchange rate

In a next step, we can form portfolios of zero coupon bonds to calculate the *forward interest rate* that we already today can lock in for an investment between two future dates.

An m-period interest rate, i_{mt}, can be written as an average of the expected short interest rates, i_t, until maturity plus a risk premium, ψ_t,

$$i_{mt} \approx (i_t + \mathrm{E}_t\, i_{t+1} + \ldots + \mathrm{E}_t\, i_{t+m-1})/m + \psi_t. \tag{8.5}$$

(The approximate equality is an equality if we use continuously compounded rates.)

The *expectations hypothesis of interest rates* says that long bonds have no or at least constant risk premia. In that case, a forward interest rate can be interpreted as the expected future short interest rate—and therefore used as a forecast.

The expectations hypothesis has been tested many times, typically by an ex post linear regression (realized interest rates regressed on lagged forward rates). The results typically reject the expectations hypothesis. It is not clear, however, if the rejection is due to systematic risk premia or to fairly small samples (compared to the long swings in interest rates). The expectations hypothesis gets more support when survey data on interest rate expectations is used instead on realized interest rates.

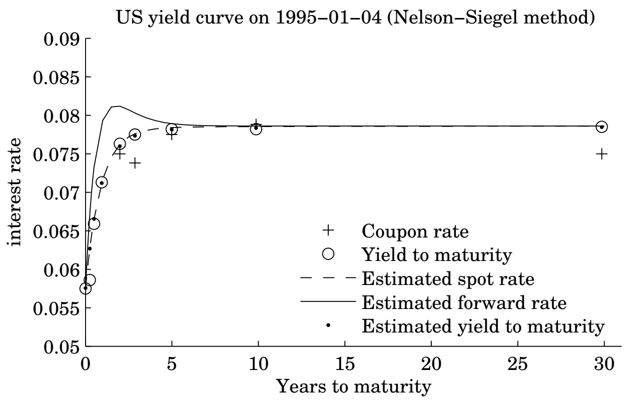


Figure 8.7: Estimated US yield curve

9 Macroeconomic Models

9.1 A Traditional Large Scale Macroeconometric Model

Main reference: Evans (2003) 12; Deutsche Bundesbank (2000)

Development of large scale macroeconometric models started in the 1960s and 1970s—and has continued since, although at a slowing pace. The purpose of the models is to forecast, but there is also a range of secondary task (which the models have proven to be better at).

A large scale macroeconometric model consists of estimated "behavioral" equations (for instance, for private consumption of nondurables) combined with accounting identities (for instance, the GDP definition), and some exogenous variables (for instance, in a

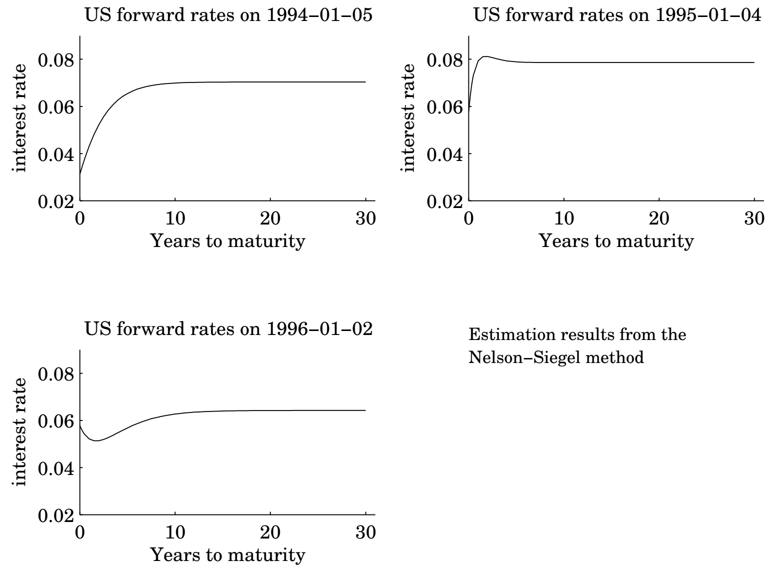


Figure 8.8: Estimated US yield curve, Nelson-Siegel method

model of a small open economy, world GDP growth is typically "fed in" to the model as an exogenous variable).

Since most of these models go back several decades, so does the philosophy of the models: they essentially capture the Keynesian idea of a demand determined economy. Most of the estimated equations are for the demand side of the economy. For instance,

- export equations are estimated with world GDP (or some other measure of demand pressure) and the relative price as regressors;
- investment equations are estimated by using output or capacity utilization and interest rates are regressors;
- private consumption equations are estimated by using disposable private income and perhaps some proxy for wealth as regressors;

- import equations are estimated by using the other GDP components (demand pressure) and a relative price as regressors;

- government consumption is typically fed manually into the model.

These regressions are often estimated by OLS (or possibly the instrumental variables or 2SLS method). Several lags of the regressors are used to capture the inertia (for instance, in the effect of relative price changes on imports).

The supply side was initially (in the 1960s and 1970s) suppressed—but has since been developed. There are two main approaches. The first focuses on the price setting of firms, the second on wage formulation. In the first approach, the supply curve (of a firm with some market power) is inverted to get an expression of the price level in terms of production costs and the demand pressure—a Phillips curve for price inflation. In practice, this means

- running a regression of price inflation on demand variables and production costs. This is sometimes done for several production sectors.

In models with many production sectors, the production costs of one sector involve the prices of many other sectors that supply intermediate goods (including imported goods). To model this, input-output matrices from historical data are used. These tell how much of the cost in sector A (say) that depends on sectors B and C (assuming 3 sectors), but also how much of the cost in sector B that depends on sectors A and C—and similarly for sector C. This results in a large number of equations—but with relatively little action. It is therefore no surprise that multi-sector models easily can have many hundreds of equations—or more.

In the second approach, the labour market is modelled by inverting a labour supply function, so that lower unemployment (higher labour supply) is associated with higher wages—a Phillips curve for wage inflation. In practice, this means

- running a regression of wage inflation on unemployment.

It is probably true to say that these large scale macroeconometric models have proved less successful as forecasting instruments than initially believed. The models often need substantial manual guidance in order to produce reasonable forecasts—at least beyond

the next few quarters. This manual guidance takes the form of either shutting down some behavioral equations or by introducing "add factors" to these equations—both of which require judgemental input.

As these models are quite expensive to maintain, the interest in them has recently decreased. Many institutions still maintain them, however. The reason is that they have proved to have positive by-products. First, such a model is probably the best tool for constructing alternative scenarios—once it has been disciplined to (re-)produce a reasonable basic (forecasting) scenario. Second, the model development teams have been good at organizing data bases—in some cases so good that the models have become side shows to the data bases. Thirdly, the model development has also generated plenty of good applied research of specific macroeconomic problems (for instance, "how does a stock market boom affect private consumption?") that has proved useful in the judgmental forecasting process.

9.2 A Modern Aggregate Macro Model

Modern macro models are often smaller than the older macroeconometric models and they pay more attention to both the supply side of the economy and the role of expectations. These models try to capture the key elements in the way central banks (and most other observers) reason about the interaction between inflation, output, and monetary policy.

In these models, inflation depends on expected future inflation (some prices are set today for a long period and will therefore be affected about expectations about future costs and competitors' prices), lagged inflation, and a "Phillips effect" where an *output gap* (output less trend output) affects price setting via demand pressure. For instance, inflation (π_t) is often modelled as

$$\pi_t = \alpha \operatorname{E}_t \pi_{t+1} + \beta \pi_{t-1} + \phi x_t + \varepsilon_{\pi t}, \tag{9.1}$$

where x_t is the output gap and $\varepsilon_{\pi t}$ can be interpreted as "cost push" shocks (wage demands, oil price shocks). This equation can be said to represent the supply side of the economy and it is typically derived from a model where firms with some market power want to equate marginal revenues and marginal costs, but choose to change prices only gradually.

The demand side of the economy is modelled from consumers' savings decision, where the trade off between consumption today and tomorrow depends on the real interest rates. Simplifying by setting consumption equal to output we get something like the following equation for the output gap

$$x_t = x_{t-1} - \gamma(i_t - \mathrm{E}_t\pi_{t+1}) + u_t, \tag{9.2}$$

where i_t is the nominal interest rate (set by the central bank) and u_t is a shock to demand. Note that the expected *real* interest rate affects demand (negatively).

In some cases, the real exchange rate is added to both (9.1) and (9.2), capturing price increases on imported goods and foreign demand for exports, respectively. The exchange rate is then linked to the rest of the model via an assumption of uncovered interest rate parity (that is, expected exchange rate depreciation equals the interest rate differential).

Some of the important features of this simple model are: *(i)* inflation expectations matter for today's inflation (think about wage inflation), *(ii)* the instrument for monetary policy, the short interest rate i_t, can ultimately affect inflation only via the output gap; *(iii)* it is the real, not the nominal, interest rate that matters for demand.

To make the model operational, two more things must be added: the monetary policy (the way the interest rate is set) and the expectations in (9.1)–(9.2) must be specified.

It is common to assume that the central bank has some instrument rule like the famous "Taylor rule"

$$i_t = \theta_0 + 0.5x_t + 1.5\pi_t + v_t. \tag{9.3}$$

The residual v_t is a "monetary policy shock," which picks up factors left out of the model (for instance, the central bank's concern for the banking sector or simply changes in the central bank's objectives). This simple reaction function has been able to track US monetary policy fairly well over the last decade or so. Another approach to find a policy rule is to assume that the central bank has some loss function that it minimizes by choosing an policy rule. This loss function is often a weighted average of the variance of inflation and the variance of the output gap. The policy rule is the solution of the minimization problem, and can often look more complicated than the Taylor rule.

The expectations in (9.1)–(9.2) can be handled in many ways. The perhaps most straightforward way is to assume that the expectations about the future equal the current value of the same variable (a "random walk"). A more satisfactory way is to use survey data on inflation expectations. Finally, many model builders assume that expectations are

"rational" (or "model consistent") in the sense that the expectation equals the best guess we could do under the assumption that the model is correct. This latter approach typically requires a sophisticated way of solving the model (as the model both generates the best guesses and depends on them).

9.3 Forecasting Inflation

The Phillips curve (9.1) provides the general idea of inflation forecasting. The inputs needed are *(i)* inflation expectations (from surveys or from nominal/real interest rates), *(ii)* business cycle conditions (that is, a GDP forecast), *(iii)* lagged inflation and wages plus *(iv)* everything that was loaded into the "cost push shock."

It is often thought that the effect of the business cycle conditions (GDP growth) on inflation is stronger and quicker than the reverse effect. For this reason, the GDP forecast is typically produced first, and then used as an input in the inflation forecast.

All the terms that were collected under the "cost push shock" label include things like productivity, wage costs and costs of imported raw material and intermediate goods (including the exchange rate effect).

Productivity is typically forecasted with inputs from the forecast of the business cycle conditions—it is often assumed that productivity picks up in the early phase of the boom (as capacity utilization increases from a low level), and then levels off (as the capacity ceiling is reached).

Wage costs are forecasted on the basis of existing wage contracts and forecasts of labour market tightness (typically lagging the business cycle with a few quarters).

The costs for imported raw material and intermediate are based on an assessment about the global business cycle, survey data and quite often also prices on futures contracts.

The Phillips curve is essentially a model for domestic producer inflation. If we want to forecast inflation at the consumer level, we have to factor in inflation on imported goods (similar to the imported raw material and intermediate goods). If in addition, we want to forecast the headline inflation (CPI), then the tax effects must also be incorporated.

9.4 Forecasting Monetary Policy

Main reference: Mozina (1998); The Economist (2000) 12
Further reading: Kettell (1999)

Monetary policy is a crucial part of the macroeconomic picture these days, so it is important to understand how monetary policy is formed. It has not always been this way: there are long periods when many countries adopted a very simple (or so it seemed) monetary policy by pegging the currency to another currency. Macroeconomic policy was then synonymous with fiscal policy. Recently, the roles have changed.

The *Taylor rule* (9.3) is surprisingly good at describing the large movements in the short interest rate—at least if it changed a bit to allow for inertia in policy: the actual short interest rate is then described as a weighted average of the right hand side of (9.3) and the short interest rate in the previous period.

Many central banks and central bank watchers use a *monetary conditions index* (MCI) to measure the stance of monetary policy. It is typically defined as

$$MCI_t = \alpha\,(i_t - i_0) + (1-\alpha)\,(s_t - s_0)\,, \tag{9.4}$$

where $i_t - i_0$ is the change in the nominal interest rate since some base period, and $s_t - s_0$ the corresponding change in the (trade weighted) exchange rate.

The idea of the MCI is to summarize the effect of monetary policy (which typically sets either the interest rate or the exchange rate, but heavily affects the other) on GDP (and less often, inflation). The weights in (9.4), α and $1-\alpha$, are therefore estimated on historical data. Typically, they come from an estimated IS curve

$$GDP_t = \beta_0 + \beta_1 i_t + \beta_2 s_t + \beta_3 \times \text{other things}, \tag{9.5}$$

by taking $\alpha = \beta_1/(\beta_1 + \beta_2)$. Typical values of α are 9/10 for U.S. and around 3/4 for most European countries.

The changes from a base period in the MCI (9.4) makes a comparison across countries problematic. Sometimes the nominal rates are replaced by real rates; and the base period values might be defined as some moving average (perhaps reflecting a perceived "equilibrium").

So, what Drives Interest Rates? There is a two-way causality: inflation and the real economy (which depend on the real interest rate) affect monetary policy (for instance, the

Taylor rule in (9.3)), and monetary policy can surely affect inflation and the real economy (for instance, as in (9.1) and (9.2)). This makes it difficult to forecast interest rates. However, for short term forecasting, the emphasis is typically on forecasting the next monetary policy move. Long run forecasting rely more on understanding the determinants of real interest rates and inflation, which depends on the general business cycle prospects, but also on the long run stance of monetary policy ("tough on inflation or not?").

9.5 VAR Models

VAR models with a few macro variables (for instance, GDP, consumption, interest rates) have often been found to forecast as well as large-scale macroeconometric models. This is impressive, especially since a small VAR model is very easy to estimate and it is a straightforward exercise to produce forecasts.

The main drawback is that VAR models cannot (without modification) explain (in an economically interesting way) why a forecast is positive, or what would happen if we instead assumed another path for fiscal policy. In practice, this means that VAR models are typically only used as early inputs to a judgemental forecast—perhaps to provide an early suggestion of the direction of the forecast.

A Details on the Financial Parity Conditions*

A.1 Expectations Hypothesis and Forward Prices

Let P_t be the price of an asset in t, F_t the *forward price* of the same asset for delivery in $t+1$, and suppose the interest rate is Y_t. If there are no dividends between t and $t+1$, then the following *forward-spot parity* (or cost-of-carry relationship) must hold

$$(1 + Y_t)P_t = F_t,$$

so the forward price is higher than the spot price if the nominal interest rate is positive.

Proof. Consider the following riskless portfolio in t: buy one asset (cash flow: $-P_t$), borrow P_t (cash flow: P_t), and promise to sell forward in $t+1$ (cash flow: 0). The payoff in $t+1$ is: sell as promised (cash flow: F_t) and repay the loan (cash flow: $-(1+Y_t)P_t$). This portfolio has a zero cash flow in t, and the payoff in $t+1$ is certain (both the forward price and the interest rate are known). The payoff in $t+1$ must therefore be zero: $F_t = (1+Y_t)P_t$. ■

Example 11 *Let the period length be a year (which is the common practice for interest rates). Suppose* $Y_t = 0.05$ *and that the spot price is* $P_t = 2.25$*. From the forward-spot parity (8.3) we then must have the forward price* $F_t = (1+0.05)2.25 \approx 2.36$.

We can always write the expected value of the future asset price as

$$F_t(1+\psi_t) = \mathrm{E}_t\, P_{t+1},$$

where we will show that ψ_t can be interpreted as a risk premium on the asset.

The *pure expectations hypothesis* (PEH) states that the forward price equals the expected asset price, that is, that $\psi_t = 0$. It is not an arbitrage condition, but rather an assumption about risk neutral pricing of the forward contract. This relation is often used as a benchmark. If ψ_t is constant, but different from zero, then the *expectations hypothesis* (EH, no P since not pure) is said to hold.

Combining gives

$$(1+Y_t)(1+\psi_t) = \frac{\mathrm{E}_t\, P_{t+1}}{P_t},$$

which says that the expected gross return on the asset (right hand side) equals the nominal interest rate times the risk premium. The pure expectations hypothesis, $\psi_t = 0$, thus amounts to assuming that the expected return equals the riskfree interest rate, or equivalently, the forward price equals the expected price.

If the asset contains positive systematic risk (in CAPM, this would show up in the form of a positive "beta"), then investors will then require a higher expected return than the riskfree interest rate. In this case $\psi_t > 0$.

A.2 Covered and Uncovered Interest Rate Parity

The forward-spot parity for foreign exchange is *called covered interest rate parity*. Suppose we want to study a 1-period investment. In this case, the forward price F_t is the price (in domestic currency) of getting one unit of foreign currency in $t+1$. Let Z_t denote the cost of the "spot way" of making sure that we have one unit of foreign currency in $t+1$

(to be discussed in detail below). The forward-spot parity is

$$(1+Y_t)Z_t = F_t.$$

The "spot way" of making sure that we have one unit of foreign currency in $t+1$ is to buy $1/(1+Y_t^*)$ units of foreign currency today, where Y_t^* is the foreign 1-period (effective, per period) interest rate. Since the price of foreign currency is S_t, the total cost is $Z_t = S_t/(1+Y_t^*)$. The covered interest rate parity condition can then be written

$$(1+Y_t)\frac{S_t}{1+Y_t^*} = F_t.$$

Extensions to other horizons are straightforward.

Example 12 *Let the period be a year. Suppose* $S_t = 9$, $Y_t^* = 0.04$*, and* $Y_t = 0.07$*. The forward price is then*

$$(1+0.07)\frac{9}{1+0.04} \approx 9.26.$$

The expectations hypothesis of exchange rates is called *uncovered interest rate parity* (UIP). It says that the forward exchange rate equals the expected future exchange rate (possibly times a constant risk premium). Using this in the CIP and rearranging gives

$$\frac{1+Y_t}{1+Y_t^*} = \frac{F_t}{S_t} = \frac{\mathrm{E}_t\, S_{t+1}}{S_t}.$$

The term F_t/S_t is called the forward premium.

Example 13 *Assume that the risk premium is zero. The expected exchange rate depreciation over the next year is*

$$\frac{\mathrm{E}_t\, S_{t+1}}{S_t} = \frac{1+0.07}{1+0.04} \approx 1.028,$$

so the expected depreciation is 2.8%.

A.3 Bonds, Zero Coupon Interest Rates

Bonds often have coupons—and similar bonds can have different coupons. This makes it difficult to directly analyze bond prices. Instead, prices of different coupon bonds are

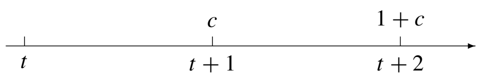


Figure A.1: Timing convention of a coupon bond

combined to calculate implied spot interest rates (on zero coupon bonds) for different maturities. Sometimes, the implied "STRIPped" zero-coupon bonds are traded on bond markets. In other cases, it is just a calculation.

Once we know the spot interest rates (on zero coupon bonds) for different maturities, they can be recombined in another way to calculate *forward interest rates*: the interest rate between two future dates that we can write a contract on already today. Forward interest rates are calculated the spot-forward parity formula.

A.3.1 Coupon Bonds and the Yield Curve

Consider a bond which pays coupons, c, for 2 periods ($t+1$, $t+2$), and one unit of account (the "face value" or "par value") in the last period $t+2$.

The coupon bond is, in fact, a portfolio of zero coupon bonds: c maturing in $t+1$, c in $t+2$, and also 1 in $t+2$. The price of the coupon bond must therefore equal the price of the portfolio

$$B_t^c(2) = B_t(1)c + B_t(2)c + B_t(2).$$

The extension to a multiperiod bond is straightforward.

Example 14 *Suppose* $B_t(1) = 0.95$ *and* $B_t(2) = 0.90$. *The price of a bond with a 6% annual coupon with two years to maturity is then*

$$0.95 \times 0.06 + 0.90 \times 0.06 + 0.90 \approx 1.01.$$

The interest rates on coupon bonds that are quoted in the financial press are typically a yield to maturity, θ, (the internal rate of return), which for the 2-period bond satisfies

$$B_t^c(2) = \frac{c}{(1+\theta)} + \frac{c}{(1+\theta)^2} + \frac{1}{(1+\theta)^2}.$$

Solving this equation for the unknown θ typically requires a non-linear (iterative) method. A first analysis of interest rates starts with the yield to maturity, but we can often extract more information from the bond prices by calculating implicit spot interest rates.

Conversely, if we knew all the spot interest rates, then it would be easy to calculate the correct price of the coupon bond. However, the situation is typically the reverse: we know prices on several coupon bonds (different maturities and coupons), and want to calculate the spot interest rates that are compatible with them. This is to *estimate the yield curve*. The implied zero coupon bond prices are often called the *discount function*.

We can sometimes calculate large portions of the yield curve directly from asset prices, that is, by the *bootstrapping method*. The following example illustrates that.

Example 15 *Suppose we know that* $B_t(1) = 0.95$ *and that the price of a bond with a 6% annual coupon with two years to maturity is 1.01. Since the coupon bond must be priced as*

$$0.95 \times 0.06 + B_t(2) \times 0.06 + B_t(2) = 1.01,$$

we can solve for the price of a two-period zero coupon bond as $B(2) = 0.90$. *The spot interest rate are then*

$$\frac{1}{0.95} = 1 + Y_t(1) \text{ or } Y_t(1) \approx 0.053$$
$$\frac{1}{0.90} = [1 + Y_t(2)]^2 \text{ or } Y_t(2) \approx 0.054.$$

Unfortunately, the bootstrapping approach is not always possible to use. In particular, there may be gaps between the available maturities, so that some of the coupon payments cannot be valued. In this case, the spot interest rates are interpolated by fitting the bond data to some model (there are many) of the yield curve.

Forward Interest Rates

Suppose we are interested in analyzing the forward price for delivery in $t+1$ of a zero coupon bond which matures in $t+2$.

Recall that the general forward-spot parity is

$$[1 + Y_t(1)]B_t(2) = F_t.$$

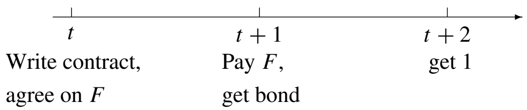


Figure A.2: Timing convention of an interest rate forward contract

This says that there are two equivalent ways of making sure (in period t) that you have a bond in $t+1$ that matures in $t+2$: the left hand side says you borrow money and buy the bond now; the right hand side says that you enter a forward contract.

It is often informative to transform these prices into interest rates instead. The forward interest rate, Y_t^f, is defined similarly to a standard interest rate

$$1/F_t = 1 + Y_t^f.$$

The forward rate is the rate of return that we get from $t+1$ (when we pay the forward price) to $t+2$ (when we get the principal). The only news is that we "lock in" this return already in period t (by entering the forward contract).

Combine to write the forward interest rate as

$$1 + Y_t^f = \frac{[1 + Y_t(2)]^2}{1 + Y_t(1)}.$$

Extensions to other maturities are straightforward.

Example 16 *Let* $m = 1$ *(one year) and* $n = 2$ *(two years), and suppose that* $Y_t(1) = 0.04$ *and* $Y_t(2) = 0.05$*. Then we have*

$$\frac{(1 + 0.05)^2}{1 + 0.04} \approx 1.06,$$

which gives $Y_t^f(1, 2) \approx 0.06$.

Bibliography

Abel, A. B., and B. S. Bernanke, 1995, *Macroeconomics*, Addison Wesley, 2nd edn.

Bodie, Z., A. Kane, and A. J. Marcus, 2002, *Investments*, McGraw-Hill/Irwin, Boston, 5th edn.

Deutsche Bundesbank, 2000, "Macro-Econometric Multi-Country Model: MEMMOD," monograph, Deutsche Bundesbank.

Evans, M. K., 2003, *Practical Business Forecasting*, Blackwell Publishing.

Harvey, C. R., 1989, "Forecasts of Economic Growth from the Bond and Stock Markets," *Financial Analysts Journal*, pp. 38–45.

Harvey, C. R., 1993, "Term Structure Forecasts Economic Growth," *Financial Analysts Journal*, pp. 6–8.

Kettell, B., 1999, *Fed-Watching: Making Sense of Market Information*, Financial Times/Prentice Hall.

Mozina, D., 1998, "U.S. Monetary Policy - A Compendium of Guides," Merrill Lynch Special Reports, http://www.ml.com/cicg/research/cbmt/spclrpts.htm.

NBER's Business Cycle Dating Committee, 2003, "The NBERs Recession Dating Procedure," National Bureau of Economic Research, http://www.nber.org/cycles/recessions.html.

OECD, 2003, http://www.oecd.org.

Surveys of Consumers, University of Michigan, 2003, Surveys of Consumers, http://www.sca.isr.umich.edu/.

The Economist, 2000, *Guide to Economic Indicators*, Profile Books Ltd, London, 4th edn.

10 Stock (Equity) Prices

More advanced material is denoted by a star (*). It is not required reading.

10.1 Returns and the Efficient Market Hypothesis

10.1.1 Prices, Dividends, and Returns

Let P_t be the price of an asset at the end of period t, after any dividends in t has been paid (an ex-dividend price). The gross return ($1 + R_{t+1}$, like 1.05) of holding an asset with dividends (per current share), D_{t+1}, between t and $t+1$ is then defined as

$$1 + R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}. \tag{10.1}$$

The dividend can, of course, be zero in a particular period, so this formulation encompasses the case of daily stock prices with annual dividend payments.

10.1.2 The Efficient Market Hypothesis

The efficient market hypothesis (EFM) casts a long shadow on every attempt to forecast asset prices. In its simplest form it says that it is not possible to forecast asset price changes (or returns), but there are several other forms with different implications. The (semi-strong form of) EFM has two building blocks: *(i)* returns should be unpredictable (because of speculation/arbitrage) on a market with rational expectations (all public information is used efficiently); *(ii)* expectations are indeed rational.

These assumptions have recently been challenged on both theoretical and empirical grounds. For instance, most asset pricing models (including the capital asset pricing model, CAPM) suggest that risk premia (expected excess returns) should vary with the volatility of the market—and we know that volatility does change (from option data and simple time series methods). Movements in expected excess returns are the same as predictability (if expectations are rational). Moreover, there is new evidence on predictability of returns—especially for medium-term and long-term investment horizons (the business

cycle frequency, a few years).

Most tests of predictability focus on excess returns, since this is easier to tie to a theory (changing risk premia)—and also because it circumvents the problem of long-run changes in inflation (excess returns are real). In practice, the results for (nominal or real) returns and excess returns are fairly similar since the movements in most asset returns are much greater than the movements in interest rates.

10.2 Time Series Models of Stock Returns

Main reference: Bodie, Kane, and Marcus (2002) 12–13
Further reading: Cuthbertson (1996) 5 and 6.1; Campbell, Lo, and MacKinlay (1997) 2 and 7; Dunis (1996) (high frequency data, volatility, neural networks); and the papers cited in the text

This section summarizes some evidence which seems to hold for both returns and returns in excess of a riskfree rate (an interest rate). For illustrations, see *Figures 10.1–10.3*.

1. The empirical evidence suggests some, but weak, positive autocorrelation in *short horizon* returns (one day up to a month)—probably too little to be able to trade on. The autocorrelation is stronger for small than for large firms (perhaps no autocorrelation at all for weekly or longer returns in large firms). This implies that equally weighted stock indices have larger autocorrelation than value-weighted indices.

2. There seems to be negative autocorrelation for *multi-year* stock returns, for instance in 5-year US returns for 1926-1985. It is unclear what drives this result, however. It could well be an artifact of just a few extreme episodes (Great Depression). Moreover, the estimates are very uncertain as there are very few (non-overlapping) multi-year returns even in a long sample—the results could be just a fluke.

3. The aggregate stock market returns, that is, a return on a value-weighted stock index, seems to be forecastable on the medium horizon by various *information variables*. This is typically studied by running a regression of the return of an investment starting in t and ending in $t+k$, $R_{t+k}(k)$, on the current value of the

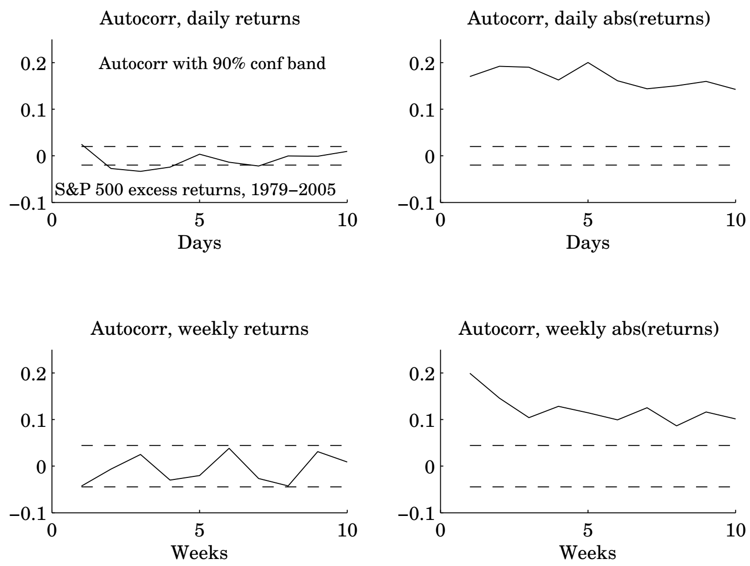


Figure 10.1: Predictability of US stock returns

information variable

$$R_{t+k}(k) = \beta_0 + \beta_1(D_t/P_t) + \varepsilon_{t+k}. \tag{10.2}$$

In particular, future stock returns seem to be predictable by the current dividend-price ratio and earnings-price ratios (positively, one to several years), or by the interest rate changes (negatively, up to a year).

4. Even if short-run returns, R_{t+1}, are fairly hard to forecast, it is often fairly easy to forecast *volatility* as measured by $|R_{t+1}|$ (the absolute value) or R^2_{t+1}. This could possibly be used for dynamic trading strategies on options which directly price volatility. For instance, buying both a call and a put option (a "straddle" or a "strangle"), is a bet on a large price movement (in any direction).

5. It is sometimes found that stock prices behave differently in periods with high

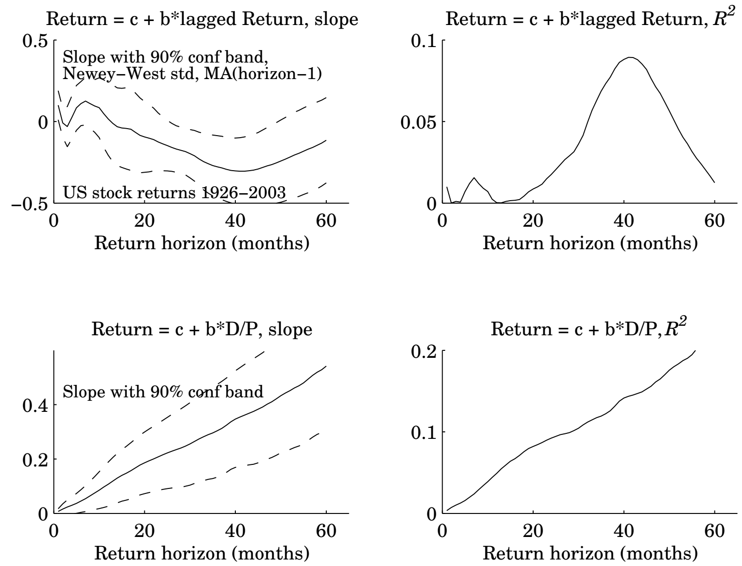


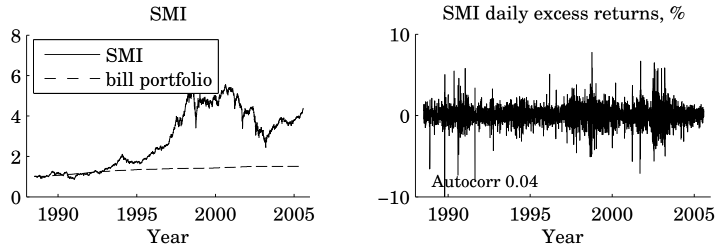
Figure 10.2: Predictability of US stock returns

volatility than in more normal periods. Granger (1992) reports that the forecasting performance is sometimes improved by using different forecasting models for these two regimes. A simple and straightforward way to estimate a model for periods of normal volatility is to simply throw out data for volatile periods (and other exceptional events).

6. There are also a number of strange patterns ("anomalies") like the small-firms-in-January effect (high returns on these in the first part of January).

10.3 Technical Analysis

Main reference: Bodie, Kane, and Marcus (2002); Neely (1997) (overview, foreign exchange market)



Daily SMI data, 1988–2005

Autocorr of returns (daily, weekly, monthly): 0.04 −0.05 0.03

Autocorr of absolute returns (daily, weekly, monthly): 0.27 0.26 0.19

Figure 10.3: SMI

Further reading: Murphy (1999) (practical, a believer's view); The Economist (1993) (overview, the perspective of the early 1990s); Brock, Lakonishok, and LeBaron (1992) (empirical, stock market); Lo, Mamaysky, and Wang (2000) (academic article on return distributions for "technical portfolios")

10.3.1 General Idea of Technical Analysis

Technical analysis is typically a data mining exercise which looks for local trends or systematic non-linear patterns. The basic idea is that markets are not instantaneously efficient: prices react somewhat slowly and predictably to news. The logic is essentially that an observed price move must be due some news (exactly which is not very important) and that old patterns can tell us where the price will move in the near future. This is an attempt to gather more detailed information than that used by the market as a whole. In practice, the technical analysis amounts plotting different transformations (for instance, a moving average) of prices—and to spot known patterns. This section summarizes some simple trading rules that are used.

10.3.2 Technical Analysis and Local Trends

Many trading rules rely on some kind of local trend which can be thought of as positive autocorrelation in price movements (also called momentum[1]).

A *filter rule* like "buy after an increase of $x\%$ and sell after a decrease of $y\%$" is clearly based on the perception that the current price movement will continue.

A *moving average rule* is to buy if a short moving average (equally weighted or exponentially weighted) goes above a long moving average. The idea is that this event signals a new upward trend. The difference between the two moving averages is called an *oscillator* (or sometimes, moving average convergence divergence[2]). A version of the moving average oscillator is the *relative strength index*[3], which is the ratio of average price level on "up" days to the average price on "down" days—during the last z (14 perhaps) days.

The *trading range break-out rule* typically amounts to buying when the price rises above a previous peak (local maximum). The idea is that a previous peak is a *resistance level* in the sense that some investors are willing to sell when the price reaches that value (perhaps because they believe that prices cannot pass this level; clear risk of circular reasoning or self-fulfilling prophecies; round numbers often play the role as resistance levels). Once this artificial resistance level has been broken, the price can possibly rise substantially. On the downside, a *support level* plays the same role: some investors are willing to buy when the price reaches that value.

When the price is already trending up, then the trading range break-out rule may be replaced by a *channel rule*, which works as follows. First, draw a *trend line* through previous lows and a *channel line* through previous peaks. Extend these lines. If the price moves above the channel (band) defined by these lines, then buy. A version of this is to define the channel by a *Bollinger band*, which is ± 2 standard deviations from a moving data window around a moving average.

A *head and shoulder* pattern is a sequence of three peaks (left shoulder, head, right shoulder), where the middle one (the head) is the highest, with two local lows in between on approximately the same level (neck line). (Easier to draw than to explain in a thousand words.) If the price subsequently goes below the neckline, then it is thought that a negative

[1]In physics, momentum equals the mass times speed.

[2]Yes, the rumour is true: the tribe of chartists is on the verge of developing their very own language.

[3]Not to be confused with relative strength, which typically refers to the ratio of two different asset prices (for instance, an equity compared to the market).

trend has been initiated. (An inverse head and shoulder has the inverse pattern.)

Clearly, we can replace "buy" in the previous rules with something more aggressive, for instance, replace a short position with a long.

The trading volume is also often taken into account. If the trading volume of assets with declining prices is high relative to the trading volume of assets with increasing prices, then this is interpreted as a market with selling pressure. (The basic problem with this interpretation is that there is a buyer for every seller, so we could equally well interpret the situations as if there is a buying pressure.)

For some simple evidence on the profitability of such trading rules, see Figure 10.4.

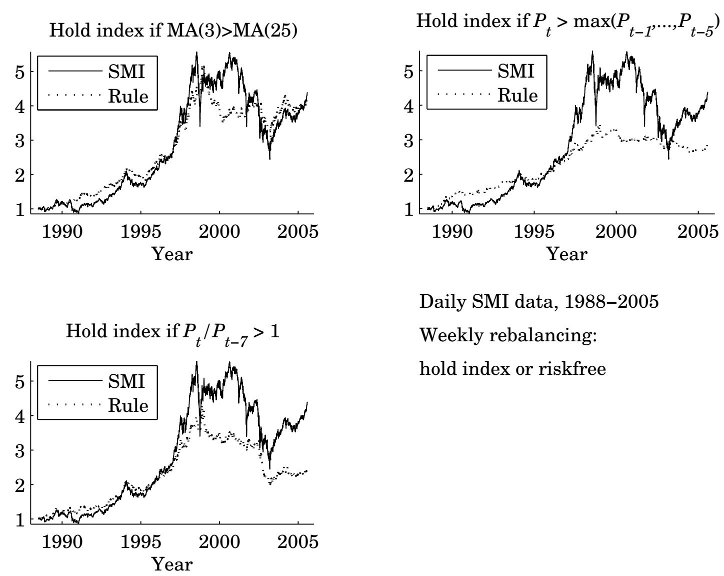


Figure 10.4: Examples of trading rules applied to SMI. The rule portfolios are rebalanced every Wednesday: if condition (see figure titles) is satisfied, then the index is held for the next week, otherwise a government bill is held. The figures plot the portfolio values.

10.3.3 Technical Analysis and Mean Reversion

If we instead believe in mean reversion of the prices, then we can essentially reverse the previous trading rules: we would typically sell when the price is high.

Some investors argue that markets show periods of mean reversion and then periods with trends—an that both can be exploited. Clearly, the concept of a support and resistance levels (or more generally, a channel) is based on mean reversion between these points. A new trend is then supposed to be initiated when the price breaks out of this band.

10.4 Fundamental Analysis

Main reference: Bodie, Kane, and Marcus (2002) 17–18 (stock returns and business cycle), Copeland, Koller, and Murrin (2000) (cash flow models)
Further reading: Kettell (2001) and papers cited in the text

10.4.1 Present Value of Future Dividends

Fundamental analysis is about using information on earnings, interest rates, and risk factors to assess the "fundamental"stock price. If this is higher than the current price, then it may be worthwhile to buy the stock.

The fundamental stock price is the present value of all expected future dividends. The discounting is made by a risk-adjusted "interest rate," which typically is higher than the riskfree rate: it corresponds to the expected return on the stock. If this discount rate is a constant R, then the fundamental price is

$$P_t = \sum_{s=1}^{\infty} \frac{\mathrm{E}_t\, D_{t+s}}{(1+R)^s}. \tag{10.3}$$

The expectations are formed using whatever information we have in t. (If we assume that dividends will grow at a constant rate, then we get the Gordon model and (10.3) can be simplified further.)[4]

[4]To derive (10.3), notice that the pricce can be written as the discounted value of next period's price plus dividends, $P_t = (\mathrm{E}_t\, D_{t+1} + \mathrm{E}_t\, P_{t+1})/(1+R)$. Substitute for P_{t+1} by using $P_{t+1} = (\mathrm{E}_{t+1}\, D_{t+2} + \mathrm{E}_{t+1}\, P_{t+2})/(1+R)$. Repeat this and use the law of iterated expectations, for instance, $\mathrm{E}_t\, \mathrm{E}_{t+1}\, D_{t+2} = \mathrm{E}_t\, D_{t+2}$.

Equation (10.3) is written in terms of the dividends (abstracting from buy-backs and other things which work through decreasing the number of shares). Of course, current dividends are typically very smooth and do not necessarily reflect the outlook for the firm—see *Figure 10.5* for US data. In addition, the dividend measure does not take into account that not all or even any available cash flows have to be paid out as dividends.

However, there exist alternative methods that also make use of the technique of discounting cash flows (see, for instance,). A widespread method in practice is the enterprise DCF model (DCF stands for "discounted cash flow") which is used for calculating the value of a whole company. It uses a broad definition of cash flow, the free cash flow (total after-tax cash flow generated by a company's operations that is available to all providers of the company's capital), which can be seen as a company's true operating cash flow. This free cash flow is discounted by an appropriate discount rate to obtain the value of operations of the company. The sum of the value of operations and of the value of nonoperating assets (whose cash flows were excluded from free cash flow) of the firm yields the value of the total enterprise from which the value of equity can be deduced. Since a sound understanding of the company's past performance provides an essential perspective for developing and evaluating forecasts of future performance, an analysis of historical performance is normally the first step in the valuation process. Here, it is very important to transform the accounting numbers into estimates of the economic performance of a company, also keeping in mind that accounting numbers like earnings and revenues can be influenced by political management decisions, as seen in recent years. (Please note that a serious discussion of the model with all its conveniences and drawbacks is not feasible in the course.)

Whenever the DCF technique is applied it is important to use consistent ingredients, i.e., if you choose a certain cash flow definition be sure to use the according growth rates for the cash flows and the according discount rate. For reasons of simplicity, let us now return to the method of discounting dividends introduced in this section above.

Fundamental analysis can then be interpreted as an attempt to calculate the right hand side of (10.3), that is, to assess the fundamental value of the stock. Factors that are likely to affect the future path of profits (and eventually dividends) are often analyzed at three levels: the macro economic situation, the sector outlooks, and firm specific aspects.

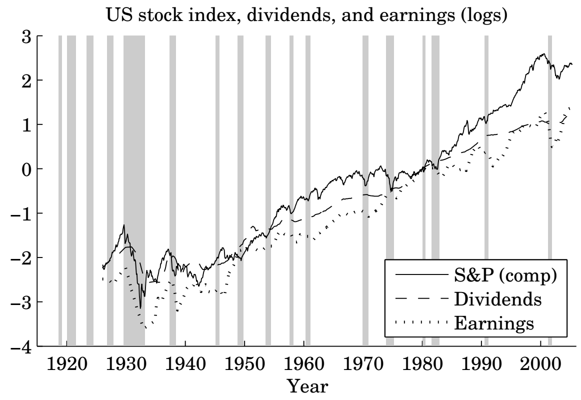


Figure 10.5: US stock index (S&P), dividends, earnings, and NBER recessions (shaded)

10.4.2 The Effect of News

It is clear from (10.3) that the asset price will change when there are changes in the discount rate and the expectations about future dividends. To highlight this, consider a very stylized case where all dividends except in $t+2$ are zero. The asset price in t is then

$$P_t = \frac{\mathrm{E}_t D_{t+2}}{(1+R)^2}. \tag{10.4}$$

In period $t+1$, the expectations about the dividend might be different, and the discount rate might be different (and is denoted R')

$$P_{t+1} = \frac{\mathrm{E}_{t+1} D_{t+2}}{1+R'}. \tag{10.5}$$

The return on holding this asset from t to $t+1$ is the capital gain since there is no dividend in $t+1$

$$\frac{P_{t+1}}{P_t} = (1+R)\frac{\mathrm{E}_{t+1} D_{t+2}}{\mathrm{E}_t D_{t+2}}\frac{1+R}{1+R'}. \tag{10.6}$$

This realized return depends on several factors, which we will discuss one at a time. First, the capital gain will depend on the *discount rate* (the first term in (10.6)): if there are no news in $t+1$, then the capital gain will equal the discount rate (risk free rate plus any risk premia).

Second, if there are *news about (future) dividends* (second term in (10.6)), this will affect the actual return already when the news arrive: news of higher dividends increases the return. It is important to remember that news is a surprise—as compared to what the market expected. Journalists (of all people) often fail to understand this definition of news when they write things like "...inexplicability, the stock market reacted negatively to the 10% earnings growth...."

Third, *news about expected (required) returns* (third term in (10.6)) also affects the actual returns. For instance, a decrease in required returns means that today's actual return is high ($R' < R$ so the last term in (10.6) is larger than unity). The intuition is that future dividends are discounted less than previously thought necessary, so the stream of future dividends is worth more. This could be due to, for instance, a surprise decrease in the nominal interest rate (for instance, because a monetary policy move) or to a decrease in the risk premium that investors require from stocks (for instance, because lower default risk as the business cycle improves).

Quite often new information affects both the expected dividends (earnings) and the discount rate. Sometimes they go in the same direction. For instance, a surprise cut in the monetary policy (interest) rate is likely to increase expected earnings and to decrease the discounting—which both tend to drive up stock prices and thereby create high realized returns in the period of the interest rate cut. (For a dramatic example, see Figure 9-5 in Siegel (1998) where FTSE-100 jumped after the UK left ERM in September 1992 and lowered interest rates.)

In other cases the two factors go in different directions. For instance, the market reaction to the strong US employment report on 5 July 1996 ("payroll up 239,000, unemployment at six-year low at 5.3%, wages up 9 cents an hour, biggest increase in 30 years") was an immediate 1.5% drop (see Figure 14-1 in Siegel (1998)). The reason is that although this was good news for earnings, it also made it much more likely that the Fed would raise interest rates to cool off any signs of inflation.

10.4.3 Stock Returns and the Business Cycle

A top-down forecast starts with an analysis of the business cycle conditions, adds industry specific factors, and works down to the individual firm (stock). It is pretty clear that stock prices react very quickly to signs of business cycle down-turns. In fact, stock returns

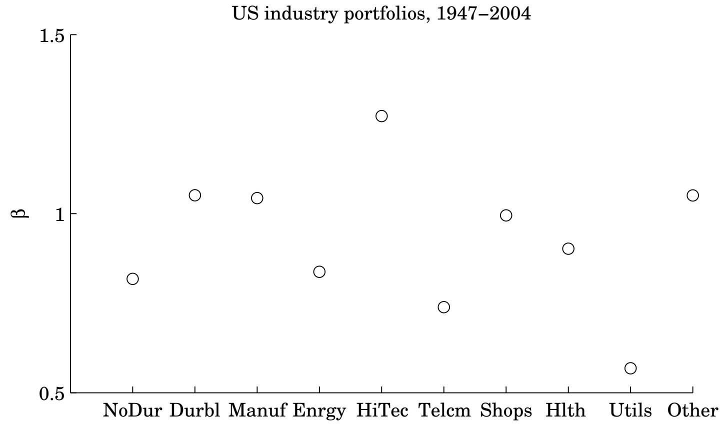


Figure 10.6: βs of US industry indices

typically lead the business cycle—see Figure 10.5. It is also clear that some industries are more cyclical than others. For instance, building companies, investment goods and car manufacturers are typically often very procyclical, whereas food and drugs are not. See 10.6 for an illustration.

However, far from all big movements in stock markets can be explained in terms of macro variables. There is a large number of jumps which seems hard to make sense of—at least if we refuse to believe in stock market bubbles. (See, for instance, Tables 13-1A and 13-1B Siegel (1998), for a listing of really large jumps on the US stock market and a discussion of what could possibly have caused them.)

10.4.4 Market Expectations versus Your Own Expectations*

Not all investors have the same beliefs—especially not about future economic conditions. There is plenty of evidence that analysts and various forecasting agencies have different opinions. How should we then interpret the expectations in the "fundamental price" (10.3)? The short answer is that the expectations in those equations reflect some kind of average expectations: the "market expectations." Consensus expectations (that is, average expectations as measured by surveys) is often used as a proxy for these market expectations.

Consider an agent i who does not share the market expectations . The "correct" price according to this agent is calculated from (10.3), but using his own expectations. Suppose that this agent actually has better information. Would that help him to trade profitably? Yes, but only when (if?) the market eventually admits that he had better information.

This highlights that the most important thing for a profitable trading may not be to make the best forecast of the fundamental value of the asset, but to make the best forecast of the future market sentiments (about the fundamental value). The idea of rational expectations (a key ingredient in the efficient market hypothesis) is that we cannot tell how we will revise the expectations in the future, but who knows if the market really is all that rational?

10.5 Security Analysts

Main reference: Makridakis, Wheelwright, and Hyndman (1998) 10.1
Further reading: papers cited in the text

10.5.1 Evidence on Analysts' Performance

Makridakis, Wheelwright, and Hyndman (1998) 10.1 shows that there is little evidence that the average stock analyst beats (on average) the market (a passive index portfolio). In fact, less than half of the analysts beat the market. However, there are analysts which seem to outperform the market for some time, but the autocorrelation in overperformance is weak. The evidence from mutual funds is similar. For them it is typically also found that their portfolio weights do not anticipate price movements.

It should be remembered that many analysts also are sales persons: either of a stock (for instance, since the bank is underwriting an offering) or of trading services. It could well be that their objective function is quite different from minimizing the squared forecast errors—or whatever we typically use in order to evaluate their performance. (The number of litigations in the US after the technology boom/bust should serve as a strong reminder of this.)

10.5.2 Do Security Analysts Overreact?

The paper by Bondt and Thaler (1990) compares the (semi-annual) forecasts (one- and two-year time horizons) with actual changes in earnings per share (1976-1984) for several hundred companies. The study is done by running regressions like

$$\text{Actual change} = \alpha + \beta(\text{forecasted change}),$$

and to study the estimates of the α and β coefficients. With rational expectations (and a long enough sample), we should have $\alpha = 0$ (no constant bias in forecasts) and $\beta = 1$ (proportionality, for instance no exaggeration).

The main findings are as follows. The main result is that $0 < \beta < 1$, so that the forecasted change tends to be too wild in a systematic way: a forecasted change of 1% is (on average) followed by less than 1% actual change in the same direction. This means that analysts in this sample tended to be too extreme—to exaggerate both positive and negative news.

10.5.3 High-Frequency Trading Based on Recommendations from Stock Analysts

Barber, Lehavy, McNichols, and Trueman (2001) give a somewhat different picture. They focus on the profitability of a trading strategy based on analyst's recommendations. They use a huge data set (some 360,000 recommendation, US stocks) for the period 1985-1996. They sort stocks in to five portfolios depending on the consensus (average) recommendation—and redo the sorting every day (if a new recommendation is published). They find that such a daily trading strategy gives a annual 4% abnormal return on the portfolio of the most highly recommended stocks, and an annual -5% abnormal return on the least favourably recommended stocks.

This strategy requires a lot of trading (a turnover of 400% annually), so trading costs would typically reduce the abnormal return on the best portfolio to almost zero. A less frequent rebalancing (weekly, monthly) gives a very small abnormal return for the best stocks, but still a negative abnormal return for the worst stocks. Chance and Hemler (2001) obtain similar results when studying the investment advise by 30 professional "market timers."

10.5.4 The Characteristics of Individual Analysts' Forecasts in Europe

Bolliger (2001) studies the forecast accuracy (earnings per share) of European (13 countries) analysts for the period 1988–1999. In all, some 100,000 forecasts are studied. It is found that the forecast accuracy is positively related to how many times an analyst has forecasted that firm and also (surprisingly) to how many firms he/she produces forecasts for. The accuracy is negatively related to the number of countries an analyst forecasts and also to the size of the brokerage house he/she works for.

10.5.5 Bond Rating Agencies versus Stock Analysts

Ederington and Goh (1998) use data on all corporate bond rating changes by Moody's between 1984 and 1990 and the corresponding earnings forecasts (by various stock analysts).

The idea of the paper by Ederington and Goh (1998) is to see if bond ratings drive earnings forecasts (or vice versa), and if they affect stock returns (prices).

1. To see if stock returns are affected by rating changes, they first construct a "normal" return by a market model:

   $$\text{normal stock return}_t = \alpha + \beta \times \text{return on stock index}_t,$$

   where α and β are estimated on a normal time period (not including the rating change). The abnormal return is then calculated as the actual return minus the normal return. They then study how such abnormal returns behave, on average, around the dates of rating changes. Note that "time" is then measured, individually for each stock, as a distance from the day of rating change. The result is that there are significant negative abnormal returns following downgrades, but zero abnormal returns following upgrades.

2. They next turn to the question of whether bond ratings drive earnings forecasts or vice versa. To do that they first note that there are some predictable patterns in revisions of earnings forecasts. They therefore fit a simple autoregressive model of earnings forecasts, and construct a variable of earnings forecast revisions (surprises) from the model. They then relate this surprise variable to the bond ratings. In short, the results are the following:

   (a) both earnings forecasts and ratings react to the same information, but there is also a direct effect of rating changes, which differs between downgrades and upgrades.

   (b) downgrades: the ratings have a strong negative direct effect on the earnings forecasts; the returns react even quicker than analysts

   (c) upgrades: the ratings have a small positive direct effect on the earnings forecasts; there is no effect on the returns

A possible reason for why bond ratings could drive earnings forecasts and prices is that bond rating firms typically have access to more inside information about firms than stock analysts and investors.

A possible reason for the observed asymmetric response of returns to ratings is that firms are quite happy to release positive news, but perhaps more reluctant to release bad news. If so, then the information advantage of bond rating firms may be particularly large after bad news. A downgrading would then reveal more new information than an upgrade.

The different reactions of the earning analysts and the returns are hard to reconcile.

10.5.6 International Differences in Analyst Forecast Properties

Ang and Ciccone (2001) study earnings forecasts for many firms in 42 countries over the period 1988 to 1997. Some differences are found across countries: forecasters disagree more and the forecast errors are larger in countries with low GDP growth, less accounting disclosure, and less transparent family ownership structure.

However, the most robust finding is that forecasts for firms with losses are special: forecasters disagree more, are more uncertain, and are more optimistic about such firms.

10.6 Expectations Hypothesis and Forward Prices

It is fairly common to use a forward (or futures) price as a rough forecast of the future asset price. The idea is that speculation should drive the forward price towards the (market) expectation of the future price.

However, this is less imaginative than first thought. The reason is that the *forward-spot parity* (a no-arbitrage relation) shows that the forward price, F_t, does not contain much more information than the current asset price, P_t. To illustrate this, suppose the forward contract expires next period (you can decide the period length...) and that there is no dividends on the asset until then. The forward price must then satisfy

$$F_t = (1 + i_t)P_t, \tag{10.7}$$

where i_t is interest rate. The intuition for this relation is simple: the forward contract is just like buying the asset on credit.

This shows that using the forward price as a forecast is virtually the same as using today's asset price: the random walk (with drift because of the interest rate) assumption.

11 Exchange Rates*

Main reference: Bodie, Kane, and Marcus (2002) 25
Further reading: Sercu and Uppal (1995) 14; Burda and Wyplosz (1997) 19; Cuthbertson (1996) 11-12; Mishkin (1997)

The exchange rate, S_t, is typically defined as units of domestic currency per unit of foreign currency, that is the price (measured in domestic currency) of foreign currency. For instance, if we take Switzerland to be the domestic economy, then we have around 1.5 CHF per EUR.

11.1 What Drives Exchange Rates?

Main reference: The Economist (1997) 11
Further reading: Kettell (2000) 5

11.1.1 Do Real Exchange Rates Determine Nominal Exchange Rates?

The *real exchange rate* is defined as the relative price of foreign goods

$$Q_t = \frac{S_t P_t^*}{P_t} = \frac{\text{“0.15 CHF per SEK} \times \text{6 SEK per Swedish jet fighter”}}{\text{“1 CHF per Swiss jet fighter”}}. \tag{11.1}$$

It is sometimes argued that the real exchange rate is also a key factor for nominal exchange rates. However, this section argues that the causality probably runs in the other direction.

You may note that the *purchasing power parity* issue is about whether Q_t is constant or not. Empirical evidence strongly suggests that there are long-run movements in the

real exchange rates (see, for instance, Burda and Wyplosz (1997) Figure 8.9b)—probably because changes in technology and preferences (things that should change a real price), but probably also medium-term changes due to changes in monetary (exchange rate) policy. The Economist's Big Mac index is an attempt to illustrate the real exchange rate by measuring the price (in a common currency) of a Big Mac in different countries.

The real exchange rate is a real price, and should therefore not depend on the nominal exchange rate—at least not in the long run. However, if nominal prices are sticky, then changes in the nominal interest rates may well have an effect on the real exchange rates.

It is a well established fact that real exchange rates are much more volatile under floating nominal exchange rates than under fixed nominal exchange rates, and that real and nominal exchange rates are strongly correlated. This makes sense if most of the movements in the real exchange rate comes from the nominal exchange rate, that is, if the price ratio in (11.1) is relatively stable. Empirical studies show that this is indeed the case. (See, for instance, Isard (1995) Figures 3.2 and 4.2; Obstfeldt and Rogoff (1996) Figures 9.1-3, Burda and Wyplosz (1997) Figure 19.1). This suggests that, for short to medium horizons, it is the nominal exchange rate that determines the real exchange rate—not the other way around.

Trade in goods is very small compared with the trade in foreign exchange—most transactions on the foreign exchange markets are due to investments, which suggests that "supply and demand" for different financial assets may be an important factor in the exchange rate determination. The UIP is one such approach, at least to the extent that we think that changing returns on bonds drive the exchange rate (the opposite could also be true).

11.1.2 The "Dividend" of Domestic Currency

Recall that the exchange rate is the price of one currency (ultimately, the bills and coins) in terms of another currency. As any other financial asset, the price of currency depends on the expected capital gains and on the "dividends." The "dividends" of a currency can be thought of as the liquidity (or payment) services it provides. The need for liquidity services typically increases with the economic activity and the price level. The price of domestic currency (the inverse of the exchange rate) will, at least in the long run, be an increasing function these variables.

With more and more cross-border equity investments, it could also be the case that changing returns on equity (for instance, due to changes in perceived risk) drive exchange rates (see The Economist (2000) for a short discussion), since trade in equity gives a demand for domestic liquidity services (you typically need cash to pay for the equity).

In the long run, both the real exchange rate and the GDP are more or less independent of the monetary policy. Why? Well, monetary policy is, in the long run, just an issue of how many zeros that should be printed on the bank notes (Japan has many zeros and Canada few zeros, but that has hardly determined their relative income level). One of the safest predictions we can make about the exchange rate is that it will probably follow the relative money supply (price level)—at least in the long run.

Note that these long run results tie in well with the Fisher equation for nominal interest rates: if long run inflation is high, then long run interest rates are typically high (leaving real interest rates unaffected), and the currency depreciates (leaving the real exchange rate unaffected). Since the real prices are unaffected, so are real savings (consumption) and exports/imports. These features are often called *long run neutrality of money*.

11.1.3 Macroeconomic News

Macroeconomic *news* are very important for exchange rate movements for two reasons: they signal current and future demand for liquidity services (money demand) and they are also likely to affect monetary policy In practice, the FX market seems to focus on seasonally adjusted (annualized) growth rates. Among the variables of interest we find the following:

1. Early information: employment report (monthly), earnings (monthly), consumer confidence, purchasing managers index, auto sales
2. The rest: GDP (quarterly), producer price index (monthly), industrial production (monthly), capacity utilization (monthly), CPI (monthly), retail sales, quit rates, government deficits

11.2 Forecasting Exchange Rates

Further reading: see papers cited in the text

Several attempts have been made to build forecasting models of exchange rates—but without much success, except possibly, for very long forecasting horizons (many years).

1. A common result is that the forward premium is a poor predictor of the future depreciation. For an example, see Figure **??**.
2. The macro approach has been to use interest rate differentials and various macro variables (prices, money supply,output, trade balance, and so forth) to forecast future exchange rates. Several authors, for instance, Meese and Rogoff (1983), have shown that this type of equations typically forecasts no better than a simple random walk (that is, assuming no expected change), at least not for short to medium run horizons (one to 12 months). This holds even when we use the actual values of the future macro series instead of the expected values. (Faust, Rogers, and Wright (2002) report somewhat better results when preliminary data is used—in an attempt to emulate the information available to the financial market.) The basic reason is probably that exchange rates are inherently forward looking and therefore contain a lot more information (for instance, about future monetary policy) than current macro data. On a more practical level, exchange rates are typically much more volatile than macro variables, and not very correlated with them. It would therefore be something of a surprise if macro data could forecast short run changes in exchange rates.
3. Macro variables are better at "explaining" (if not forecasting) long-run exchange rate depreciation. Countries with high inflation, rapidly expanding money supply, and weak current accounts typically experience exchange rate depreciation.
4. Macro and political variables are sometimes reasonably good at explaining the interest rate differential (expected depreciation under UIP), in particular, under fixed exchange rate regimes with credibility problems (see Lindberg, Söderlind, and Svensson (1993) for an example).
5. Several studies show that professional forecasters do not, on average, predict exchange rates better than simple econometric models, at least not when evaluated in terms of the mean squared error (MSE). There are some indications that they are better at predicting the direction of change, however. Good forecasting performance does not seem to last: few, if any, forecasters are able to outperform simple econometric models for a long period. (See, for instance, Sercu and Uppal (1995) 15.)

6. Surveys of foreign exchange traders (see, for instance, Cheung and Chinn (1999) and Cheung, Chinn, and Marsh (2000)) show several interesting things.

   (a) News about macroeconomic variables is very rapidly incorporated in rates (often within a minute or less).

   (b) The effects of macroeconomic announcements shift over time: the focus moves from one variable to another.

   (c) Fundamentals (including PPP) are typically thought to have very little importance for intraday price movements, a fairly high importance for medium run movements (up to six months) and very large importance for longer movements. In the short to medium run, "speculation," "overreaction to news," and "bandwagon" effects are thought to be important.

   (d) When asked to describe their trading method, the answers are fairly evenly distributed among the following four categories: technical trading, customer order, fundamentals, jobbing (continuous and small "speculation").

7. Taylor (1994) discusses "channel rules" (see Technical Analysis) for foreign exchange futures and argues that an appropriate channel rule can mimic the rule from a time series model (similar to an AR model) and therefore exploit this type of predictability of asset prices. An application to a sample exchange rates of the late 1980s suggest that the rule may generate profits.

12 Interest Rates*

12.1 Interest Rate Analysts

Further reading: papers cited in the text

12.1.1 Interest Rate Forecasts by Analysts

Kolb and Stekler (1996) use a semi-annual survey of (12 to 40) professional analysts' interest rate forecasts published in Wall Street Journal. The (6 months ahead) forecasts are for the 6-month T-bill rate and the yield on 30-year government bonds. The paper studies four questions, and I summarize the findings below.

1. Q. Is the distribution of the forecasts (across forecasters) at any point in time symmetric? (Analyzed by first testing if the sample distribution could be drawn from a normal distribution; if not, then checking asymmetry (skewness).) A. Yes, in most periods. (The authors argues why this makes the median forecast a meaningful representation of a "consensus forecast.")

2. Q. Are all forecasters equally good (in terms of ranking of (absolute?) forecast error)? A. Yes for the 90-day T-bill rate; No for the long bond yield.

3. Q. Are some forecasters systematically better (in terms of absolute forecast error)? (Analyzed by checking if the absolute forecast error is below the median more than 50% of the time) A. Yes.

4. Q. Do the forecasts predict the direction of change of the interest rate? (Analyzed by checking if the forecast gets the sign of the change right more than 50% of the time.) A. No.

12.1.2 Market Positions as Interest Rate Forecasts

Hartzmark (1991) has data on daily futures positions of large traders on eight different markets, including futures on 90-day T-bills and on government bonds. He uses this data to see if the traders changed their position in the right direction compared to realized prices (in the future) and if they did so consistently over time.

The results indicate that these large investors in T-bills and bond futures did no better than an uninformed guess of the direction of change of the bill and bond prices. He get essentially the same results if the size of the change in the position and in the price are also taken into account.

There is of course a distribution of how well the different investors do, but it looks much like one generated from random guesses (uninformed forecasts). The investors change places in this distribution over time: there is very little evidence that successful investors continue to be successful over long periods.

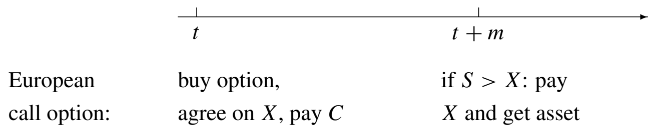


Figure 13.1: Timing convention of option contract

13 Options

Further reading: Bodie, Kane, and Marcus (2002) 20–21; Bahra (1996); McCauley and Melick (1996b); McCauley and Melick (1996a); Söderlind and Svensson (1997) (academic)

This section discusses how option prices can be used to gauge market beliefs and uncertainty.

13.1 Risk Neutral Pricing of a European Call Option

A European call option contract traded in t may stipulate that the buyer of the contract has the right (not the obligation) to buy one unit of the underlying asset (from the issuer of the option) in $t+m$ at the strike price X. If this is a pure option, then the option price, $\tilde{C}_t$, is actually not paid until $t+m$. In contrast, a standard option requires payment of the option price at t. See *Figure 13.1* for the timing convention.

Suppose investors are risk neutral. The price of a pure call option is then

$$\tilde{C}_t = \mathrm{E}_t \max\left(0, S_{t+m} - X\right), \tag{13.1}$$

where S_{t+m} is the asset price, and X and the strike price. Of course, investors are not risk neutral, but we will use this as a convenient simplification.

Example 17 *Suppose S_{t+m} only can take three values: 90, 100, and 110; and that the probabilities for these events are: 0.5, 0.4, and 0.1, respectively. We consider three European call option contracts with the strike prices 89, 99, and 109. From (13.1) their prices are*

$$\tilde{C}\,(X = 89) = 0.5(90-89) + 0.4(100-89) + 0.1(110-89) = 7$$
$$\tilde{C}\,(X = 99) = 0.5 \times 0 + 0.4(100-99) + 0.1(110-99) = 1.5$$
$$\tilde{C}\,(X = 109) = 0.5 \times 0 + 0.4 \times 0 + 0.1(110-109) = 0.1.$$

13.2 Black-Scholes

In the Black-Scholes model, we assume that the logarithm of the future asset price is normally distributed with mean $\bar{s}$ and variance σ_{ss}

$$\ln S_{t+m} \text{ is distributed as } N\left(\bar{s}, \sigma_{ss}\right). \tag{13.2}$$

This distribution could vary over time (even though no time subscripts have been put on it—to minimize notational clutter). Note that the assumption in the Black-Scholes model, that $\ln S_t$ is a random walk, implies (13.2).

Calculating the expectation in (13.1) and manipulating the results a bit gives the Black-Scholes formula.[1] (We can arrive at the same formula by using no-arbitrage arguments instead—which do not assume risk neutrality.)

13.3 Implied Volatility: A Measure of Market Uncertainty

The Black-Scholes formula contains only one unknown parameter: the variance σ_{ss} in the distribution of $\ln S_{t+m}$ (see (13.2)). With data on the option price, spot price, interest rate, and strike price, we can solve for the variance. The term $\sqrt{\sigma_{ss}}$ is often called the *implied volatility*. Often this is expressed as standard deviation per unit of time until expiry, σ, which obeys $\sigma\sqrt{m} = \sqrt{\sigma_{ss}}$. Note that we can solve for one implied volatility for each

[1] According to this formula, the price of a European call option with strike price X is

$$C_t = S_t \Phi\left(\frac{m \ln\left[1 + Y_t(m)\right] + \ln\left(S_t/X\right)}{\sqrt{\sigma_{ss}}}\right) - \left[1 + Y_t(m)\right]^{-m} X \Phi\left(\frac{m \ln\left[1 + Y_t(m)\right] + \ln\left(S_t/X\right) - \sigma_{ss}/2}{\sqrt{\sigma_{ss}}}\right),$$

where $\Phi(z)$ is the probability of a value lower than z according to a standard normal distribution.

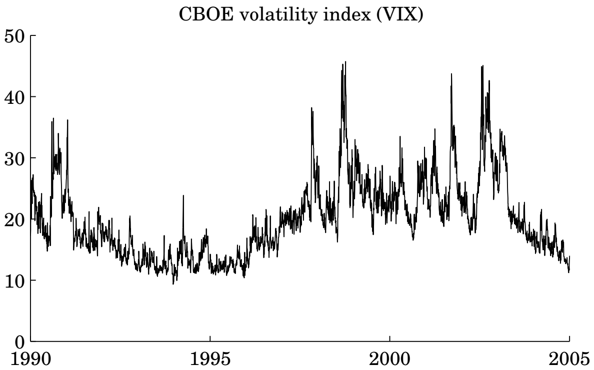


Figure 13.2: CBOE VIX, summary measure of implied volatilities (30 days) on US stock markets

available strike price—which can be used as an indicator of market uncertainty about the future asset price, S_{t+m}. See Figure 13.2 for an example.

If the Black-Scholes formula is correct, that is, if the assumption in (13.2) is correct, then these volatilities should be the same across strike prices. It is often found that the implied volatility is a U-shaped function of the strike price. One possible explanation is that the (perceived) distribution of the future asset price has relatively more probability mass in the tails ("fat tails") than a normal distribution has.

13.4 Subjective Distribution: The Shape of Market Beliefs

In (13.2) we assumed that the distribution of the future log asset price is normal, which is the same assumption as in the Black-Scholes model. However, we could very well assume some other distribution and then use option prices to estimate its form by choosing the parameters in the distribution to minimize, say, the sum (across strike prices) of squared differences between observed and predicted prices. (This is like the minimization problem behind the least squares method in econometrics.) This allows the possibility to pick up skewed (downside risk different from upside risk?) and even bi-modal distributions.

Example 18 *Suppose we observe the option prices in Example 17, and want to use these*

to recover the probabilities. We know the possible states, but not their probabilities. Let $\Pr(x)$ *denote the probability that* $S_{t+m} = x$*. From Example 17, we have that the option price for* $X = 109$ *equals*

$$\begin{aligned}\tilde{C}\,(X = 109) &= 0.1\\ &= \Pr(90) \times 0 + \Pr(100) \times 0 + \Pr(110)(110 - 109),\end{aligned}$$

which we can solve as $\Pr(110) = 0.1$*. We now use this in the expression for the option price for* $X = 99$

$$\begin{aligned}\tilde{C}\,(X = 99) &= 1.5\\ &= \Pr(90) \times 0 + \Pr(100)(100 - 99) + 0.1(110 - 99),\end{aligned}$$

which we can solve as $\Pr(100) = 0.4$*. Since probabilities sum to one, it follows that* $\Pr(90) = 0.5$.

It is important to note that the distribution we can estimate directly from option prices are risk neutral distributions, which are different from the true distributions if there is a risk premium, but it turns out that the Black-Scholes formula holds even if there are risk premia. Therefore, if the underlying asset price, S_t, has a risk premium then this is automatically incorporated into the option price (in a non-linear way). It is straightforward to show that the effect of the risk premium in this case is to shift the mean of the normal distribution. Similar, but potentially more complicated, things happen in other types of distributions. A meaningful interpretation of a shift in the estimated distribution there requires an estimate (of some sort) of the risk premium. It is common to assume that it can be disregarded.

Figure 13.3 shows some data and results for German bond options around the announcement of the very high money growth rate on 2 march 1994.

Bibliography

Ang, J. S., and S. J. Ciccone, 2001, "International Differences in Analyst Forecast Properties," mimeo, Florida State University.

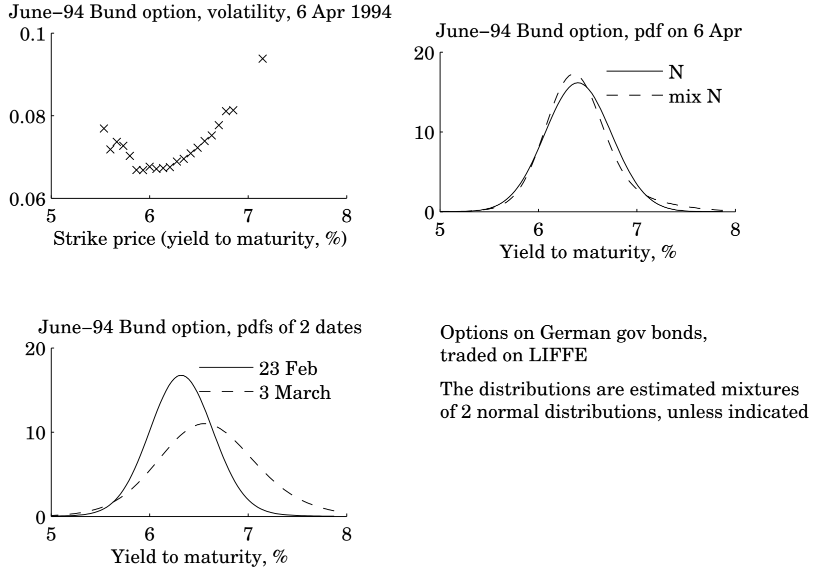


Figure 13.3: Bund options 23 February and 3 March 1994. Options expiring in June 1994.

Bahra, B., 1996, "Probability Distributions of Future Asset Prices Implied by Option Prices," *Bank of England Quarterly Bulletin*, 36, August 1996, 299–311.

Barber, B., R. Lehavy, M. McNichols, and B. Trueman, 2001, "Can Investors Profit from the Prophets? Security Analyst Recommendations and Stock Returns," *Journal of Finance*, 56, 531–563.

Bodie, Z., A. Kane, and A. J. Marcus, 2002, *Investments*, McGraw-Hill/Irwin, Boston, 5th edn.

Bolliger, G., 2001, "The Characteristics of Individual Analysts' Forecasts in Europe," mimeo, University of Neuchatel.

Bondt, W. F. M. D., and R. H. Thaler, 1990, "Do Security Analysts Overreact?," *American Economic Review*, 80, 52–57.

Brock, W., J. Lakonishok, and B. LeBaron, 1992, "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns," *Journal of Finance*, 47, 1731–1764.

Burda, M., and C. Wyplosz, 1997, *Macroeconomics - A European Text*, Oxford University Press, 2nd edn.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, New Jersey.

Chance, D. M., and M. L. Hemler, 2001, "The Performance of Professional Market Timers: Daily Evidence from Executed Strategies," *Journal of Financial Economics*, 62, 377–411.

Cheung, Y.-W., and M. D. Chinn, 1999, "Macroeconomic Implications of the Beliefs and Behavior of Foreign Exchange Traders," NBER Working Paper 7417.

Cheung, Y.-W., M. D. Chinn, and I. W. Marsh, 2000, "How Do UK-Based Foreign Exchange Dealers Think Their Market Operates," NBER Working Paper 7524.

Copeland, T., T. Koller, and J. Murrin, 2000, *Valuation: Measuring and Managing the Value of Companies*, Wiley John & Sons Ltd, Ney York, 3rd edn.

Cuthbertson, K., 1996, *Quantitative Financial Economics*, Wiley, Chichester, England.

Dunis, C., 1996, *Forecasting Financial Markets*, Wiley, Chichester.

Ederington, L. H., and J. C. Goh, 1998, "Bond Rating Agencies and Stock Analysts: Who Knows What When?," *Journal of Financial and Quantitative Analysis*, 33.

Faust, J., J. H. Rogers, and J. H. Wright, 2002, "Exchange Rate Forecasting: The Errors We've Really Made," International Finance Discussion Paper 714, Board of Governors of the Federal Reserve System.

Granger, C. W. J., 1992, "Forecasting Stock Market Prices: Lessons for Forecasters," *International Journal of Forecasting*, 8, 3–13.

Hartzmark, M. L., 1991, "Luck versus Forecast Ability: Determinants of Trader Performance in Futures Markets," *Journal of Business*, 64, 49–74.

Isard, P., 1995, *Exchange Rate Economics*, Cambridge University Press.

Kettell, B., 2000, *What Drives Currency Markets: Making Sense of Market Information*, Financial Times/Prentice Hall.

Kettell, B., 2001, *Economics for Financial Markets: Making Sense of Information in Financial Markets*, Financial Times/Prentice Hall.

Kolb, R. A., and H. O. Stekler, 1996, "How Well Do Analysts Forecast Interest Rates," *Journal of Forecasting*, 15, 385–394.

Lindberg, H., P. Söderlind, and L. E. Svensson, 1993, "Devaluation Expectations: The Swedish Krona 1985-1992," *The Economic Journal*, pp. 1180–1189.

Lo, A. W., H. Mamaysky, and J. Wang, 2000, "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation," *Journal of Finance*, 55, 1705–1765.

Makridakis, S., S. C. Wheelwright, and R. J. Hyndman, 1998, *Forecasting: Methods and Applications*, Wiley, New York, 3rd edn.

McCauley, R., and W. Melick, 1996a, "Propensity and Density," *Risk*, 9, 52–54.

McCauley, R., and W. Melick, 1996b, "Risk Reversal Risk," *Risk*, 9, 54–58.

Meese, R. A., and K. Rogoff, 1983, "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample," *Journal of International Economics*, 14, 3–24.

Mishkin, F. S., 1997, *The Economics of Money, Banking, and Financial Markets*, Addison-Wesley, Reading, Massachusetts, 5th edn.

Murphy, J. J., 1999, *Technical Analysis of the Financial Markets*, New York Institute of Finance.

Neely, C. J., 1997, "Technical Analysis in the Foreign Exchange Market: A Layman's Guide," *Federal Reserve Bank of St. Louis Review*.

Obstfeldt, M., and K. Rogoff, 1996, *Foundations of International Macroeconomics*, MIT Press.

Sercu, P., and R. Uppal, 1995, *International Financial Markets and the Firm*, South-Western College Publishing, Cinicinnati, Ohio.

Siegel, J. J., 1998, *Stocks for the Long Run*, McGraw-Hill, 2nd edn.

Söderlind, P., and L. E. O. Svensson, 1997, "New Techniques to Extract Market Expectations from Financial Instruments," *Journal of Monetary Economics*, 40, 383–429.

Taylor, S. J., 1994, "Trading Futures Using a Channel Rule: A Study of the Predictive Power of Technical Analysis with Currency Examples," *Journal of Futures Markets*, 14, 215–235.

The Economist, 1993, "Frontiers of Finance," pp. 5–20.

The Economist, 1997, *Guide to Economic Indicators*, John Wiley and Sons, New York.

The Economist, 2000, "Test-Driving a New Model," pp. 81–82.