Trend Models

• A trend model is

\[ T_t = g(\text{Time}_t) \]

where \( \text{Time}_t \) is the time index.

• In STATA, \( \text{Time}_t \) is an integer sequence, normalized to be zero at first observation of 1960.

• Most common models
  – Linear Trend
  – Exponential Trend
  – Quadratic Trend
  – Trends with Changing Slope
Warning: Be skeptical of Trend Models

- While in some cases, trend forecasting can be useful.
- In many cases, it can be hazardous.
- We will examine some of the trend examples in Chapter 5 of Diebold’s text.
- They did not forecast well out of sample.
- A constructive alternative is to forecast growth rates, as we did for consumption expenditure.
Example 1
Labor Force Participation Rate

• From BLS
• Monthly, 1948-2009, Seasonally adjusted
• Men and Women, ages 25+
• Percentage of population in labor force (employed plus unemployed divided by population)
• Diebold estimates on 1948-1992
• We will estimate on 1948-1992, forecast 1993-2009
Women’s Labor Participation Rate
1948-1992
Men’s Labor Participation Rate
1948-1992

[Graph showing a declining trend in men's labor participation rate from 1948 to 1992.]
Linear Trend Model

- The labor force participation rates have been smoothly and linearly increasing (for women) and smoothly and linearly decreasing (for men) over 1948-1992.
- This suggests a linear trend

\[ T_t = \beta_0 + \beta_1 Time_t \]

- In this model, \( \beta_1 \) is the expected period-to-period change in the trend \( T_t \).
Example 2
Retail Sales, Current Dollars

- From Census Bureau
  - This particular series discontinued after 2001
- Diebold estimates up to 1991
- We will forecast 1992-current
Retail Sales
1955-1993
Quadratic Trends

• The retail sales series has been increasing smoothly over 1955-1993, but not linearly.
• To model this we will use a quadratic trend

\[ T_t = \beta_0 + \beta_1 Time_t + \beta_2 Time_t^2 \]
Example 3
Transaction Volume, S&P Index

• From Yahoo Finance
• (Similar to NYSE series in Diebold)
• Weekly, 1950-current
• Diebold estimates on 1955-1993, forecasts 1994
• We will forecast 1994-2001
Transaction Volume
Exponential Trend

- To model this we will use an exponential trend

\[ T_t = e^{\beta_0 + \beta_1 Time_t} \]

- The exponential trend is linear after taking (natural) logarithms

\[ \ln(T_t) = \beta_0 + \beta_1 Time_t \]

- This is typically estimated by a linear model after taking logs of the variable to forecast
• In logarithms, trend is roughly linear.
Exponential Trends

• Most economic series which are growing (aggregate output, such as GDP, investment, consumption) are exponentially increasing
  – Percentage changes are stable in the long run
• These series cannot be fit by a linear trend
• We can fit a linear trend to their (natural) logarithm
Linear Models

• The linear and quadratic trends are both linear regression models of the form

$$\mu_t = \beta_0 + \beta_1 x_{1t}$$

or

$$\mu_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t}$$

where

- $x_{1t} = Time_t$
- $x_{2t} = Time_t^2$
Example 4
Real GDP

- From BEA
- Quarterly, 1947-2009
- We will estimate on 1947-1990, forecast 1991-2009
- Also use an exponential trend
Real GDP
Ln(Real GDP)
Linear Forecasting

- The goal is to forecast future observations given a linear function of observables.
- In the case of trend estimation, these observables are functions of the time index.
- In other cases, they will be other functions of the data.
- In the model \( \mu_t = \beta_0 + \beta_1 x_t \),
  the forecast for \( y_{t+h} \) is \( \hat{y}_{t+h} = b_0 + b_1 x_t \) where \( b_0 \) and \( b_1 \) are estimates.
Estimation

• How should we select $b_0$ and $b_1$?

• The goal is to produce a forecast with low mean square error (MSE)

• The best linear forecast is the linear function $\beta_0 + \beta_1 x_t$ that minimizes the MSE

\[
E(y_{t+h} - \hat{y}_{t+h})^2 = E(y_{t+h} - \beta_0 - \beta_1 x_t)^2
\]

• We do not know the MSE, but we can estimate it by a sample average
Sum of Squared Errors

• Sample estimate of mean square error is the sum of squared errors

\[ S_n(\beta_0, \beta_1) = \frac{1}{n} \sum_{t=1}^{n} (y_{t+h} - \beta_0 - \beta_1 x_t)^2 \]

• The best linear forecast is the linear function \( \beta_0 + \beta_1 x_t \) that minimizes the MSE, or expected sum of squared errors.

• Our sample estimate of the best linear forecast is the linear function which minimizes the (sample) sum of squared errors.

• This is called the least-squares estimator
Least Squares

- The least-squares estimates \((b_0, b_1)\) are the values which minimize the sum of squared errors

\[
S_n(\beta_0, \beta_1) = \frac{1}{n} \sum_{t=1}^{n} (y_{t+h} - \beta_0 - \beta_1 x_t)^2
\]

- This produces estimates of the best linear predictor – the linear function \(\beta_0 + \beta_1 x_t\) that minimizes the MSE
Multiple Regressors

• There are multiple regressors

\[ \mu_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} \]

• For example, the quadratic trend

\[ T_t = \beta_0 + \beta_1 \text{Time}_t + \beta_2 \text{Time}_t^2 \]

• The best linear predictor is the linear function

\[ \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} \] that minimizes the MSE

\[ E(y_{t+h} - \hat{y}_{t+h})^2 = E(y_{t+h} - \beta_0 - \beta_1 x_{1t} - \beta_2 x_{2t})^2 \]
Multiple Regression

• The sample estimate of the best linear predictor are the values \((b_0, b_1, b_2)\) which minimize the sum of squared errors

\[
S_n(\beta_0, \beta_1, \beta_2) = \frac{1}{n} \sum_{t=1}^{n} (y_{t+h} - \beta_0 - \beta_1 x_{1t} - \beta_2 x_{2t})^2
\]

• In STATA, use the `regress` command
Example 1

Women’s Labor Force Participation Rate
Regression Estimation

. use "C:\Users\Bruce Hansen\Documents\docs\classdocs\390\participation.dta"

. regress women t if t<=tm(1992m12)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 540</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>33879.9082</td>
<td>1</td>
<td>33879.9082</td>
<td>F( 1, 538) = 19575.55</td>
</tr>
<tr>
<td>Residual</td>
<td>931.130559</td>
<td>538</td>
<td>1.73072595</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>34811.0387</td>
<td>539</td>
<td>64.5844874</td>
<td>R-squared = 0.9733</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.9732</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 1.3156</td>
</tr>
</tbody>
</table>

| women   | Coef.     | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|---------|-----------|-----------|-------|-------|---------------------|
| t       | 0.0508126 | 0.0003632 | 139.91| 0.000 | 0.0500992  0.0515261 |
| _cons   | 36.02153  | 0.0726804 | 495.62| 0.000 | 35.87876  36.1643  |
In-Sample Fit
Residuals

- Residuals are difference between data and fitted regression line

\[ \hat{e}_t = y_{t+h} - T_t \]
\[ = y_{t+h} - b_0 - b_1 Time_t \]

`. predict e if t<=tm(1992m12), residuals
(204 missing values generated)`. 
Residual Plot

The Residual Plot shows the residuals of a model over time, with a linear prediction line (red) and a linear prediction (green). The residuals appear to follow a pattern, with some fluctuations around the mean value.
In-Sample Fit

- Compute with **predict** command
- Fit looks good
Forecast

- Forecast is the linear function with estimated coefficients

\[ T_{T+h} = b_0 + b_1 Time_{T+h} \]

- Compute with **predict** command
Forecast Intervals

• Compute residuals

$$\hat{e}_t = y_{t+h} - \hat{\mu}_t$$

$$= y_{t+h} - b_0 - b_1 Time_t$$

• Compute quantiles of residuals
  – These are constant over time

• Add to predicted values
  – Identical to constant mean case
• Out of sample prediction might be too low.
Out-of-Sample
Women’s Labor Force Participation

• No: Prediction was way too high!
Men’s Labor Force Participation Rate
Estimation

```
regress men t if t<=tm(1992m12)
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>9332.01915</td>
<td>1</td>
<td>9332.01915</td>
</tr>
<tr>
<td>Residual</td>
<td>280.157915</td>
<td>538</td>
<td>.520739619</td>
</tr>
<tr>
<td>Total</td>
<td>9612.17706</td>
<td>539</td>
<td>17.8333526</td>
</tr>
</tbody>
</table>

Number of obs = 540
F(1, 538) = 17920.70
Prob > F = 0.0000
R-squared = 0.9709
Adj R-squared = 0.9708
Root MSE = 0.72162

| men | Coef.  | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-----|--------|-----------|-------|-----|----------------------|
| t   | -.0266679 | .0001992 | -133.87 | 0.000 | -.0270592 - .0262766 |
| _cons | 85.59182 | .039867 | 2146.94 | 0.000 | 85.51351 - 85.67013 |
In-Sample Fit
Residuals
Forecast

• End of Sample looks worrying
Out-of-Sample
Men’s Labor Force Participation

• Linear Trend Terrible
Example 2
Retail Sales
## Linear and Quadratic Trend

```
.regress sales t if t<=tm(1993m12)
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 468</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1.0423e+12</td>
<td>1</td>
<td>1.0423e+12</td>
<td>F( 1,   466) = 4167.23</td>
</tr>
<tr>
<td>Residual</td>
<td>1.1655e+11</td>
<td>466</td>
<td>250107680</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>1.1588e+12</td>
<td>467</td>
<td>2.4814e+09</td>
<td>R-squared = 0.8994</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.8992</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 15815</td>
</tr>
</tbody>
</table>

| sales    | Coef.       | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-------------|-----------|-------|-----|---------------------|
| t        | 349.3091    | 5.41111   | 64.55 | 0.000 | 338.6759 - 359.9423 |
| _cons    | 4983.413    | 1189.88   | 4.19  | 0.000 | 2645.218 - 7321.609 |

```
.generate t2=t^2
.regress sales t t2 if t<=tm(1993m12)
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 468</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1.1553e+12</td>
<td>2</td>
<td>5.7763e+11</td>
<td>F( 2,   465) =75689.11</td>
</tr>
<tr>
<td>Residual</td>
<td>3.5487e+09</td>
<td>465</td>
<td>7631593.27</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>1.1588e+12</td>
<td>467</td>
<td>2.4814e+09</td>
<td>R-squared = 0.9969</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.9969</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 2762.5</td>
</tr>
</tbody>
</table>

| sales    | Coef.       | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|----------|-------------|-----------|-------|-----|---------------------|
| t        | 19.01707    | 2.874205  | 6.62  | 0.000 | 13.36903 - 24.66511 |
| t2       | 9518503     | 0.0078223 | 121.68| 0.000 | 9364789 - 9672217   |
| _cons    | 16263.16    | 227.5825  | 71.46 | 0.000 | 15815.94 - 16710.37 |
Linear and Quadratic Trend
Forecast
Residuals
Example 3: Volume

![Graph showing the trend of volume over time, with a significant increase towards the end of the 1980s.](image)
**Estimating Logarithmic Trend**

```stata
.use "C:\Users\Bruce Hansen\Documents\docs\classdocs\390\s&p.dta"

.generate lvolume=ln(volume)

.regress lvolume t if t<tw(1992w1)
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5122.49788</td>
<td></td>
<td>5122.49788</td>
</tr>
<tr>
<td>Residual</td>
<td>199.78817</td>
<td>2190</td>
<td>0.091227475</td>
</tr>
<tr>
<td>Total</td>
<td>5322.28605</td>
<td>2191</td>
<td>2.4291584</td>
</tr>
</tbody>
</table>

- Number of obs = 2192
- F( 1, 2190) = 56150.82
- Prob > F = 0.0000
- R-squared = 0.9625
- Adj R-squared = 0.9624
- Root MSE = 0.30204

| lvolume   | Coef.      | Std. Err. | t     | P>|t| | [95% Conf. Interval] |
|-----------|------------|-----------|-------|-----|----------------------|
| t         | 0.0024242  | 0.0000102 | 236.96| 0.000| 0.0024041 0.0024442 |
| _cons     | 15.05281   | 0.0087069 | 1728.83| 0.000| 15.03574 15.06989  |
Fitted Trend

- Blue line: Ivolume
- Red line: Fitted values
Residuals
Forecast
Out-of-Sample
Forecasting Levels from a Forecast of Logs

• Let $Y_t$ be a series and $y_t = \ln(Y_t)$ its logarithm
• Suppose the forecast for the log is a linear trend:
  $E(y_{t+h} \mid \Omega_t) = T_t = \beta_0 + \beta_1 Time_t$
• Then a forecast for $Y_t$ is $\exp(T_t)$
• If $[L_T, U_T]$ is a forecast interval for $y_{T+h}$
• Then $[\exp(L_T), \exp(U_T)]$ is a forecast interval for $Y_{T+h}$
• In other words, just take your point and interval forecasts, and apply the exponential function.
  – In STATA, use `generate` command
Forecast in Levels
Out-of-Sample
Example 4: Real GDP
Ln(Real GDP)
. regress y t if t<=tq(1990q4)

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>34.6808119</td>
<td>1</td>
<td>34.6808119</td>
</tr>
<tr>
<td>Residual</td>
<td>.257101262</td>
<td>174</td>
<td>.001477593</td>
</tr>
<tr>
<td>Total</td>
<td>34.9379132</td>
<td>175</td>
<td>.199645218</td>
</tr>
</tbody>
</table>

Number of obs = 176
F( 1, 174) = 23471.15
Prob > F = 0.0000
R-squared = 0.9926
Adj R-squared = 0.9926
Root MSE = 0.03844

| y       | Coef.    | Std. Err. | t      | P>|t|  | [95% Conf. Interval] |
|---------|----------|-----------|--------|------|---------------------|
| t       | 0.0087372 | 0.000057  | 153.20 | 0.000| 0.0086247 0.0088498 |
| _cons   | 7.964978  | 0.0035347 | 2253.35| 0.000| 7.958001 7.971954  |
Fitted Trend
Residuals
Forecast of $\ln(R\text{GDP})$
Forecast of RGDP (in levels)
Out-of-Sample

![Graph showing time series data with labels rgdp, Forecast, ep1, and ep2 over time from 1950q1 to 2010q1.]
Problems with Pure Trend Forecasts

- **Trend forecasts understate uncertainty**
- Actual uncertainty increases at long forecast horizons.
- Short-term trend forecasts can be quite poor unless trend lined up correctly.
- Long-term trend forecasts are typically quite poor, as trends change over long time periods.
- It is preferred to work with growth rates, and reconstruct levels from forecasted growth rates (more on this later).
Trend Models

• I hope I’ve convinced you to be skeptical of trend-based forecasting.
• The problem is that there is no economic theory for constant trends, and “changes” in the trend function are not apparent before they occur.
• It is better to forecast growth rates, and build levels from growth.
Final Trend Forecast

World Record – 100 meter sprint