# Contents

# Chapter 5

# Asymptotics and Hypothesis Testing and Dynamic Common Factor Models

## 5.1 Asymptotics

- The asymptotic properties are really straightforward once we have bought the normality assumption and are in the maximum likelihood environment.
- In the case of weakly stationary data the asympotics are the same as the classical regression model (only large sample results are typically attainable)
- We know that these estimators are asymptotically efficient (achieve the Cramer-Rao Lower bound), are root $T$ consistent and asymptotically normal.
- The variance-covariance matrix is the inverse of the information matrix and from this we may apply standard methods for hypothesis testing.
- We briefly review the asymptotic theory of maximum likelihood with special attention to the theory for calculating the variance covariance matrix and consider some hypothesis tests.
- The basic results are stated in Chapter 1. We repeat the argument here.
- The joint density function for observations $y_1, \ldots, y_T$ is assumed to depend upon $n$ unknown parameters in the vector $\mathbf{\Psi} = (\psi_1, \ldots, \psi_n)^T$. We denote the likelihood function given the data $y_1, \ldots, y_T$ as

$$L(\mathbf{\Psi}; y_1, \ldots, y_T) \text{ or } L(\mathbf{\Psi}) \tag{5.1}$$

- We maximize the likelihood w.r.t. $\mathbf{\Psi}$ by setting the score to zero

$$\frac{\partial logL}{\partial \mathbf{\Psi}} = 0 \qquad \} \text{ likelihood equations} \tag{5.2}$$

- The likelihood equations are in general (when there are moving average coefficients to estimate) nonlinear in the parameters which will require an iterative procedure which we discussed in Chapter 4.
- The properties of $MLE$ estimators are well known. Let $\tilde{\Psi} = \{\tilde{\psi}_1, \ldots, \tilde{\psi}_n\}$ be the $MLE$ estimate

$$\sqrt{T}(\tilde{\Psi} - \Psi) \overset{d}{\sim} N(0, T\ AVar(\Psi)) \text{ asymptotically} \tag{5.3}$$

  where **AVar** is the asymptotic variance covariance matrix obtained in the usual way from the probability limit of the *inverse information matrix*:

$$\mathbf{IA}(\Psi) = p\lim T^{-1}\{-\frac{\partial^2 log(L)}{\partial\Psi\partial\Psi^T}\}. \tag{5.4}$$

- That is $\mathbf{Avar}(\tilde{\Psi}) = T^{-1}\mathbf{IA}(\Psi)^{-1}$ and in practice we estimate the variance covariance matrix by evaluating the **Avar** at the MLE : $\tilde{\Psi}$.
- The information matrix can also be obtained from taking the **outer product of the gradients** (hence the idea behind information matrix tests) :

$$\mathbf{IA}(\Psi) = p\lim T^{-1}\{\frac{\partial log(L)}{\partial\Psi}\frac{\partial log(L)}{\partial\Psi}^T\}. \tag{5.5}$$

- If we let $\epsilon_t(Y;\Psi)$ be the error term $NID(0,\sigma^2)$ for either an $AR(p), MA(q)$ or ARMA$(p,q)$ then $\Psi = (\phi^T, \theta^T)$

$$\mathbf{IA}(\mathbf{\Psi}) = \sigma^{-2}\text{plim}\quad T^{-1}\sum_t \left(\frac{\partial\epsilon_t}{\partial\mathbf{\Psi}}\right)\left(\frac{\partial\epsilon_t}{\partial\mathbf{\Psi}}\right)^T \tag{5.6}$$

- Most programs like STATA/RATS will simply take the inverse of the Hessian evaluated at the **final iteration** and use $\tilde{\sigma}^2$.
- Such estimators do not impose diagonality of the information matrix and so present additional issues in finite samples

## 5.2 Asymptotics of Time Series Models

- Perhaps the simplest time series models from the point of view of both estimation of the parameters and the estimation of the variance covariance matrix is the $AR(p)$ process.
- Recall that the parameters can be simply estimated by OLS (in STATA that is the regress command).
- It also turns out that the estimation of the variance covariance follows the same basic formula and since the model is linear it is easy to obtain.

### 5.2.1 The $AR(1)$ Case

$$y_t \;=\; \phi y_{t-1} + \epsilon_t \tag{5.7}$$

where $\epsilon_t \sim NID(0, \sigma^2)$. For this case, it can be shown (see Fuller pp. 327-332) that:

$$\mathbf{Avar}(\tilde{\phi}, \tilde{\sigma}^2) \;=\; T^{-1}\mathbf{I}\mathbf{A}^{-1}(\phi, \sigma^2) \;=\; \begin{bmatrix} (1-\phi^2)/T & 0 \\ 0 & 2\sigma^4 T \end{bmatrix} \tag{5.8}$$

- Also, notice carefully that the variance-covariance matrix is block diagonal indicating the MLE $'s$ of $\phi$ and the $\sigma^2$ are distributed independently and hence is identical for all orders of the autoregressive process.
- Off diagonal elements are zero which allows efficient sequential estimation of the $\phi$ and then conditional on $\tilde{\phi}$, $\sigma^2$ can be estimated
- If the null hypothesis is $\phi \;=\; 0$ then the variance is simply $1/T$ (and its standard error is $1/\sqrt{T}$). We have used this result in testing the autocorrelation function to identify the time series model.
- Finally we could estimate $Avar(\tilde{\phi})$ simply by taking the OLS estimate $\hat{\phi}$, calculating

  $$s^2 \;=\; \frac{1}{T}\sum_{t=2}^{T}(y_t - \hat{\phi}y_{t-1})^2$$

  and:

$$\widehat{AVAR}(\hat{\phi}) = \frac{s^2}{\sum_{t=2}^{T} y_t^2} \text{ so that } T\widehat{Avar}(\hat{\phi}) = p\lim \frac{Ts^2}{\sum_{t=2}^{T} y_t^2} = 1 - \phi^2. \tag{5.9}$$

### 5.2.2 The $AR(p)$ case

- The above results extend directly to $AR(p)$:

$$y_t \;=\; \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t \tag{5.10}$$

- Let

  $$A_T = \frac{1}{T-p}\sum_{t=p+1}^{T} X_t^T X_t$$

  and

  $$X_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p})$$

  then the OLS estimator (equivalent to the MLE) is

$$\tilde{\phi} = A_T^{-1}\frac{1}{T-p}\sum_{t=p+1}^{T} X_t^T y_t. \tag{5.11}$$

- **Subject to the stationarity requirements**:

$$T^{1/2}(\tilde{\phi} - \phi) \underset{\sim}{d} N(\mathbf{0}, A^{-1}\sigma^2) \tag{5.12}$$

where $A^{-1} = p\lim\ A_T$ .

- We may also show that

$$\tilde{\sigma}^2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\left(y_t - \tilde{\phi}_0 - \tilde{\phi}_1 y_{t-1} - \tilde{\phi}_2 y_{t-2} - \ldots - \tilde{\phi}_p y_{t-p}\right)^2 \tag{5.13}$$

is consistent for $\sigma^2$.

### 5.2.3 $MA(1)$

- The moving-average error process is estimated by maximum likelihood and therefore the same asymptotic results obtain.
- In view of the nonlinearity in estimation, the calculation of the variance covariance matrix is more complex.
- As indicated above, typically the inverse of the Hessian is employed. The $MA(1)$ case also shows an interesting parallel with the $AR(1)$ model

$$y_t \ = \ \epsilon_t + \theta\epsilon_{t-1} \tag{5.14}$$

- We have seen (equation 4.35) that the gradient for the $MA(1)$ process follows the first-order autoregressive process (letting $z_t \ = \ \frac{\partial \epsilon_t(Y;\theta)}{\partial\theta}$ ) :

$$z_t \ = \ -\theta z_{t-1} + \epsilon_{t-1}(Y;\theta) \tag{5.15}$$

- Hence using:

$$IA(\psi) \ = \ \sigma^{-2} p\lim T^{-1}\sum\left(\frac{\partial\epsilon_t}{\partial\theta}\right)^2 \ = \ \sigma^{-2} p\lim T^{-1}\sum z_t^2 \tag{5.16}$$

and

$$p\lim T^{-1}\sum z_t^2 \ = Var(z_t) = \frac{\sigma^2}{1-\theta^2}. \tag{5.17}$$

- Therefore

$$Avar(\tilde{\theta}) \ = \ \frac{1-\theta^2}{T} \tag{5.18}$$

- We see that the structure of the $Avar$ is the same for the $AR(1)$ and $MA(1)$.

- In fact, if we recall tests for serially correlated errors in the presence of lagged dependent variables based on the Lagrange Multiplier principle (Godfrey, 1978), have as an implicit alternative: $AR$ or $MA$ or ARMA .
- The test cannot discriminate between these alternatives. Higher order $MA$ processes follow the same arguments (see Fuller pp. 351).
- We may obtain the $Avar$ for the ARMA(1,1) process in an analogous manner (just as we could extend the estimation problem fashion). For more complete details see Harvey pp. 151-152.

## 5.3 Hypothesis Tests

- One advantage of being in the maximum likelihood framework is that the testing principles are well understood.
- Hence tests may be calculated using Wald, Likelihood ratio and Lagrange Multiplier principles.
- All our tests are asymptotic (central $\chi^2$ under the null hypothesis). The usual guidelines apply: namely, **we do the test procedure that is simple**r for the problem at hand.
- For example, in Chapter 6 where we discuss model building, we find that likelihood ratio criteria are often applied to determine the order of the process –be it autoregressive, moving average or mixed.
- Indeed, a common empirical strategy is to read $t - tests$ off the Hessian (for individual significance testing) and other kinds of tests are handled through likelihood ratio tests.
- The discussion below is in terms of testing linear restrictions..
- It should be obvious that **nonlinear restrictions** can easily be incorporated.

### 5.3.1 Testing Linear Restrictions

- Suppose we wish to test the linear restrictions:

$$H_0 : R\Psi \;=\; r \quad m \text{ restrictions} \tag{5.19}$$

  where $R$ is an $m \times n$ matrix (fixed and known) and $\Psi$ is a $n \times 1$ where $n$ is at most equal to $p + q + 1$ (the variance).
- Assume that $m$, the number of restrictions is strictly less than the number of parameters.
- The restrictions may be tested in three asymptotically equivalent (although not numerically) ways.

- In time series models it is common that exclusion restrictions are tested so that $r = 0$ and $R$ is appropriately defined

### 5.3.2 Likelihood Ratio Tests (LR)

- This is often the simplest way. We simply estimate the unrestricted model, denote the maximized value of the log **likelihood** as $\log L(\Omega)$.

- Next we estimate the restricted model (*i.e.* the restrictions are imposed) and obtain the restricted log likelihood $\log L(\omega)$.

- Then **asymptotically**

$$-2(\log L(\omega) - \log L(\Omega)) \sim \chi^2_m \text{ under } H_0. \tag{5.20}$$

### 5.3.3 Wald Tests (W)

- While the likelihood ratio test is usually the easiest, there are occasions in which the restricted model is difficult to estimate.

- For instance, in many dynamic models with forward looking agents, the restricted model is highly non-linear and difficult to estimate

- However, the unrestricted vector autoregressive model (VAR) and test the over-identifying restrictions:

$$W \;=\; (R\tilde{\Psi} - r)^T \left[R\; I(\tilde{\Psi})\; R^T\right]^{-1} \;(R\tilde{\Psi} - r) \sim \chi^2_m \text{ under } H_0 \tag{5.21}$$

  where we have evaluated the information matrix $I(\tilde{\Psi})$ at the **unrestricted** maximum likelihood estimates.

### 5.3.4 Lagrange Multiplier Tests (LM)

- There are occasions in which the restricted model is the simplest to estimate and test.

- For example, suppose that we have an $AR(p)$ process and we wish to know whether the error terms are correlated, say possibly due to a moving average error process.

- In this case, the restricted model can be estimated by OLS (recall the $AR$ model is linear) and the restrictions tested in the manner suggested by Godfrey (see Chapter 6)

- The general form of the $LM$ test is:

$$LM = \left(\frac{\partial logL}{\partial \Psi}\right)^T I(\tilde{\Psi})^{-1} \left(\frac{\partial logL}{\partial \Psi}\right) \sim \chi^2_m \text{ under } H_0 \tag{5.22}$$

  where $I(\tilde{\Psi})$ the information matrix and the $\left(\frac{\partial logL}{\partial \Psi}\right)$ are evaluated at the **restricted estimates**.

- For the test for serial correlation, we would regress the residuals on the lagged residuals of whatever order (say $q$) we wished to test and

$$LM = T \times R^2 \sim \chi^2_q \text{ under } H_0. \tag{5.23}$$

## 5.4 Nested Hypothesis Tests

- All three forms of these tests presume that the models are **nested** (the null hypothesis is a limiting case of the more general model)
- Hence, it is not possible with standard classical methods to test an $AR(1)$ against an $MA(1)$. These models are **nonnested**.
- On the other hand, an $AR(1)$ or and $MA(1)$ may be tested against the more general form ARMA(1,1). In the former case there are information criteria available to **select** one model specification over the other and there are **nonnested** hypothesis tests that may be applied.
- We discuss model selection methods in some detail in Chapter 6.

## 5.5 Warning About Testing Hypotheses

- One important item to keep in mind while reading this chapter is the difference between the **limiting distribution** (a large sample consideration) of a test or estimator and its **finite sample** properties.
- We may have a test that, in the limit is, say, $\chi^2_k$ under the null hypothesis, but for the sample size we have, it can be quite different.
- **All our theoretical test results are asymptotic (as $T \to \infty$).**
- In time series tests the general finite sample result is tests tend to be oversized, meaning that they reject the null more frequently than the nominal size
- Given that more of the tests tend to be exclusion restrictions (making the model smaller by omitting variables), this over-rejection leads to large models
- As a result, many applied researchers refrain from asymptotic comparisons against central $\chi^2$ and favour doing some finite sample checks
- Certainly in this context,. I would advise all tests be at the 1% level of significance since this at least raises the bar for rejection of the null

### 5.5.1 Checking the Finite Sample Distribution of a Test

- How do we know when our tests are biased? In general, we can't be sure.
- One way applied econometricians check the finite sample properties of the tests that they preform is to do a simple **Monte Carlo** experiment *calibrated* to their particular example.
- Another procedure is to use the **Bootstrap** which is a resampling technique (large sample justification is required) in the hopes if obtaining confidence intervals, p-values or whatever that are more reliable than standard asymptotic checks
- In the program boot_monte_int.do we illustrate both techniques using the $AR(1)$ model in the U.S. 1-month treasury bill

$$r_t = \alpha + \phi r_{t-1} + \epsilon_t$$

- Suppose we wish to answer the question how likely is it that we would obtain an estimate of $\phi \geq 1$ when the true value of $\phi$ is equal to our estimated value obtained from the actual data

**Monte Carlo**

- In a Monte Carlo example, you can generate data (using perhaps a normal distribution assumption and a random number generator) repeatedly under the null hypothesis (in this case use our estimated value of $\phi$) and then construct your statistic of interest.

1. Estimate the model using the actual data obtain an estimate of $\alpha, \phi$ and $\sigma^2$
2. With these estimates and the assumption of normality generate say $R = 1000$ replications data sets of with the sample size $T$ equal to the number of observations in your data set. (In general you need to *impose the null hypothesis*, but you can set all other unknown coefficients equal to the values you have estimated from the data)
3. We can look at the number of times we generate a $\phi \geq 1$
4. in the case of the U.S. interest data we see that this is unlikely (we note that the OLS coefficient tends to be biased downwards in the $AR(1)$ model)

- In general testing situation

1. If we are testing at the $\alpha$ level (say 1%), then we should expect to reject the null hypothesis $\alpha \times R$ (.01*1000=10) times.
2. We could construct a confidence interval for the number of rejections since this is a binomial distribution with independent draws.

3. If we find the number of rejections is excessive, then we have a **biased test**.

4. In circumstances in which we have biased tests, we can construct **size-adjusted critical values** by ordering our $R$ test statistics (denote test statistic by $S$) from smallest to largest, say $S_1 \dots S_R$ .

5. The size-adjusted critical values is obtained by taking the test statistic $S_{\alpha \times R}$ (the nearest integer). For $R = 1000$, $\alpha = .01$, this would mean that $S_{10}$ would be the size-adjusted critical value.

6. We would then look at the test statistic that we got from the data, say $S_{data}$ and see if it is large or smaller than our size-adjusted critical value.

7. If $S_{data} > S_{\alpha \times R}$ , we would reject the null hypothesis at the $\alpha$ level of significance. This would be a **size-adjusted test**.

**Bootstrapping**

- Bootstrapping is a resampling technique based on independent identically distributed draws
- With dependent data we typically do not have this assumption so some accomodation is required
- There have been two typical strategies
    1. Block Bootstrapping where data is resampled in blocks (variable of fixed block lengths) with the aim that with sufficiently large blocks the data is asymptotically independent
    2. Parametric Bootstrapping where the model is assumed to create a situation where say the error structure is *iid* and so the errors can be resampled and bootstrap replications of data can be constructed
- In the example we estimate the AR model and then resample from the residuals of the $AR(1)$
- In light of the $ARCH$ errors that are present, the *iid* assumption clearly is a problem

## 5.6 Confidence Intervals

- Confidence intervals may be calculated in the usual way using the asymptotic normal results earlier.
- Note that we have assumed stationarity conditions in all the maximum likelihood theory (as well as invertibility).

- It is certainly quite possible for our confidence intervals to extend into the nonstationary region.
- This is something we shall return to when we discuss nonstationary time series models.

## 5.7 A Simple Common Factor model

- In Chapter 4 we discussed how to estimate models by Kalman filtering and in this chapter we have investigated hypothesis tests for $MLE$
- In Stata $ARIMA$ models are estimated by $MLE$ but there is a class of models called dynamic factor models which are estimated via the Kalman filter
- In Stata using the $d$ factor command it is easy to estimate these kinds of models
- Variables (Monthly) $y_{it}, i = 1, 2, 3$ $t = 1967$ to most recent (logged difference for stationarity)
  - Industrial Production
  - Real Personal Income-transfer payments
  - Non Farm payroll employment
  - Average initial claims (not significant)

$$\begin{aligned} y_{it} &= \beta_i f_t + u_{it} \\ f_t &= \phi f_{t-1} + v_t \\ u_{it} &= \rho_i u_{it-1} + \epsilon_{it} \end{aligned}$$

- We think of $f_t$ as the unobservable component which we loosely call "economic activity" and the goal here is to estimate this as well as the unknown paramters under the assumption

$$Cov[f_t u_{is}] = 0 \qquad \forall i \text{ and } t \neq s$$

  and all errors are contemporaenously uncorrelated as well as zero cross corelation with normality and constant variances
- With these assumtions we may estimate this model and program *com.do* executes the commands
- If we assume that there is no persistence to the model ($\phi = \rho_i = 0 \forall i$ then this is identical to principal components
  - leading eigenvalue in the $y_{it}$ along with its eigenvector as a measure of the common component