Estimating the Single Index Model

Eric Zivot

August 15, 2012

Estimating the Single Index Model

Sharpe's Single (SI) model:

$$
\begin{aligned}
R_{it} &= \alpha_i + \beta_i R_{Mt} + \varepsilon_{it},\ t = 1, \ldots, T \\
\varepsilon_{it} &\sim \text{iid } N(0, \sigma^2_{\varepsilon,i}),\ R_{M,t} \sim \text{iid } N(\mu_M, \sigma^2_M) \\
\text{cov}(R_{Mt}, \varepsilon_{is}) &= 0 \text{ for } t, s \\
E[R_{it}] &= \mu_i = \alpha_i + \beta_i \mu_M,\ \text{var}(R_{it}) = \beta_i^2 \sigma^2_M + \sigma^2_{\varepsilon,i} \\
\alpha_i &= \mu_i - \beta_i \mu_M \\
\beta_i &= \frac{\text{cov}(R_{it}, R_{Mt})}{\text{var}(R_{Mt})} = \frac{\sigma_{iM}}{\sigma^2_M}
\end{aligned}
$$

Main parameters to estimate: α_i, β_i and $\sigma^2_{\varepsilon,i}$

Plug-in Principle Estimators

Plug-in principle: Estimate model parameters using sample statistics

$$
\begin{aligned}
\hat{\beta}_i &= \frac{\hat{\sigma}_{iM}}{\hat{\sigma}^2_M} \\
\hat{\sigma}_{iM} &= \frac{1}{T-1}\sum_{t=1}^{T}(R_{it}-\hat{\mu}_i)(R_{Mt}-\hat{\mu}_M) \\
\hat{\sigma}^2_M &= \frac{1}{T-1}\sum_{t=1}^{T}(R_{Mt}-\hat{\mu}_M)^2 \\
\hat{\mu}_i &= \frac{1}{T}\sum_{t=1}^{T}R_{it}, \\
\hat{\mu}_M &= \frac{1}{T}\sum_{t=1}^{T}R_{Mt}
\end{aligned}
$$

Plug-in principle estimator for $\alpha_i = \mu_i - \beta_i \mu_M$:

$$\hat{\alpha}_i = \hat{\mu}_i - \hat{\beta}_i \hat{\mu}_M$$

Plug-in principle estimator of ε_{it} :

$$\hat{\varepsilon}_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{Mt}$$

Plug-in principle estimator for $\sigma^2_{\varepsilon,i} = \text{var}(\varepsilon_{it})$:

$$\begin{aligned} \hat{\sigma}^2_{\varepsilon,i} &= \frac{1}{T-2} \sum_{t=1}^{T} \hat{\varepsilon} \\ &= \frac{1}{T-2} \sum_{t=1}^{T} \left(R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{Mt} \right)^2 \end{aligned}$$

Least Squares Estimation of SI Model Parameters

Idea: SI model postulates a linear relationship between R_{it} and R_{Mt} with intercept α_i and slope β_i :

$$R_{it} = \alpha_i + \beta_i R_{Mt} + \varepsilon_{it}$$

Estimate α_i and β_i by finding the "best fitting line" to the scatterplot of data

- Problem: How to define the "best fitting line"?

- Least Squares solution: minimize the sum of squared residuals (errors)

Least Squares Algorithm

$$\begin{aligned}\hat{\alpha}_i &= \text{initial guess for } \alpha_i \\ \hat{\beta}_i &= \text{initial guess for } \beta_i \\ \hat{R}_{it} &= \hat{\alpha}_i + \hat{\beta}_i R_{Mt} = \text{fitted line} \\ \hat{\varepsilon}_{it} &= R_{it} - \hat{R}_{it} \\ &= R_{it} - (\hat{\alpha}_i + \hat{\beta}_i R_{Mt}) = \text{residual}\end{aligned}$$

Determine the best fitting line by minimizing the *Sum of Squared Residuals* (SSR)

$$\begin{aligned}\mathrm{SSR}(\hat{\alpha}_i, \hat{\beta}_i) &= \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 \\ &= \sum_{t=1}^{T} \left(R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{Mt}\right)^2\end{aligned}$$

That is, the least squares estimates solve

$$\min_{\hat{\alpha}_i, \hat{\beta}_i} \mathsf{SSR}(\hat{\alpha}_i, \hat{\beta}_i) = \sum_{t=1}^{T} \left(R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{Mt} \right)^2$$

Note: Because $\mathsf{SSR}(\hat{\alpha}_i, \hat{\beta}_i)$ is a quadratic function in $\hat{\alpha}_i, \hat{\beta}_i$, the first order conditions for a minimum give two linear equations in two unknowns and so there is an analytic solution to the minimization problem that we can find using calculus.

Calculus Solution

The first order conditions for a minimum are

$$
\begin{aligned}
0 &= \frac{\partial \mathsf{SSR}(\hat{\alpha}_i, \hat{\beta}_i)}{\partial \hat{\alpha}_i} = -2\sum_{t=1}^{T}(R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{Mt}) = -2\sum_{t=1}^{T}\hat{\varepsilon}_{it} \\
0 &= \frac{\partial \mathsf{SSR}(\hat{\alpha}_i, \hat{\beta}_i)}{\partial \hat{\beta}_i} = -2\sum_{t=1}^{T}(R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{Mt})R_{Mt} = -2\sum_{t=1}^{T}\hat{\varepsilon}_{it}R_{Mt}
\end{aligned}
$$

These are two linear equations in two unknowns. Solving for $\hat{\alpha}_i$ and $\hat{\beta}_i$ gives

$$
\begin{aligned}
\hat{\alpha}_i &= \hat{\mu}_i - \hat{\beta}_i \hat{\mu}_M \\
\hat{\beta}_i &= \frac{\hat{\sigma}_{iM}}{\hat{\sigma}^2_M}
\end{aligned}
$$

which are exactly the plug-in principle estimators!

Estimators for $\sigma^2_{\varepsilon,i}$ and $R - square$

Utilize plug-in principle

$$\begin{aligned}
\hat{\varepsilon}_{it} &= R_{it} - \hat{\alpha} - \hat{\beta}_i R_{Mt} \\
\hat{\sigma}^2_{\varepsilon,i} &= \frac{1}{T-2}\sum_{t=1}^{T}\hat{\varepsilon}^2_{it} \\
\hat{\sigma}_{\varepsilon,i} &= \sqrt{\hat{\sigma}^2_{\varepsilon,i}} = \text{SER} \\
&= \text{ standard error of regression}
\end{aligned}$$

Remarks

- $\hat{\sigma}_{\varepsilon,i}$ typical magnitude of residual $=$ standard error of regression (SER)

- Divide by $T-2$ to get unbiased estimate of $\sigma^2_{\varepsilon,i}$

- $T-2 =$ degrees of freedom $=$ sample size - number of estimated parameters (α_i and β_i)

Recall

$$\begin{aligned} R_i^2 &= \frac{\beta_i^2 \sigma_M^2}{\sigma_i^2} \\ &= 1 - \frac{\sigma_{\varepsilon,i}^2}{\sigma_i^2} \\ &= \text{\% of variability due to market} \end{aligned}$$

Estimate using plug-in principle

$$\begin{aligned} \hat{R}_i^2 &= \frac{\hat{\beta}_i^2 \hat{\sigma}_M^2}{\hat{\sigma}_i^2} \\ &= 1 - \frac{\hat{\sigma}_{\varepsilon,i}^2}{\hat{\sigma}_i^2} \end{aligned}$$

Least Squares Estimation Using R

R command

`lm` - linear model estimation

Syntax

```
lm.fit = lm(y~x,data=my.data.df)
```

`my.data.df` = data frame with columns named `y` and `x`

Note: `y~x` is formula notation in R. It translates as the linear model

$$y = \alpha + \beta x + \varepsilon$$

For multiple regression, the notation `y~x1+x2` implies

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Important method functions for lm objects

- `summary()`: summarize model fit
- `plot()`: plot results
- `residuals()`: extract residuals
- `fitted()`: extract fitted values
- `coef()`: extract estimated coefficients
- `confint()`: extract confidence intervals

Least Squares Estimates are Maximum Likelihood Estimates Under Normal Distribution Assumption

$$
\begin{gathered}
R_{it} = \alpha_i + \beta_i R_{Mt} + \varepsilon_{it},\ t = 1, \ldots, T \\
\varepsilon_{it} \sim \text{iid } N(0, \sigma^2_{\varepsilon,i}),\ R_{M,t} \sim \text{iid } N(\mu_M, \sigma^2_M)
\end{gathered}
$$

Then

$$
\begin{aligned}
R_{it}|R_{Mt} &\sim \ N(\alpha_i + \beta_i R_{Mt}, \sigma^2_{\varepsilon,i}) \\
f(R_{it}|R_{Mt}) &= (2\pi\sigma^2_{\varepsilon,i})^{-1/2} \exp\left(\frac{-1}{2\sigma^2_{\varepsilon,i}} (R_{it} - \alpha_i + \beta_i R_{Mt})^2\right) \\
\ln L(\theta|\mathbf{R}, \mathbf{R}_M) &= \frac{-T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2_{\varepsilon,i}) \\
&\quad - \frac{1}{2\sigma^2_{\varepsilon,i}} \sum_{t=1}^{T} (R_{it} - \alpha_i + \beta_i R_{Mt})^2
\end{aligned}
$$

Maximizing ln $L(\theta|\mathbf{R}, \mathbf{R}_M)$ with respect to $\theta = (\alpha_i, \beta_i, \sigma^2_{\varepsilon,i})'$ gives the least squares estimates!

Statistical Properties of Least Squares Estimates

Assuming the SI model generates the observed data, the estimators

$$\hat{\alpha}_i,\ \hat{\beta}_i \text{ and } \hat{\sigma}^2_{\varepsilon,i}$$

are random variables.

Properties

- $\hat{\alpha}_i$, $\hat{\beta}_i$ and $\hat{\sigma}^2_{\varepsilon,i}$ are unbiased estimators

$$\begin{aligned} E[\hat{\alpha}_i] &= \alpha_i \\ E[\hat{\beta}_i] &= \beta_i \\ E[\hat{\sigma}^2_{\varepsilon,i}] &= \sigma^2_{\varepsilon,i} \end{aligned}$$

- Analytic standard errors are available for $\widehat{\mathsf{SE}}(\hat{\alpha}_i)$ and $\widehat{\mathsf{SE}}(\hat{\beta}_i)$

$$\widehat{\mathsf{SE}}(\hat{\alpha}_i) = \frac{\hat{\sigma}_{\varepsilon,i}}{\sqrt{T \cdot \hat{\sigma}_M^2}} \cdot \sqrt{\frac{1}{T}\sum_{t=1}^{T} R_{Mt}^2}$$
$$\widehat{\mathsf{SE}}(\hat{\beta}_i) = \frac{\hat{\sigma}_{\varepsilon,i}}{\sqrt{T \cdot \hat{\sigma}_M^2}}$$

These are routinely reported in standard regression ouput (e.g. by R `summary` command)

  - $\widehat{\mathsf{SE}}(\hat{\alpha}_i)$ and $\widehat{\mathsf{SE}}(\hat{\beta}_i)$ are smaller the smaller is $\hat{\sigma}_{\varepsilon,i}$
  - $\widehat{\mathsf{SE}}(\hat{\beta}_i)$ is smaller the larger is $\hat{\sigma}_M^2$
  - $\widehat{\mathsf{SE}}(\hat{\alpha}_i)$ and $\widehat{\mathsf{SE}}(\hat{\beta}_i) \rightarrow 0$ as T gets large $\Rightarrow$ $\hat{\alpha}_i$ and $\hat{\beta}_i$ are consistent estimators

- Standard errors for $\hat{\sigma}^2_{\varepsilon,i}$, $\hat{\sigma}_{\varepsilon,i}$ or $R-$square can be computed using the bootstrap

- For T large enough, the central limit theorem (CLT) tells us that

$$\hat{\alpha}_i \sim N(\alpha_i, \widehat{\mathsf{SE}}(\hat{\alpha}_i)^2)$$
$$\hat{\beta}_i \sim N(\beta_i, \widehat{\mathsf{SE}}(\hat{\beta}_i)^2)$$

- Approximate 95% confidence intervals

$$\hat{\alpha}_i \pm 2 \cdot \widehat{\mathsf{SE}}(\hat{\alpha}_i)$$
$$\hat{\beta}_i \pm 2 \cdot \widehat{\mathsf{SE}}(\hat{\beta}_i)$$

SI Model Using Matrix Algebra

$$R_{it} = \alpha_i + \beta_i R_{Mt} + \varepsilon_{it},\ t = 1, \ldots, T$$

Stack over observations $t = 1, \ldots, T$

$$\begin{pmatrix} R_{i1} \\ \vdots \\ R_{iT} \end{pmatrix} = \alpha_i \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_i \begin{pmatrix} R_{M1} \\ \vdots \\ R_{MT} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}$$

or

$$\begin{aligned}
\mathbf{R}_i &= \alpha_i \cdot \mathbf{1} + \beta_i \cdot \mathbf{R}_M + \varepsilon_i = \begin{pmatrix} \mathbf{1} & \mathbf{R}_M \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} + \varepsilon_i \\
&= \mathbf{X}\gamma_i + \varepsilon_i \\
\mathbf{X} &= \begin{pmatrix} \mathbf{1} & \mathbf{R}_M \end{pmatrix},\ \gamma_i = \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix}
\end{aligned}$$

Recall the least squares normal equations

$$0 = \frac{\partial \mathsf{SSR}(\hat{\alpha}_i, \hat{\beta}_i)}{\partial \hat{\alpha}_i} = -2 \sum_{t=1}^{T} (R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{Mt})$$

$$0 = \frac{\partial \mathsf{SSR}(\hat{\alpha}_i, \hat{\beta}_i)}{\partial \hat{\beta}_i} = -2 \sum_{t=1}^{T} (R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{Mt}) R_{Mt}$$

Using matrix algebra these equations are

$$\begin{pmatrix} \sum_{t=1}^{T} R_{it} \\ \sum_{t=1}^{T} R_{it} R_{Mt} \end{pmatrix} = \begin{pmatrix} T & \sum_{t=1}^{T} R_{Mt} \\ \sum_{t=1}^{T} R_{Mt} & \sum_{t=1}^{T} R_{Mt}^2 \end{pmatrix} \begin{pmatrix} \hat{\alpha}_i \\ \hat{\beta}_i \end{pmatrix}$$

Equivalently,

$$\begin{pmatrix} \mathbf{1}'\mathbf{R}_i \\ \mathbf{R}_M'\mathbf{R}_i \end{pmatrix} = \begin{pmatrix} \mathbf{1}'\mathbf{1} & \mathbf{1}'\mathbf{R}_M \\ \mathbf{1}'\mathbf{R}_M & \mathbf{R}_M'\mathbf{R}_M \end{pmatrix} \begin{pmatrix} \hat{\alpha}_i \\ \hat{\beta}_i \end{pmatrix}$$

or

$$\mathbf{X}'\mathbf{R}_i = \mathbf{X}'\mathbf{X}\hat{\gamma}_i$$

Solving for $\hat{\gamma}_i$ gives the least squares estimates

$$\hat{\gamma}_i = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{R}_i$$

Estimating SI Model Covariance Matrix

Recall, in the SI model

$$\mathbf{\Sigma} = \sigma_M^2 \beta\beta' + \mathbf{D}$$

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}, \ \mathbf{D} = \begin{pmatrix} \sigma_{\varepsilon,1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_{\varepsilon,n}^2 \end{pmatrix}$$

Estimate $\mathbf{\Sigma}$ using plug-in principle

$$\hat{\mathbf{\Sigma}} = \hat{\sigma}_M^2 \hat{\beta}\hat{\beta}' + \hat{\mathbf{D}}$$

where

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{pmatrix}, \ \hat{\mathbf{D}} = \begin{pmatrix} \hat{\sigma}_{\varepsilon,1}^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \hat{\sigma}_{\varepsilon,n}^2 \end{pmatrix}$$

Hypothesis Testing in SI Model

Single Index Model and Assumptions

$$R_{it} = \alpha_i + \beta_i R_{Mt} + \varepsilon_{it}$$
$$\text{cov}(R_{Mt}, \varepsilon_{it}) = 0, \ \text{cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0, \ \text{cov}(\varepsilon_{it}, \varepsilon_{i,t-j}) = 0$$
$$R_{Mt} \sim iid\ N(\mu_M, \sigma_M^2)$$
$$\varepsilon_{it} \sim iid\ N(0, \sigma_{\varepsilon,i}^2)$$
$$\alpha_i, \beta_i, \mu_M, \sigma_M^2, \sigma_{\varepsilon,i}^2 \text{ are constant over time}$$

Hypotheses of Interest

- Basic significance test

$$H_0 : \beta_i = 0 \text{ vs. } H_1 : \beta_i \neq 0$$

- Test for specific value

$$H_0 : \beta_i = \beta_i^0 \text{ vs. } H_1 : \beta_i = \beta_i^0$$

- Test of constant parameters

$$H_0 : \beta_i \text{ is constant over entire sample}$$
$$H_1 : \beta_i \text{ changes in some sub-sample}$$

Basic significance test

$$H_0 : \beta_i = 0 \text{ vs. } H_1 : \beta_i \neq 0$$

Test statistics: t-statistics

$$t_{\beta_i=0} = \frac{\hat{\beta}_i - 0}{\widehat{SE}(\hat{\beta}_i)} = \frac{\hat{\beta}_i}{\widehat{SE}(\hat{\beta}_i)}$$

Intuition:

- If $|t_{\beta_i=0}| \approx 0$ then $\hat{\beta}_i \approx 0$, and $H_0 : \beta_i = 0$ should not be rejected

- If $|t_{\beta_i=0}| > 2$, say, then $\hat{\beta}_i$ more than 2 values of $\widehat{SE}(\hat{\beta}_i)$ away from 0. This is very unlikely if $\beta_i = 0$, so $H_0 : \beta_i = 0$ should be rejected.

Distribution of test statistics under H_0

Under the assumptions of the SI model, and $H_0 : \beta_i = 0$

$$t_{\theta=0} = \frac{\hat{\beta}_i}{\widehat{SE}(\hat{\beta}_i)} \sim t_{T-2}$$

where

$$t_{T-2} = \text{Student t distribution with}$$
$$T-2 \text{ degrees of freedom (d.f.)}$$

Remarks:

- t_{T-2} is bell-shaped and symmetric about zero (like normal)
- d.f. $=$ sample size - number of estimated parameters. In SI model there are two estimated parameters (α_i and β_i)
- Degrees of freedom determines kurtosis (tail thickness)

$$\text{d.f.} = T - 2 < 10,\ kurt(t_{T-2}) >> 3$$
$$\text{d.f.} = T - 2 > 60,\ kurt(t_{T-2}) \approx 3$$

- For $T \geq 60$, $t_{T-2} \sim N(0,1)$. Therefore, for $T \geq 60$

$$t_{\beta_i=0} = \frac{\hat{\beta}_i}{\widehat{SE}(\hat{\beta}_i)} \sim N(0,1)$$

Test for specific value

$$H_0 : \beta_i = \beta_{i0} \text{ vs. } H_1 : \beta_i \neq \beta_{i0}$$

Test statistics: t-statistics

$$t_{\beta_i=0} = \frac{\hat{\beta}_i - \beta_{i0}}{\widehat{SE}(\hat{\beta}_i)}$$

Intuition:

- If $|t_{\beta_i=\beta_{i0}}| \approx 0$ then $\hat{\beta}_i \approx \beta_{i0}$, and $H_0 : \beta_i = \beta_{i0}$ should not be rejected

- If $|t_{\beta_i=\beta_{i0}}| > 2$, say, then $\hat{\beta}_i$ more than 2 values of $\widehat{SE}(\hat{\beta}_i)$ away from β_{i0}. This is very unlikely if $\beta_i = \beta_{i0}$, so $H_0 : \beta_i = \beta_{i0}$ should be rejected.

Diagnostic for constant parameters: rolling Regression

Idea: Compute estimates of α_i and β_i from SI model over rolling windows of length $n < T$

$$R_{it}(n) = \alpha_i(n) + \beta_i(n) R_{Mt}(n) + \varepsilon_{it}(n)$$

If $\hat{\alpha}_i(n)$, $\hat{\beta}_i(n)$ are roughly constant over the rolling windows then the hypothesis that α_i and β_i are constant is supported by the data.