A Primer in Econometric Theory

Lecture 12: Large Samples and Dependence

John Stachurski
Lectures by Akshay Shanker

March 26, 2017

Large Sample Least Squares

Large samples allow us to drop parametric assumptions on the error term we made for finite sample inference

Theory developed below also useful for cross-sectional environments with no correlation between observations

Assume data $(y_1, \mathbf{x}_1), \ldots, (y_T, \mathbf{x}_T)$ generated by the linear model

$$y_t = \mathbf{x}_t^\mathsf{T} \boldsymbol{\beta} + u_t, \qquad t = 1, \ldots, T \tag{1}$$

- $\boldsymbol{\beta}$ is a K-vector of unknown coefficients, and u_t is an unobservable shock
- observations indexed by t rather than n to remind us that observations are dependent
- sample size will be denoted by T

Let:

- $\mathbf{y}$ be the $T \times 1$ vector of observed outputs
- y_t is the tth element of $\mathbf{y}$
- $\mathbf{u}$ is the vector of shocks
- u_t is the tth element of $\mathbf{u}$

Let $\mathbf{X}$ be the $T \times K$ matrix $\mathbf{X} := (x_{tk})$, where $1 \leq t \leq T$ and $1 \leq k \leq K$

Estimate the parameter vector $\boldsymbol{\beta}$ via least squares

The OLS estimate:

$$\hat{\boldsymbol{\beta}}_T = \left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{\intercal} \right]^{-1} \cdot \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t y_t$$

Expression for the sampling error in (12.4) can be expanded into sums to obtain

$$\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} = \left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{\intercal} \right]^{-1} \cdot \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t u_t \tag{2}$$

Drop the exogeneity assumption $\mathbb{E}[\mathbf{u} \mid \mathbf{X}] = \mathbf{0}$

For example, exogeneity fails when we estimate AR(1) model $y_{t+1} = \beta y_t + u_{t+1}$

Setting $x_t = y_{t-1}$ produces the regression model

$$y_t = \beta x_t + u_t, \qquad t = 1, \ldots, T$$

Regressor correlated with lagged values of the shock

Assumption.(13.1.1) The matrix $\mathbf{X}$ is full column rank with probability one and the sequence $\{\mathbf{x}_t\}$ is stationary. Moreover

1. $\boldsymbol{\Sigma}_{\mathbf{x}} := \mathbb{E}\left[\mathbf{x}_t \mathbf{x}_t^\intercal\right]$ exists and is positive definite, and
2. the sequence $\{\mathbf{x}_t\}$ satisfies $\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^\intercal \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{x}}$ as $T \to \infty$.

Example. Let $\{x_t\}$ be the Markov process in example 7.2.11

To repeat

$$x_{t+1} = a|x_t| + (1 - a^2)^{1/2} w_{t+1}$$

$$\text{with} \quad -1 < a < 1 \quad \text{and} \quad \{w_t\} \overset{\text{IID}}{\sim} \text{N}(0, 1)$$

The model has a unique, globally stable stationary distribution π_∞

If $\mathcal{L}(x_0) = \pi_\infty$, then the process $\{x_t\}$ is stationary and all of the conditions in assumption 13.1.1 are satisfied (see ex. 13.4.3)

Assumption.(13.1.2)[Weak exogeneity]

The shocks $\{u_t\}$ are IID

Moreover

1. $\mathbb{E}[u_t] = 0$ and $\mathbb{E}[u_t^2] = \sigma^2$ for all t, and
2. u_t is independent of $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t$ for all t

Example. (13.1.2) In the AR(1) regression (6), assumption 13.1.2 holds if shocks $\{u_t\}$ are IID

- contemporaneous and lagged regressors $x_1, \ldots, x_t$ are equal to the lagged state variables $y_0, \ldots, y_{t-1}$
- $y_0, \ldots, y_{t-1}$ are functions of only y_0 and $u_1, \ldots, u_{t-1}$, and therefore independent of u_t

A consequence of assumption 13.1.2

$$\mathbb{E}\left[u_s u_t \,|\, \mathbf{x}_1, \ldots, \mathbf{x}_t\right] = \begin{cases} \sigma^2 & \text{if} \quad s = t \\ 0 & \text{if} \quad s < t \end{cases}$$

The proof is an exercise (ex. 13.4.4)

Implication of assumptions 13.1.1 and 13.1.2: linear functions of $\{\mathbf{x}_t u_t\}$ form a martingale difference sequence (MDS)

Lemma. (13.1.1) if assumptions 13.1.1 and 13.1.2 both hold, then, for any constant vector $\mathbf{a} \in \mathbb{R}^K$, the sequence $\{m_t\}$ defined by $m_t = \mathbf{a}^\top \mathbf{x}_t u_t$ is

1. stationary with $\mathbb{E}[m_t^2] = \sigma^2 \mathbf{a}^\top \mathbf{\Sigma}_{\mathbf{x}} \mathbf{a}$ for all t, and
2. an MDS with respect to the filtration defined by

$$\mathscr{F}_t := \{\mathbf{x}_1, \ldots, \mathbf{x}_t, \mathbf{x}_{t+1}, u_1, \ldots, u_t\}$$

Proof.

First let's check part 1.

That $\{m_t\}$ is stationary follows from the assumption that $\{u_t\}$ and $\{\mathbf{x}_t\}$ are stationary

Regarding the second moment $\mathbb{E}[m_1^2]$, we have

$$\mathbb{E}[m_1^2] = \mathbb{E}[\mathbb{E}[u_1^2(\mathbf{a}^\top \mathbf{x}_1)^2 \mid \mathbf{x}_1]] = \mathbb{E}[(\mathbf{a}^\top \mathbf{x}_1)^2 \mathbb{E}[u_1^2 \mid \mathbf{x}_1]]$$

From independence of u_1 and $\mathbf{x}_1$, the inner expectation is σ^2

Moreover

$$(\mathbf{a}^\top \mathbf{x}_1)^2 = \mathbf{a}^\top \mathbf{x}_1 \mathbf{a}^\top \mathbf{x}_1 = \mathbf{a}^\top \mathbf{x}_1 \mathbf{x}_1^\top \mathbf{a}$$

$$\therefore \quad \mathbb{E}[m_1^2] = \mathbb{E}[\mathbf{a}^\top \mathbf{x}_1 \mathbf{x}_1^\top \mathbf{a}\, \sigma^2] = \sigma^2 \mathbf{a}^\top \mathbb{E}[\mathbf{x}_1 \mathbf{x}_1^\top] \mathbf{a} = \sigma^2 \mathbf{a}^\top \mathbf{\Sigma}_x \mathbf{a}$$

To check part 2., note $\{m_t\}$ is adapted to $\{\mathscr{F}_t\}$, since $m_t := u_t \mathbf{a}^\top \mathbf{x}_t$ is a function of variables in $\mathscr{F}_t$

Moreover we have

$$\begin{aligned}\mathbb{E}\left[m_{t+1} \mid \mathscr{F}_t\right] = \mathbb{E}\left[u_{t+1} \mathbf{a}^\top \mathbf{x}_{t+1} \mid \mathscr{F}_t\right] &= \mathbf{a}^\top \mathbf{x}_{t+1} \mathbb{E}\left[u_{t+1} \mid \mathscr{F}_t\right] \\ &= \mathbf{a}^\top \mathbf{x}_{t+1} \mathbb{E}\left[u_{t+1}\right] = 0\end{aligned}$$

This confirms $\{m_t\}$ is an MDS with respect to $\{\mathscr{F}_t\}$

Consistency

Under the conditions of §13.1.1, the OLS estimator $\hat{\boldsymbol{\beta}}_T$ is consistent for $\boldsymbol{\beta}$:

Theorem. (13.1.1) If assumptions 13.1.1 and 13.1.2 hold, then

$$\hat{\boldsymbol{\beta}}_T \xrightarrow{p} \boldsymbol{\beta} \quad \text{as} \quad T \to \infty$$

Proof. Recall equation (13.2):

$$\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} = \left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^\top \right]^{-1} \cdot \; \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t u_t$$

We show the expression on the right-hand converges in probability to $\mathbf{0}$

First, let's show $\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t u_t \xrightarrow{p} \mathbf{0}$. In view of fact 6.1.1, it suffices to show that, for any $\mathbf{a} \in \mathbb{R}^K$,

$$\mathbf{a}^\top \left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t u_t \right] \xrightarrow{p} \mathbf{a}^\top \mathbf{0} = 0 \tag{3}$$

Define $m_t := \mathbf{a}^\top \mathbf{x}_t u_t$. The left-hand side of (3) can be written as $T^{-1} \sum_{t=1}^{T} m_t$

Proof.(cont.) Since $\{m_t\}$ is a stationary MDS (lemma 13.1.1), the convergence $T^{-1}\sum_{t=1}^{T} m_t \xrightarrow{p} 0$ follows from Theorem 7.3.1

Return to the expression on the right-hand side of (13.2)

By assumption 13.1.1 and fact 6.2.1, we see that

$$\left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{\intercal}\right]^{-1} \xrightarrow{p} \mathbf{\Sigma_x}^{-1} \quad \text{as} \quad T \to \infty \tag{4}$$

Appealing to fact 6.2.1 once more, we obtain

$$\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} = \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{\intercal}\right]^{-1} \cdot \frac{1}{T}\sum_{t=1}^{T} u_t \mathbf{x}_t \xrightarrow{p} \mathbf{\Sigma_x}^{-1} \mathbf{0} = \mathbf{0}$$

Theorem. (13.1.2) If assumptions 13.1.1 and 13.1.2 hold, then

$$\hat{\sigma}_T^2 \xrightarrow{p} \sigma^2 \quad \text{as} \quad T \to \infty$$

Proof. By the definition of $\hat{\sigma}_T^2$ and the linear model assumption 1,

$$\hat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T}(y_t - \mathbf{x}_t^\intercal \hat{\boldsymbol{\beta}}_T)^2 = \frac{1}{T}\sum_{t=1}^{T}\left[u_t + \mathbf{x}_t^\intercal (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)\right]^2$$

Proof.(cont.) Expand out the square

$$\hat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T} u_t^2 + 2(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)^\top \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t u_t + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)^\top \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^\top\right] (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)$$

By assumption 13.1.2 and the law of large numbers, the first term on the right-hand side converges in probability to σ^2

Show the second and third term converge in probability to zero as $T \to \infty$ — exercise using convergence results we have already established (refer to fact 6.2.1)

Asymptotic Normality

Theorem. (13.1.3) If assumptions 13.1.1 and 13.1.2 hold, then

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\right) \quad \text{as} \quad T \to \infty$$

Proof. Expression (2) gives

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) = \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{\intercal}\right]^{-1} \cdot \; T^{-1/2} \sum_{t=1}^{T} u_t \mathbf{x}_t$$

Let $\mathbf{z}$ be a random variable satisfying $\mathcal{L}(\mathbf{z}) = \mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_{\mathbf{x}})$

Proof.(cont.)

Suppose we can show

$$T^{-1/2} \sum_{t=1}^{T} u_t \mathbf{x}_t \xrightarrow{d} \mathbf{z} \quad \text{as} \quad T \to \infty \tag{5}$$

If (5) is valid, then, applying assumption 13.1.1 along with fact 6.2.2, we obtain

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) = \left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^{\top} \right]^{-1} \cdot T^{-1/2} \sum_{t=1}^{T} u_t \mathbf{x}_t \xrightarrow{d} \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{z}$$

Proof.(cont.)

Clearly $\mathbf{\Sigma}_{\mathbf{x}}^{-1}\mathbf{z}$ is Gaussian with zero mean

By symmetry of $\mathbf{\Sigma}_{\mathbf{x}}^{-1}$ (since $\mathbf{\Sigma}_{\mathbf{x}}$ is symmetric) the variance of $\mathbf{\Sigma}_{\mathbf{x}}^{-1}\mathbf{z}$ is

$$\mathbf{\Sigma}_{\mathbf{x}}^{-1} \operatorname{var}[\mathbf{z}] \, \mathbf{\Sigma}_{\mathbf{x}}^{-1} = \mathbf{\Sigma}_{\mathbf{x}}^{-1} \, \sigma^2 \, \mathbf{\Sigma}_{\mathbf{x}} \, \mathbf{\Sigma}_{\mathbf{x}}^{-1} = \sigma^2 \mathbf{\Sigma}_{\mathbf{x}}^{-1}$$

This completes the proof of theorem 13.1.3, conditional on (5)

Let's now check that (5) is valid

By the Cramér–Wold device (fact 6.1.6), suffices to show that for any $\mathbf{a} \in \mathbb{R}^K$, we have

$$\mathbf{a}^\mathsf{T} \left[T^{-1/2} \sum_{t=1}^{T} u_t \mathbf{x}_t \right] \xrightarrow{d} \mathbf{a}^\mathsf{T} \mathbf{z} \tag{6}$$

Proof.(cont.) Fix $\mathbf{a}$ and let $m_t := u_t \mathbf{a}^\top \mathbf{x}_t$; the expression on the left of (6) can be rewritten as

$$T^{-1/2} \sum_{t=1}^{T} m_t$$

Since $\mathcal{L}(\mathbf{z}) = \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{\Sigma}_{\mathbf{x}})$, to establish (6) we need to show

$$T^{-1/2} \sum_{t=1}^{T} m_t \xrightarrow{d} \mathrm{N}(0, \sigma^2 \mathbf{a}^\top \mathbf{\Sigma}_{\mathbf{x}} \mathbf{a}) \tag{7}$$

From lemma 13.1.1, we know $\{m_t\}$ is stationary with $\mathbb{E}[m_t^2] = \sigma^2 \mathbf{a}^\top \mathbf{\Sigma}_{\mathbf{x}} \mathbf{a}$ and an MDS with respect to the filtration given in (2)

Proof.(cont.)

By the martingale difference CLT, (7) holds whenever

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[m_t^2 \mid \mathscr{F}_{t-1}\right] \xrightarrow{p} \sigma^2 \mathbf{a}^\top \mathbf{\Sigma}_{\mathbf{x}} \mathbf{a} \quad \text{as } T \to \infty \tag{8}$$

Since $\mathbf{x}_t \in \mathscr{F}_{t-1}$, we have

$$\mathbb{E}\left[m_t^2 \mid \mathscr{F}_{t-1}\right] = \mathbb{E}\left[u_t^2 (\mathbf{a}^\top \mathbf{x}_t)^2 \mid \mathscr{F}_{t-1}\right]$$

$$= (\mathbf{a}^\top \mathbf{x}_t)^2 \mathbb{E}\left[u_t^2 \mid \mathscr{F}_{t-1}\right] = \sigma^2 (\mathbf{a}^\top \mathbf{x}_t)^2$$

Proof.(cont.)

Another way to write the last expression is $\sigma^2 \mathbf{a}^\top \mathbf{x}_t \mathbf{x}_t^\top \mathbf{a}$

The left-hand side of (8) is therefore

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[m_t^2 \mid \mathscr{F}_{t-1}\right] = \frac{1}{T}\sum_{t=1}^{T}(\sigma^2 \mathbf{a}^\top \mathbf{x}_t \mathbf{x}_t^\top \mathbf{a}) = \sigma^2 \mathbf{a}^\top \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t^\top\right] \mathbf{a}$$

which converges in probability to $\sigma^2 \mathbf{a}^\top \mathbf{\Sigma}_{\mathbf{x}} \mathbf{a}$ by assumption 13.1.1 and fact 6.2.1

This verifies (8), completing the proof □

Example. Consider again the scalar linear Gaussian AR(1) model $x_{t+1} = ax_t + w_{t+1}$ with $|a| < 1$ and $\{w_t\}$ IID and standard normal

Let $\{x_t\}$ be stationary

As discussed in §12.2.2, the OLS estimator of a is

$$\hat{a}_T := \frac{\mathbf{x}^\top \mathbf{y}}{\mathbf{x}^\top \mathbf{x}} \quad \text{where} \quad \mathbf{y} := (x_1, \ldots, x_T) \text{ and } \mathbf{x} := (x_0, \ldots, x_{T-1})$$

Both assumption 13.1.1 and assumption 13.1.2 are satisfied, so $\sqrt{T}(\hat{a}_T - a)$ converges in distribution to $\mathrm{N}(0, \sigma^2 \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})$

Example. (cont.) In this case, $\sigma^2 = 1$ because the shocks are standard normal

Furthermore $\mathbf{\Sigma}_{\mathbf{x}}^{-1}$ reduces to $1/\mathbb{E}\,[x_1^2]$, where the expectation is under the stationary distribution

The stationary distribution is $\mathrm{N}(0, 1/(1 - a^2))$ (recall our discussion in chapter 7 of ET, particularly surrounding Equation (7.18))

Hence the inverse of $\mathbb{E}\,[x_1^2]$ is $1 - a^2$, and

$$\sqrt{T}(\hat{a}_T - a) \xrightarrow{d} \mathrm{N}(0, 1 - a^2) \tag{9}$$

Large Sample Tests

In the large sample setting, the hypothesis to be tested:

$$H_0\colon \beta_k = \beta_k^0$$

Recall if the error terms are normally distributed, then the expression $(\hat{\beta}_k - \beta_k) / \operatorname{se}(\hat{\beta}_k)$ is t-distributed with $N - K$ degrees of freedom

- in the large sample case, we can use the CLT to show the same statistic is asymptotically normal

Theorem. (13.1.4) Let assumptions 13.1.1 and 13.1.2 hold, and let

$$\mathrm{se}(\hat{\beta}_k^T) := \sqrt{\hat{\sigma}_T^2 v_k(\mathbf{X})}$$

Under the null hypothesis H_0, we have

$$z_k^T := \frac{\hat{\beta}_k^T - \beta_k^0}{\mathrm{se}(\hat{\beta}_k^T)} \xrightarrow{d} \mathrm{N}(0,1) \quad \text{as} \quad T \to \infty \tag{10}$$

Proof. Recall from theorem 13.1.3 that $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \xrightarrow{d} \mathbf{z}$, where $\mathbf{z}$ is a random vector with distribution $\mathrm{N}(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_{\mathbf{x}}^{-1})$ and $\boldsymbol{\beta}$ is the true parameter vector

Hence

$$\sqrt{T}(\hat{\beta}_k^T - \beta_k) = \mathbf{e}_k^\top [\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})] \xrightarrow{d} \mathbf{e}_k^\top \mathbf{z}$$

The distribution of $\mathbf{e}_k^\top \mathbf{z}$ is $\mathrm{N}(0, \mathbf{e}_k^\top \operatorname{var}[\mathbf{z}] \mathbf{e}_k) = \mathrm{N}(0, \sigma^2 \mathbf{e}_k^\top \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{e}_k)$, so

$$\frac{\sqrt{T}(\hat{\beta}_k^T - \beta_k)}{\sqrt{\sigma^2 \mathbf{e}_k^\top \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \mathbf{e}_k}} \xrightarrow{d} \mathrm{N}(0, 1) \tag{11}$$

Proof.(cont.) Since

$$\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^\top\right]^{-1} \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{x}}{}^{-1} \quad \text{as} \quad T \to \infty$$

Now refer to our rules for convergence of random matrices, in particular, 5. of fact 6.2.1. We have

$$T v_k(\mathbf{X}) = T\mathbf{e}_k^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{e}_k = \mathbf{e}_k^\top\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t^\top\right]^{-1}\mathbf{e}_k \xrightarrow{p} \mathbf{e}_k^\top\boldsymbol{\Sigma}_{\mathbf{x}}{}^{-1}\mathbf{e}_k$$

By theorem 13.1.2 we have $\hat\sigma_T^2 \xrightarrow{p} \sigma^2$, and hence

$$\sqrt{\hat\sigma_T^2\, T v_k(\mathbf{X})} \xrightarrow{p} \sqrt{\sigma^2\mathbf{e}_k^\top\boldsymbol{\Sigma}_{\mathbf{x}}{}^{-1}\mathbf{e}_k}$$

Proof.(cont.) Combine the above with (11) to arrive at

$$\frac{\sqrt{T}(\hat{\beta}_k^T - \beta_k)}{\sqrt{\hat{\sigma}_T^2 \, T v_k(\mathbf{X})}} \xrightarrow{d} \mathrm{N}(0,1)$$

Assuming H_0 and canceling $\sqrt{T}$ gives (10)□

MLE for Markov Processes

Now turn to nonlinear estimation in a time series setting, using maximum likelihood

Consider a Markov process. Suppose:

- transition density $p_{\boldsymbol{\theta}}$ depends on some unknown parameter vector $\boldsymbol{\theta} \in \Theta$
- process has a unique stationary density $\pi_\infty^{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta}$, and that $\mathbf{x}_1$ is a draw from this stationary density

Log-likelihood function

$$\ell(\boldsymbol{\theta}) = \ln \pi_{\infty}^{\boldsymbol{\theta}}(\mathbf{x}_1) + \sum_{t=1}^{T-1} \ln p_{\boldsymbol{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{x}_t)$$

In practice drop the first term in this expression

- influence of a single element is likely to be negligible

Abusing notation slightly, write

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{T-1} \ln p_{\boldsymbol{\theta}}(\mathbf{x}_{t+1} \mid \mathbf{x}_t) \tag{12}$$

The ARCH Case

Recall the ARCH model

Suppose $x_t = \sigma_t w_t$ where $\sigma_{t+1}^2 = \alpha_0 + \alpha_1 x_t^2$

Combining these equations:

$$x_{t+1} = (\alpha_0 + \alpha_1 x_t^2)^{1/2} w_{t+1} \quad \text{with} \quad \{w_t\} \overset{\text{IID}}{\sim} \text{N}(0,1) \tag{13}$$

where $\alpha_0 > 0$, $\alpha_1 \geq 0$

By (12), the log-likelihood function is

$$\ell(a,b) = \sum_{t=1}^{T-1} \left\{ -\frac{1}{2} \ln(2\pi(a + bx_t^2)) - \frac{x_{t+1}^2}{2(a + bx_t^2)} \right\} \tag{14}$$

Rearranging, dropping terms that don't depend on a or b, and multiplying by 2 (an increasing transformation), rewrite as

$$\ell(a,b) = -\sum_{t=1}^{T-1} \left\{ \ln z_t + \frac{x_{t+1}^2}{z_t} \right\} \quad \text{where} \quad z_t := a + b x_t^2 \qquad (15)$$

Solution method

- no analytical expressions for the MLEs
- need to use numerical routines — `R`'s inbuilt optimization routines

Sequence of observations $x_1, \dots, x_T$ stored in a vector `xdata`

the function `arch_like` can be optimized numerically via the commands:

```
start_theta <- c(0.65, 0.35)  # An initial guess of (a,b)
neg_like <- function(theta) {
    return(-arch_like(theta, xdata))
}
opt <- optim(start_theta, neg_like, method="BFGS")
```

Code to define function `arch_like` and simulate observations on following slide

```r
arch_like <- function(theta, data) {
    Y <- data[-1]              #  All but first element
    X <- data[-length(data)]   #  All but last element
    Z <- theta[1] + theta[2] * X^2
    return(-sum(log(Z) + Y^2 / Z))
}

sim_data <- function(a, b, n=500) {
    x <- numeric(n)
    x[1] = 0
    w = rnorm(n)
    for (t in 1:(n-1)) {
        x[t+1] = sqrt(a + b * x[t]^2) * w[t]
    }
    return(x)
}

xdata <- sim_data(0.5, 0.5)  #  True parameters
```

The Newton–Raphson Algorithm

The Newton–Raphson algorithm is a *root-finding* algorithm

- given a function $g \colon \mathbb{R} \to \mathbb{R}$, the algorithm searches for points $\bar{s} \in \mathbb{R}$ such that $g(\bar{s}) = 0$

Optimize differentiable functions

- for differentiable functions, interior optimizers are always roots of the objective function's first derivative

Let

- $g \colon \mathbb{R} \to \mathbb{R}$
- s_0 be some initial point in $\mathbb{R}$ that we think (hope) is somewhere near a root

We know how to move to the root of the function that forms the *tangent line* to g at s_0

Replace g with its linear approximation around s_0, given by

$$\tilde{g}(s) := g(s_0) + g'(s_0)(s - s_0) \qquad (s \in \mathbb{R})$$

and solve for the root of $\tilde{g}$

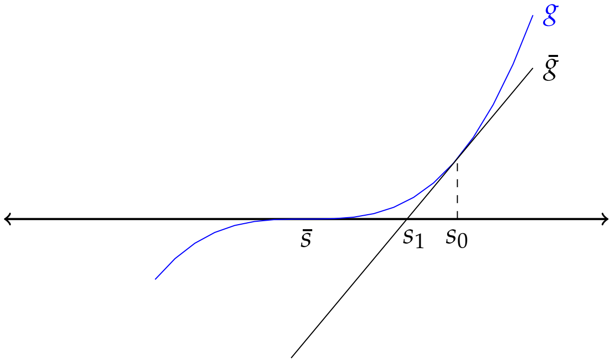

Figure: First step of the Newton–Raphson algorithm

Next guess of the root $s_1 := s_0 - g(s_0)/g'(s_0)$

Procedure is repeated, taking the tangent of g at s_1

Generates a sequence of points $\{s_k\}$ satisfying

$$s_{k+1} = s_k - \frac{g(s_k)}{g'(s_k)}$$

Various results telling us that when g is suitably well-behaved and s_0 is sufficiently close to a given root $\bar{s}$, then sequence $\{s_k\}$ will converge to $\bar{s}$

In practical situations we often have no way of knowing whether the conditions are satisfied, and there have been many attempts to make the procedure more robust

The R optimization routine described above is a child of this process

Optimization

Suppose $g \colon \mathbb{R} \to \mathbb{R}$ is a twice differentiable function we wish to maximize

If s^* is a maximizer of g, then $g'(s^*) = 0$

Apply the Newton–Raphson algorithm to g', giving the sequence

$$s_{k+1} = s_k - \frac{g'(s_k)}{g''(s_k)}$$

Multivariate case: suppose g is twice differentiable and $g \colon \mathbb{R}^2 \to \mathbb{R}$s

The **gradient vector** and **Hessian** of g at $(x, y) \in \mathbb{R}^2$ are defined as

$$\nabla g(x,y) := \begin{pmatrix} g_1'(x,y) \\ g_2'(x,y) \end{pmatrix}$$

and

$$\nabla^2 g(x,y) := \begin{pmatrix} g_{11}''(x,y) & g_{12}''(x,y) \\ g_{21}''(x,y) & g_{22}''(x,y) \end{pmatrix}$$

Here g_i' is the first partial of g with respect to its ith argument, g_{ij}'' is the second cross-partial, and so on

Newton–Raphson algorithm generates the sequence $\{(x_k, y_k)\}$ defined by

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) - [\nabla^2 g(x_k, y_k)]^{-1} \nabla g(x_k, y_k)$$

from some initial guess (x_0, y_0)

(Assuming the Hessian matrix is nonsingular)

Consider maximization of the log-likelihood function for the ARCH model — Equation (15) above

Let z_t be as defined in (15)

The first partials are

$$\frac{\partial \ell}{\partial a}(a,b) = \sum_{t=1}^{T-1} \left[\frac{x_{t+1}^2}{z_t^2} - \frac{1}{z_t} \right], \quad \frac{\partial \ell}{\partial b}(a,b) = \sum_{t=1}^{T-1} x_t^2 \left[\frac{x_{t+1}^2}{z_t^2} - \frac{1}{z_t} \right]$$

The second partials are

$$\frac{\partial^2 \ell}{\partial a^2}(a,b) = \sum_{t=1}^{T-1} \left[\frac{1}{z_t^2} - 2\frac{x_{t+1}^2}{z_t^3} \right]$$

and

$$\frac{\partial^2 \ell}{\partial b^2}(a,b) = \sum_{t=1}^{T-1} x_t^4 \left[\frac{1}{z_t^2} - 2\frac{x_{t+1}^2}{z_t^3} \right]$$

The cross-partial is

$$\frac{\partial^2 \ell}{\partial a \partial b}(a,b) = \sum_{t=1}^{T-1} x_t^2 \left[\frac{1}{z_t^2} - 2\frac{x_{t+1}^2}{z_t^3} \right]$$

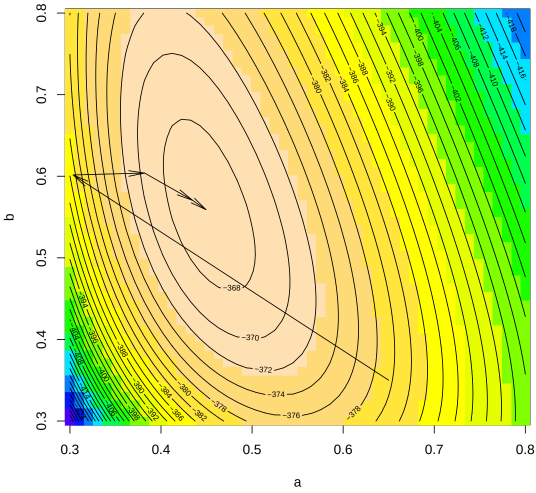

Figure: Newton–Raphson iterates