Inference with dependent observations

Javier Mencía

CEMFI 2011/2012

1 Pseudo-Maximum likelihood

1.1 Set up

- Data generating process:

$$x_t = \frac{u_t + \theta u_{t-1}}{\sqrt{1+\theta^2}}$$

  where $u_t \; iid \; \sim \; N(0,1)$.

- Goal: Estimate the mean $\mu = E(x_t)$ as well as its standard error for different values of $\theta \in (-1,1)$.

  - We will compute standard errors that are robust and non-robust to autocorrelation.
  - We will compare these estimators with the optimal one that exploits the fact that $x_t \sim MA(1)$.
  - Finally, we will consider a Monte Carlo estimator of the standard error.

- Estimator of the mean:

$$\hat{\mu}_T = \bar{x}_T = \frac{1}{T}\sum_{t=1}^{T} x_t$$

- Standard errors: $\sqrt{T}\hat{\mu}_T \xrightarrow{d} N(\mu, \mathbf{?})$

  1. Non-robust: assuming that x_t is *iid*:

$$\sqrt{T}\hat{\mu}_T \xrightarrow{d} N(0, \sigma_1^2)$$

     where

$$\hat{\sigma}_1^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{\mu}_T)^2$$

2. Robust to autocorrelation (Newey-West)

$$\sqrt{T}\hat{\mu}_T \xrightarrow{d} N(0, \sigma_2^2)$$

where σ_2^2 is estimated as

$$\begin{aligned}\hat{\sigma}_2^2 &= \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{\mu}_T)^2 \\ &\quad + 2\sum_{j=1}^{q}\left[1 - \frac{j}{q+1}\right]\left[\frac{\sum_{t=j+1}^{T}(x_t - \hat{\mu}_T)(x_{t-j} - \hat{\mu}_T)}{T-j}\right]\end{aligned}$$

This estimator does **NOT** assume that we know the true process of x_t.

3. Estimator when we know that the true DGP is a MA(1) (see appendix)

$$\sqrt{T}\hat{\mu}_T \xrightarrow{d} N(0, \sigma_3^2) \tag{1}$$

where

$$\sigma_3^2 = \frac{(1+\theta)^2}{1+\theta^2} = 1 + 2\gamma_1$$

with

$$\gamma_1 = \mathrm{cor}(x_t, x_{t-1}) = \frac{\theta}{1+\theta^2}.$$

1.2 Exercise 1

Compute the standard errors for a given value of θ.

Hint:

- Choose some $\theta \in (-1, 1)$.
- Generate x_t, for $t = 1, \cdots, 1000$.
- Compute the sample mean $\hat{\mu}_{T,j}$
- Compute $\hat{\sigma}_1^2, \hat{\sigma}_2^2$ y $\hat{\sigma}_3^2$:

1.3 Exercise 2

Compute the standard errors for a grid of values of θ.

Steps:

For each $\theta = \theta_i, i = 1, \cdots, n_1$

- Generate x_t, for $t = 1, \cdots, 1000$.
- Compute $\hat{\sigma}_1^2(\theta_i), \hat{\sigma}_2^2(\theta_i)$, and $\hat{\sigma}_3^2(\theta_i)$.
- Generate $j = 1, \cdots, n_2$ more independent draws of x_t, for $t = 1, \cdots, 1000$.
  - Estimate the sample mean $\hat{\mu}_{T,j}$ for each iteration.
  - Compute the Monte Carlo variance of $\hat{\mu}_{T,j}$:

$$\sigma_{rnd}^2(\theta_i) = \frac{1}{n_2} \sum_{j=1}^{n_2} \hat{\mu}_{T,j}^2 - \left[\frac{1}{n_2} \sum_{j=1}^{n_2} \hat{\mu}_{T,j} \right]^2$$

- Plot $\sigma_{rnd}^2(\theta)$, $\hat{\sigma}_1^2(\theta), \hat{\sigma}_2^2(\theta)$ and $\hat{\sigma}_3^2(\theta)$ against θ.

To estimate σ_3^2 we need a consistent estimator of θ. From the autocorrelation,

$$\sigma_3^2 = 1 + 2\gamma_1.$$

2 Specification tests

2.1 Set up

- Consider the regression:

$$y_t = \beta x_t + u_t \tag{2}$$

where

$$\begin{aligned} x_t &= \alpha y_{t-1} + v_t \\ u_t &= \rho u_{t-1} + w_t \end{aligned}$$

and v_t iid $\sim N(0, \sigma_v^2)$, w_t iid $\sim N(0, \sigma_w^2)$.

- Goal:
  - Test $H_0 : \rho = 0$.
  - Compute the Likelihood Ratio (LR), Wald (W) and Lagrange Multiplier (LM) tests.
  - Compute the size and power of these tests.

2.2 Restricted model (M_0)

- Estimate (2) subject to $\rho = 0$.
- The log-likelihood can be obtained using Bayes rule:

$$\log[f_0(y_t, x_t|I_{t-1})] = \log[f_0(y_t|x_t, I_{t-1})] + \log[f_0(x_t|I_{t-1})] \tag{3}$$

where $I_{t-1} = \{y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \cdots\}$ and

$$\begin{aligned} \log[f_0(y_t|x_t, I_{t-1})] &= -\frac{1}{2}\log(2\pi\sigma_w^2) - \frac{(y_t - \beta x_t)^2}{2\sigma_w^2} \\ \log[f_0(x_t|I_{t-1})] &= -\frac{1}{2}\log(2\pi\sigma_v^2) - \frac{(x_t - \alpha y_{t-1})^2}{2\sigma_v^2} \end{aligned}$$

- If we differentiate (3) with respect to $\boldsymbol{\theta} = (\beta, \alpha, \sigma_w^2, \sigma_v^2, \rho)'$, we obtain the score function

$$s_{0,t}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{(y_t-\beta x_t)x_t}{\sigma_w^2} \\ \frac{(x_t-\alpha y_{t-1})y_{t-1}}{\sigma_v^2} \\ \frac{(y_t-\beta x_t)^2-\sigma_w^2}{2\sigma_w^4} \\ \frac{(x_t-\alpha y_{t-1})^2-\sigma_v^2}{2\sigma_v^4} \\ \cdot \end{bmatrix}$$

- Hence, by solving $\sum_t s_t(\boldsymbol{\theta}) = \mathbf{0}$, we can obtain the maximum likelihood estimators:

$$
\begin{aligned}
\tilde{\beta} &= \frac{\sum_{t=2}^{T} y_t x_t}{\sum_{t=2}^{T} x_t^2} \\
\tilde{\alpha} &= \frac{\sum_{t=2}^{T} x_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} \\
\tilde{\sigma}_w^2 &= (1/T)\sum_{t=2}^{T}(y_t - \tilde{\beta}x_t)^2 \\
\tilde{\sigma}_v^2 &= (1/T)\sum_{t=2}^{T}(x_t - \tilde{\alpha}y_{t-1})^2 \\
\tilde{\rho} &= 0
\end{aligned}
$$

2.3 Unrestricted model (M_1)

- Unrestricted model (M_1): estimate (2) allowing for $\rho \neq 0$.
- Again, the log-likelihood can be obtained using Bayes rule:

$$
\log[f_1(y_t, x_t|I_{t-1})] = \log[f_1(y_t|x_t, I_{t-1})] + \log[f_1(x_t|I_{t-1})] \tag{4}
$$

where $I_{t-1} = \{y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, \cdots\}$ and

$$
\begin{aligned}
\log[f_1(y_t|x_t, I_{t-1})] &= -\frac{1}{2}\log(2\pi\sigma_w^2) - \frac{[y_t - \beta x_t - \rho(y_{t-1} - \beta x_{t-1})]^2}{2\sigma_w^2} \\
\log[f_1(x_t|I_{t-1})] &= -\frac{1}{2}\log(2\pi\sigma_v^2) - \frac{(x_t - \alpha y_{t-1})^2}{2\sigma_v^2}
\end{aligned}
$$

- If we differentiate (4) with respect to $\boldsymbol{\theta} = (\beta, \alpha, \sigma_w^2, \sigma_v^2, \rho)'$, we obtain the score function

$$
s_{1,t}(\boldsymbol{\theta}) = \begin{bmatrix}
\frac{[y_t - \beta x_t - \rho(y_{t-1} - \beta x_{t-1})](x_t - \rho x_{t-1})}{\sigma_w^2} \\
\frac{(x_t - \alpha y_{t-1})y_{t-1}}{\sigma_v^2} \\
\frac{[y_t - \beta x_t - \rho(y_{t-1} - \beta x_{t-1})]^2 - \sigma_w^2}{2\sigma_w^4} \\
\frac{(x_t - \alpha y_{t-1})^2 - \sigma_v^2}{2\sigma_v^4} \\
\frac{[y_t - \beta x_t - \rho(y_{t-1} - \beta x_{t-1})](y_{t-1} - \beta x_{t-1})}{\sigma_w^2}
\end{bmatrix}
$$

- If we knew ρ, we could express the estimates of the remaining parameters as

$$\begin{aligned}
\hat{\beta}(\rho) &= \frac{\sum_{t=2}^{T}(y_t-\rho y_{t-1})(x_t-\rho x_{t-1})}{\sum_{t=2}^{T}(x_t-\rho x_{t-1})^2} \\
\hat{\alpha} &= \frac{\sum_{t=2}^{T} x_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} \\
\hat{\sigma}_w^2(\rho) &= (1/T)\sum_{t=2}^{T}(\hat{u}_t-\rho\hat{u}_{t-1})^2 \\
\hat{\sigma}_v^2 &= (1/T)\sum_{t=2}^{T}(x_t-\hat{\alpha} y_{t-1})^2
\end{aligned}$$

- This is known as concentrating the log-likelihood on ρ.
- We can obtain $\hat{\rho}$ by finding the value of ρ that maximises this concentrated log-likelihood.

$$\max_{\rho}\sum_{t=2}^{T}\log[f_1(y_t,x_t|I_{t-1};\boldsymbol{\theta}=(\hat{\beta}(\rho),\hat{\alpha},\hat{\sigma}_w^2(\rho),\hat{\sigma}_v^2,\rho)')]$$

2.4 Tests

- LR requires estimating both M_0 and M_1:

$$LR = 2[L(M_1)-L(M_0)] \overset{H_0}{\to} \chi_1^2$$

  where $L(M_i)$ is the log-likelihood of the model M_i.

- LM only requires M_0:

$$LM = (1/T)\left[\sum_t s_{1,t}'(\tilde{\boldsymbol{\theta}})\right]\mathcal{I}_1^{-1}(\tilde{\boldsymbol{\theta}})\left[\sum_t s_{1,t}(\tilde{\boldsymbol{\theta}})\right] \overset{H_0}{\to} \chi_1^2$$

  where

$$\mathcal{I}_1(\tilde{\boldsymbol{\theta}}) = \frac{1}{T}\sum_t s_{1,t}(\tilde{\boldsymbol{\theta}})s_{1,t}'(\tilde{\boldsymbol{\theta}})$$

- W only requires M_1:

$$W = \frac{T\hat{\rho}^2}{se^2(\sqrt{T}\hat{\rho}_T)} \to \chi_1^2$$

  where $se(\sqrt{T}\hat{\rho})$ the square root of the $(5,5)$ element of the inverse of:[1]

$$\mathcal{I}_1(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_t s_{1,t}(\hat{\boldsymbol{\theta}})s_{1,t}'(\hat{\boldsymbol{\theta}})$$

[1]Notice that $\sqrt{T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_{true}) \overset{H_1}{\to} N(\mathbf{0},\mathcal{I}_1^{-1}(\boldsymbol{\theta}_{true}))$, where $\boldsymbol{\theta}_{true}$ is the true value of $\boldsymbol{\theta}$.

2.5 Exercise 1

- Set $\beta = .5, \alpha = .3, \sigma_w = 1, \sigma_v = 1$
- Choose some $\rho \in (-1, 1)$.
- Simulate x_t and y_t with T=2000.
- Burn the first 1001 observations.
- Compute LR, W and LM.
- Compute the p-values:

$$\begin{aligned} p_{LR} &= 1 - F_{\chi^2_1}(LR) \\ p_{LM} &= 1 - F_{\chi^2_1}(LM) \\ p_{Wald} &= 1 - F_{\chi^2_1}(Wald) \end{aligned}$$

  where $F_{\chi^2_1}(\cdot)$ is the cdf of a chi-square with 1 degree of freedom.

2.6 Exercise 2

- Set $\beta = .5, \alpha = .3, \sigma_w = 1, \sigma_v = 1$
- Choose some $\rho \in (-1, 1)$.
- Repeat from $i = 1$ to n:

  1. Simulate x_t and y_t with T=2000.
  2. Burn the first 1001 observations.
  3. Compute LR_i, W_i and LM_i and store the values in the $i-th$ rows of the vectors LR, LM and $Wald$, respectively.

- Check that the means of LR, LM and $Wald$ are approximately 1 if $\rho = 0$ and greater than 1 for any $\rho \neq 0$.
- Compute the empirical p-value at the 90% confidence level of the three tests:

  - Proportion of times that the LR, LM and $Wald$ tests are greater than the 90% critical value of the chi-square of 1 degree of freedom.

- Check that these empirical p-values are close to .1 if $\rho = 0$ and greater than .1 for any $\rho \neq 0$.
- P-value discrepancy plot:

- Consider a grid of nominal p-values from 0 to 1: α_i, $i = 1, \cdots, m$.
- For each α_i, compute the empirical p-value of each test.
- Plot the nominal p-values on the x-axis and the empirical ones on the y-axis.
- Check that the plots are on the diagonal under H_0, and above the diagonal under H_1.

A Standard error of a MA(1)

We can always write $x_1, \cdots, x_t$ as:

$$\begin{aligned}
\sqrt{1+\theta^2} x_T &= u_T + \theta u_{T-1} \\
\sqrt{1+\theta^2} x_{T-1} &= u_{T-1} + \theta u_{T-2} \\
\sqrt{1+\theta^2} x_{T-2} &= u_{T-2} + \theta u_{T-3} \\
&\vdots \\
\sqrt{1+\theta^2} x_2 &= u_2 + \theta u_1 \\
\sqrt{1+\theta^2} x_1 &= u_1.
\end{aligned}$$

Hence, we can express the sample mean as:

$$\sqrt{T}\hat{\mu}_T = \frac{1}{\sqrt{T}\sqrt{1+\theta^2}} u_T + \frac{1+\theta}{\sqrt{T}\sqrt{1+\theta^2}} \sum_{t=1}^{T-1} u_t$$

whose variance is

$$V(\sqrt{T}\hat{\mu}_T) = \frac{1}{T(1+\theta^2)} + \frac{(1+\theta)^2(T-1)}{(1+\theta^2)T}$$

If we take the limit $T \to \infty$, we obtain (1).