# Applied Econometric Time Series
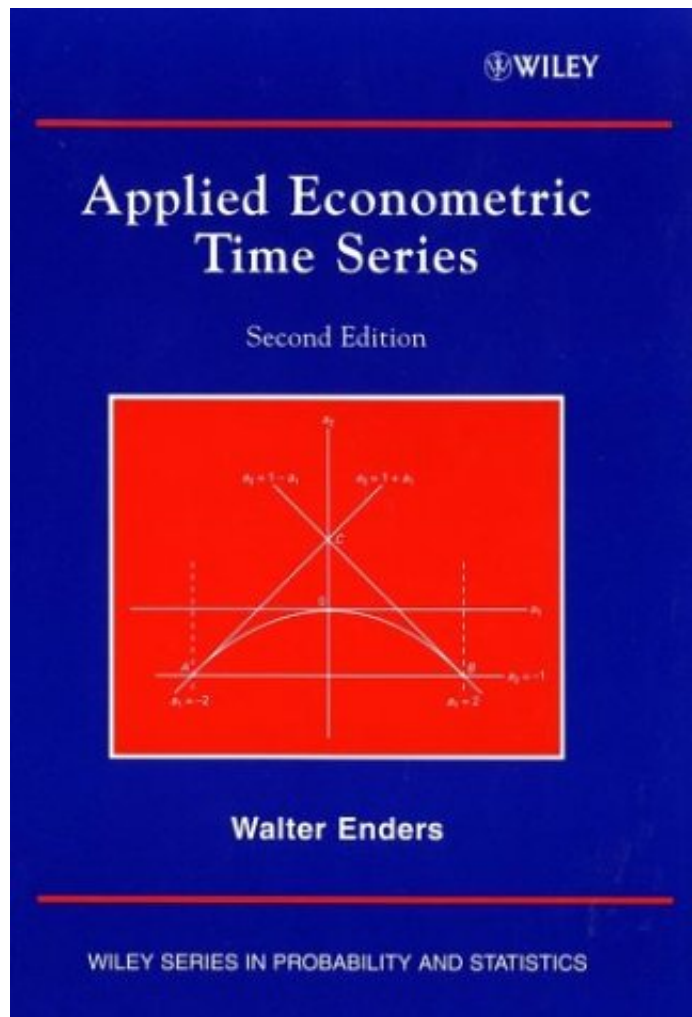## Book Summary

RK

November 10, 2013

**Abstract**

The purpose of this document is to summarize the main points of the book "Applied Econometric Time Series".

# 1   Difference Equations

The book starts off by stating,

> Time series econometrics is concerned with the estimation of difference equations containing stochastic components.

Hence the book naturally begins with a full-fledged chapter on difference equations. A few examples of difference equations are given such as Random walk model, Structural equation, Reduced form equation, Error correction model to show the reader that difference equations are everywhere in econometrics.

Any time series model indeed is trying to explain a univariate variable or a multivariate vector in terms of lagged values , lagged differences, exogenous variables, seasonality variables etc. The representative structure for the time series model is a difference equation. Any difference equation can be solved by repeated iteration, given an initial value. If the initial value is not given, it can be chosen in the form that involves infinite summation and the solution thus obtained by repeated iteration is just one of the many solutions that the difference equation can possess. However this method of *repeated iteration* breaks down for higher order difference equations. The chapter then talks about systematically finding the solutions to a difference equation using the following four steps :

1. Form the homogeneous equation and find all $n$ homogeneous solutions
2. Find a particular solution
3. Obtain a general solution as the sum of the particular solution and a linear combination of all homogeneous solutions
4. Eliminate the arbitrary constants by imposing the initial conditions on the general solution.

Thus the solutions to difference equations are usually written as a combination of *homogeneous solutions* and a *specific solution.* In all the algebraic jugglery that one needs to do for finding solutions, the thing that becomes important is the stability of the algebraic equations. The solutions to difference equations could remain stable or explode depending on the structure of the difference equation, i.e. the coefficients of the difference equation. The chapter then illustrates the difference equation machinery using *cobweb model* and derives the solution to a first-order linear difference equation with constant coefficients. In the process it shows the importance of *stability of the solution* and the relevance of *impulse response analysis.* The example makes a reader realize that building an econometric model might be an exercise that is taken up for various reasons. Well Forecasting is the obvious one. There are other aspects like granger causality analysis, instantaneous causality analysis, impulse response function analysis, that fall under the umbrella of structural analysis. These kind of analysis help a modeler in inferring the various relationships amongst various time series.

Solving a homogeneous difference equation involves writing out the *characteristic equation* and finding the characteristic roots. These roots decide the stability of the process. If all the roots are with in a unit circle, then the process is stable. If there are a few roots on the unit circle and rest are all with in the circle, then the process is called a unit root process or a process with order of integration $d$, where $d$ is the number of roots on the circle. If there are roots outside the circle, then the process explodes. These roots are nothing but the eigen values of a specific matrix that results out of a difference equation. One subtle point to note is that very often a similar statement is made about the roots of a reverse characteristic equation. The inverse characteristic equation is probably a natural way to write and hence an analogous statement for the stability or instability of the process is : if all the roots are outside the unit circle , then the process is stable. If there

are a few roots on the unit circle and rest are outside the unit circle, then the process is an unit root process. If there are roots with in the circle, then the process is explosive.

Finding a *particular solution* for a difference equation is often a matter of ingenuity and perseverance. The chapter cites some common difference equations and tricks to solve for particular solution. The chapter concludes by introducing two ways to solve for a particular solution, one is by the *method of undetermined coefficients* and second is via *Lag operators*. Lag operator method is intuitively appealing than the undetermined coefficient method.

# 2    Stationary Time Series Models

This chapter touches upon all the aspects of time series modeling that a typical undergraduate course would cover. It starts off with a basic representation of ARMA models, discusses stability of the various ARMA models. It then introduces ACF and PACF tools to get an idea of the underlying process. Box Pierce and Ljung and Box statistics for diagnostic testing are mentioned. In terms of model selection, the two most common measures AIC, SBC are highlighted where the former is effective in small samples whereas the latter is effective in large samples. SBC penalizes against over parametrization more than AIC. The appendix also mentions FPE(finite prediction error) criterion that seeks to minimize the one-step-ahead mean squared prediction error. A few examples are given where in two different ARMA processes are fit to the same dataset and the entire tool box containing ACF, PACF,Diagnostic tests, model selection criteria are used to select the best representative process. Box Jenkins model selection framework is introduced and the three main stages, i.e. identification stage, estimation stage and diagnostic checking, are shown via several examples.

The section on forecasts is probably the most interesting aspect of this chapter. Not many books highlight explicitly, the difference between the forecast based on known parameters and forecasts based on estimates. If you look at any time series model that is estimated, the estimated forecast error is more than the forecast error of a model with known parameters. However in most of the softwares that are available, you typically get confidence intervals based on the theoretical forecast variance rather than estimated forecast variance. In any case one can make an argument that if the sample size increases, the theoretical forecast error dominates the error component arising from the uncertainty of estimates.

The question relating to evaluating the forecast coming from competing models is answered well in this chapter. Two popular tests, Granger-Newbold test and Diebold-Mariano test are explained. The former overcomes the problem of contemporaneously correlated forecast errors. These tests have been mentioned because they relax the harsh restrictions of a typical forecast performance technique, i.e forecast errors have zero mean and are normally distributed, the forecast errors are serially and contemporaneously uncorrelated. The chapter ends with a discussion of addition of seasonality component to ARIMA models denoted by $\text{ARIMA}(p, d, q)(P, D, Q)$ where $p$ and $q$ are the number of nonseasonal ARMA components, $d$ is the number of nonseasonal differences, $P$ is the number of multiplicative autoregressive coefficients, $D$ is the number of seasonal differences, $Q$ is the number of multiplicative moving average coefficients and $s$ is the seasonal period.

# 3    Modeling Volatility

There is a certain notoriety associated with ARCH/GARCH models that are the topics of this chapter. These models have been the criticized by many people who claim the volatility modeling with gaussian models is too

naive in this highly complex and non linear world. May be that is the case. But who am I to pass a judgment on models which have made people get a Nobel Prize? I will just do my job of summarizing the models with out passing any value judgments on them. The first type of models presented are the ARCH models. All said and done, the story factor behind these models is solid, i.e. there are periods of homoscedasticity followed violent persistent shocks. One of the ways to model this kind of behavior is to build a conditional variance model and that's exactly what ARCH is. The conditional variance is modeled as a auto regressive process. GARCH is generalization of ARCH that includes an MA component for the conditional variance process. Infact these models are quite a hit in the academic community that there is an entire family of related models such as IGRACH, TGARCH, EGARCH etc. I think the problem with spending too much time with these models is that you unknowingly start believing that one of the family members should be the only ones representing the volatility of an instrument. As long as you are aware of that meta problem, I think its perfectly fine to have a working knowledge of these models.

# 4    Models with Trend

This chapter presents unit root tests sans all the complicated math that goes behind it. It starts off by defining a few models and explaining the difference between trend-stationary models and difference-stationary models. In the former the deterministic trend removal makes the series a stationary series whereas in the latter, a differencing of the series makes it a stationary series. One way to differentiate between the two is this : detrend the series and check the PACF, ACF of residuals for any thing fishy. If everything is fine, then it is likely to be a stationary process. However if residuals refuse to show any sane structure, then it is likely that it is a difference stationary process. To be more certain about it, fit a model to the differenced data then do residual diagnostics.

To answer the question more rigorously, the author introduces the idea of spurious regression. This happens when two integrated random variables are regressed against each other. The $t$ statistic and $F$ statistic for the coefficients are usually high. This does not mean that we must be sure about the coefficient values. Infact the reason for the test statistics taking high values is that the residuals are not stationary as is required for OLS. The derivation of the coefficient estimate in a closed form is not presented in the book. However the end result is presented, i.e. the test statistics are proportional to the square root of sample size. This means by merely taking a bigger sample, one can get a statistically significant estimate. So, one needs to be think through four cases that typically arise when you regress two series, let's say $\{y_t\}$ and $\{x_t\}$

1. $\{y_t\}$ and $\{x_t\}$ are stationary series - OLS is perfect and all the relevant principles from asymptotic theory is valid
2. $\{y_t\}$ and $\{x_t\}$ are integrated of different orders - Regression is meaningless
3. $\{y_t\}$ and $\{x_t\}$ are integrated of the same order and the residuals are non-stationary - Spurious regression
4. $\{y_t\}$ and $\{x_t\}$ are integrated of the same order and the residual is stationary - The variable are said to be cointegrated

The chapter deals with the univariate case where a $\{y_t\}$ is tested for the presence of unit root. The first thing that should be highlighted but has not been in this section is that,

> There is no unit root test available to check for the presence of one in a generalized stationary series. You have got to assume a certain data generating process and you can only check the null hypothesis that certain coefficients of the process that make the process a unit root process, take

up certain values. So, there is no one size fit all test.

The book assume the DGP as AR(1) and introduces Dickey-Fuller testing framework. When you run a regression between an integrated variable with its lags, the coefficients are a realization of non standard distributions, i.e. a Brownian motion functional. Hence somebody had to run simulations to give critical values for the coefficients and that job was done by Dickey-Fuller. So, whenever you see a table of critical values that help you test your hypothesis, you must thank the guys who took the effort to run simulations and publish these results for everyone to use. The plain vanilla Dickey-Fuller tests considers AR(1) in three forms. First form has a lagged values, the second form has lagged values and an intercept, the third form has a lagged value, intercept and linear time trend component. Critical value of $t$ statistics to test the parameter estimates are tabulated for each of the three forms. There are also critical values for statistics that test joint hypotheses on coefficients.

An interesting example is presented that shows the apparent dilemma that commonly occurs when analyzing time series with roots close to unity in absolute value. Unit root tests do not have much power in discriminating between characteristic roots close to unity and actual unit roots. Hence one needs to do two kinds of tests to be more certain about the process. First type of tests are the Dickey-Fuller types where the unit root is null hypothesis. Second type of tests are the ones where the null is stationary(KPSS test) and alternate has a unit root.

One has to note that Dickey-Fuller tests assume that the errors are independent and have a constant variance. The chapter then raises 6 pertinent questions and answers them systematically.

1. The DGP may contain both autoregressive and moving average components. We need to know how to conduct the test if the order of the moving average terms is unknown
   - This problem was cracked way back in 1984 where it was shown that an unknown $\text{ARIMA}(p, 1, q)$ can be well approximated by an $\text{ARIMA}(n, 1, 0)$ autoregression of order $n$ where $n \leq T^{1/3}$ where $T$ is the sample size.
2. We need to know the correct lag length of the AR process to be included
   - Too few lags means that the regression residuals do not behave like white noise processes and too many lags reduces the power of the test to reject the null of unit root.
   - This problem is solved by invoking the result from Sims, Stock and Watson(1990) paper: *Consider a regression equation containing a mixture of I(1) and I(0) variables such that the residuals are white noise. If the model is such that the coefficients of interest can be written as a coefficient on zero-mean stationary variables, then asymptotically , the OLS estimator converges to a normal distribution.* As such,a $t$-test is appropriate.
   - The way to solve the lag issue is to start with a higher lag and then reduce the lag until the appropriate lag appears statistically significant.
3. What if there are multiple roots in the characteristic equation ?
   - The solution is to perform Dickey-Fuller tests on successive differences of $\{y_t\}$
4. There might be roots that require first differences and others that necessitate seasonal differencing.
   - There are methods to distinguish between these two types of unit roots
5. What about structural breaks in the data that can impart an apparent trend to the data ?
   - The presence of structural break might make the unit root testing biased in favor of unit root. There is a nice example that goes to illustrate the reason behind it.
   - Phillip Perron's framework is suggested to remedy the situation. In fact the book shows an example

that economic variable that showed difference stationary behavior started showing trend stationary behavior in the presence of known structural breaks

- You can easily simulate a stationary series that contains a structural break and convince for yourself Dickey-Fuller tests are biased.

6. It might not be known whether an intercept and/or time trend belongs in the equation?

- Monte Carlo simulations have shown that the power of various Dickey-Fuller tests can be very low. These tests will too often indicate that a series contains unit root.
- It is important to use a regression equation that mimics the actual DGP. Inappropriately omitting the intercept or time trend can cause the power of the test to go to zero. On the other hand, extra regressors increase the critical values so that you may fail to reject the null of a unit root.
- The key problem is that the tests for unit roots are conditional on the presence of the deterministic regressors and tests for the presence of the deterministic regressors are conditional on the presence of a unit root.
- To crack this problem , the author shows that a result from the paper from Sims, Stock and Watson can be used. The result goes like this : *If the data-generating process contains any deterministic regressors (i.e., an intercept or a time trend) and the estimating equation contains these deterministic regressors, inference on all coefficients can be conducted using a t-test or an F-test.* This is because a test involving a single restriction across parameters with different rates of convergence is dominated asymptotically by the parameters with the slowest rates of convergence.(Read the book on time series by Hamilton to understand this statement better).

The models introduced in this book are of two types, first type are the ones where three is only trend component or only stationary component, the second type being that it contains both components. In the case of models that have both trend and stationary component, there is a need to decompose the series in to its components. Beveridge and Nelson show how to recover the trend and stationarity component from the data. I went through the entire procedure given in the book and found it rather tedious. In any case, state space modeling provides a much more elegant way to address this decomposition. The chapter ends with a section on panel unit root tests.

# 5  MultiEquation Time-Series Models

This chapter deals with multivariate time series. Instead of taking all the series at once and explaining the model, the chapter progresses step by step, or should I say model by model. It starts off with *Intervention analysis* which is a formal test of a change of mean for a time series. The intervention variable is assumed to be exogenous to the system and the whole point of analysis is to understand the effect of the observation variable on the time series. There is a subjectivity involved in choosing the type of intervention process. It could be a pulse function or gradually changing function or prolonged pulse function. Interesting examples like estimating the effects of metal detectors on Skyjacking and effect of Libyan bombings are given in this section.

The chapter then moves on to *transfer function model* that is a generalized version of intervention model. Here the exogenous variable is not constrained to follow a deterministic path, but is a stationary process. This kind of systematic explanation of model by model helps a reader understand about the approaches tried out before VAR was adopted. You assemble different exogenous processes in to one process and estimate the whole process. As far as estimating individual component processes were concerned, one could use the

standard Box-Jenkins methodology. However for estimating the coefficients of the final equation, it was more art and a lot of subjectivity was involved. In this context, the author says

> Experienced econometricians would agree that the procedure is a blend of skill, art and perseverance that is developed through practice.

There is a nice case study involving transfer function that analyses the impact of terrorist attacks on tourism industry in Italy. Examples like these makes this book a very interesting read. These examples serve as a anchor points for remembering the important aspects of various models.

Out of the many problems with the transfer function is the assumption of NO FEEDBACK. A simple example of "thermostat and room temperature" is given in the book to explain "reverse causality". In economic variable scenario, most of the variables are always in a feedback loop. Transfer function modeling assumes all the subprocesses are independent and hence it is limited in its usage.

This set the stage for the evolution of the next type of model, VAR(vector auto regressive processes). My exposure to VAR modeling was via the Standard VAR(p) representation. This book showed me that there is another form of VAR(p), i.e. structural VAR that is the focus of an analyst. Standard VAR(p) is a transformed version of Structural VAR and is a computational convenience . Going from Structural VAR to Standard VAR means reducing the number of parameters in the model and hence there is an IDENTIFICATION PROBLEM. You can estimate the parameters of the Standard VAR but to map it back uniquely to Structural VAR would not be possible unless you impose restrictions on the error structure of the variables involved. The link between the two kinds of VARs are presented at the beginning of the chapter so that the reader knows that there is going to be some subjectivity in choosing the error structure. Again this matters less if the researcher is only interested in forecasting.

Stability issues for VAR(p) models are discussed. Frankly I felt this aspect is wonderfully dealt in Helmut Lutkepohol's book on multiple time series. In fact I came across VAR in multiple places and the book that gave me a solid understanding of VAR from a math perspective was the book by Lutkepohl. However the intuition and application of such models is what this chapter stands out for. Also the math behind VAR is slightly daunting with the `vec` operators, matrix differentials used all over the place. Hence one can consider this chapter as a gentle introduction to VAR modeling. As far as estimation is concerned, one can use OLS for each equation and estimate the parameters. If there are varying lag lengths, SUR can be explored. The more I think of these models, the more I realize that all these models were constructed to give convenient answers to questions like "If a unit shock is applied to this variable, what happens to the system?". Needless to say any answer provided to such a question at least in the economic variable scenario is merely a story. It is hard to capture a nonlinear world in a linear form. Nate Silver's book on signal vs. noise has a chapter on forecasting performance on economic variables. I guess books such as Signal vs. Noise help us in not getting carried away by notions such as impulse response functions etc. Well, all these concepts such IR functions are good on paper but how well they stand up to economic realities is a big question. In any case the author has to do the job of presenting the literature and my job is summarizing it. So, let me go ahead.

Once the estimation is done, one might be interested in doing a structural analysis. The chapter presents the definition of impulse response functions,the way to compute them and estimate their uncertainties. Whenever you make a forecast using VAR, one can decompose the forecast error by chunking it and attributing it to the variables in the system. One needs to use Moving average representation of Standard VAR model to analyze forecast error variance decomposition. IR functions and Forecast error variance decomposition fall under the

category of innovation accounting.

The other kinds of structural analysis involved are hypothesis testing,granger causality, tests with non stationary variables, etc. Each of these topics is intuitively explained with a few examples. For testing hypothesis, the LR ratio test is suggested. If all VAR variables are stationary, then testing granger causality can be done via the usual F test route. When testing with non stationary variables, the book presents the result from Sims, Stock and Watson paper: *If the coefficient of interest can be written as a coefficient on a stationary variable, then a t-test is appropriate.* You can take this result at face value and do all the hypothesis testing. However for the curious ones, it pays to understand the statement better. My first exposure to such a statement came while reading Hamilton's book on time series. I realized that when you mix stationary and non stationary variables in to one regression equation, the concept of rate of convergence becomes very important. I think this is where the whole field of time series math differs from the usual regression models. For every variable that you include in the model, you have to think about the rate of convergence and in some cases, your standard OLS regression estimates are good enough despite having non stationary variables in the equation. For more clarity on this, I think its better to read chapter 8 from Hamilton's book. In the context of VAR, the author presents a model to capture the relationship between terrorism and tourism in Spain.
What's the flip side of VAR ?

> The VAR approach has been criticized as being devoid of any economic content. The sole role of the economist is to suggest the appropriate variables to be included in VAR. From that point, the procedure is almost mechanical. Since there is so little economic input in to VAR, it should not be surprising that there is little economic content in the results. Of course, innovation accounting does require an ordering of the variables, but the selection of the ordering is generally ad hoc.

There are some examples given that impose conditions on the ordering of the variables in VAR to generate impulse response functions. Somehow after reading all those examples, given that I am highly skeptical about medium to long range economic system analysis, I feel most of the literature on VAR was useful to publish papers and nothing else. Whatever fancy decompositions that one reads, I think they fall flat in terms of explaining macroeconomic realities. After all the basic model runs on gaussian errors and you are trying to predict stuff in a non-linear world. Ok, if there is somebody publishing GDP on a daily basis, then may be sheer magnitude of data, some averaging takes place and one can use gaussian models. But applying such models for quarterly data, annual data seems a futile exercise.

In any case, my interest in going through this book was to read some general stuff on cointegration and VECM. I felt the treatment of VECM model estimation in Lutkepohl was extremely rigorous to the point that I realized that I had to take a break and revisit the stuff at a later point in time.

# 6   Cointegration and Error Correction Models

This section is probably the most relevant to developing trading strategies. "Cointegration" is a term that is often heard in the context of pairs trading. Broadly the term captures the situation where a linear combination of similar ordered integrated series exhibits a lower order integrated series. More precisely, components of a vector time series are said to be cointegrated of order $CI(d, b)$ , if all the components of the vector are integrated of order $d$ and there exists a linear combination of the vector components that is integrated of order $d - b$. This definition is slightly tweaked depending in different books. For example Helmut Lutkepohl tweaks this

definition a little bit so that stationary series can also be included in a cointegrated system. There are four important points that needs to noted about the definition

- The emphasis is on *linear* combination. Theoretically one can think of nonlinear combination. But that's an area of active research and not dealt in this book. Also, cointegrating vector is not unique. But it can be made unique by normalizing the vector.
- Even though the original definition is restricted to series of order $d$, it is perfectly possible that only a certain set of variables are integrated. This book introduced me to the concept of "Multicointegration", that refers to a situation where there is a equilibrium relationship between groups of variables that are integrated of different orders. Is there ever a multicointegration amongst a set of stocks in the real world? I don't know. Some one would have done some research on this aspect.
- In a set of $n$ vectors there can be as high as $n-1$ cointegrating vectors
- Most of the cointegration literature focuses on the case in which each of the component has a single unit root because there are hardly any variables that are integrated of an order higher than unity.

In simple terms, any cointegrated system has a common stochastic term. The job of researcher is to tease out that relationship. The crucial insight that helps in doing this is the connection between a cointegrated system and error correction model. What's an error correction model ? It looks similar to VAR but with an additional lagged level variable in the equation. So, if one goes ahead and builds a VAR model with differenced data of set of $I(1)$ processes, there could be a risk of misspecification. Why ? If there are a subset of variables that are cointegrated, the correct model to use is the error correction model rather than a VAR model. The advantage of using an error correction model is that one can tease out the speed of adjustment parameters, that help us understand the way in which each of individual series responds to the deviations from the common stochastic trend.

The chapter also explains a crucial connection between VAR model and error correction model by casting a simple bivariate VAR(1) in to two univariate second order difference equations that have the same characteristic equation. The eigen values of the characteristic equations cannot be some independent values if the system is cointegrated. It is shown that one of the eigen values has to be one and other less than 0. This ensures that the bivariate VAR(1) in level variables are cointegrated. This means that there are restrictions on the coefficients that make the system a cointegrated system. So at once the reader understands that for any set of I(1) variables, the ECM and cointegration are equivalent and that the rank of the matrix of coefficients in ECM model can be taken as the number of cointegrating relations in the system. The chapter generalizes the findings to a n variable system.

There are two standard methods of finding the cointegration amongst variables. One is the Engle Granger method and the second is the Johansen test. The former is computationally easy but has some problems. In the latter, the math(canonical correlation analysis) is a little challenging but the reward for the slog is that you get a consistent system. In the Engle Granger case, all you do is this : At first you check whether the series are $I(1)$ . In the second step, you regress two I(1) series, the $t$ statistics are meaningless as nobody has told us that the variable on the rhs is the independent variable and the one on lhs is the dependent variable. All one can do with such a regression is to use the estimated residuals and check for stationarity. The important thing to realize is that you can't use Dickey-Fuller critical values to do a null hypothesis testing of unit root process ? Why ? The reason being that the residuals are estimated unlike the case of a unit root testing of a given level variable. Hence there are some kind souls who have tabulated the critical values for such estimated residuals and hence one can just go ahead and use them. If you use R, the author would have taken care of

this and you can just invoke a function. Ok, the logical step after you realize that the system is cointegrated is to build a VECM to get a sense of the rate of adjustment of series, granger causality etc. The chapter also has an example where an analysis of a multicointegration system is presented.

Engle Granger procedure, though easy to implement, has a few defects.

- The estimation of the long-run equilibrium regression requires that the researcher place one variable as regressand and use the other variables as regressors. In large sample systems the analysis of residuals is independent on what variable is chosen as regressand. However for small samples, one can face a situation where one regression indicates the variables are cointegrated whereas reversing the equation indicates the variables are not cointegrated. This problem is compounded in a three variable or multi variable system.
- Engle Granger relies on two steps. This means any error in the first step will make the second step meaningless

The chapter introduces Johansen procedure to remedy the above defects. Intuitively the procedure is a multivariate generalization of Dickey-Fuller test. The math behind requires some sweat from the reader who has never been exposed to canonical correlation analysis. The output of the procedure is basically two test statistics based on the eigen values of the coefficient matrix appearing in the multivariate Dickey-Fuller test. Thanks to Johansen's work, there are critical values that are provided to infer the number of cointegrating relations in a system. The other beauty of Johansen procedure is that the normalized eigen vectors actually provide you with a set of equilibrium relations. The other goody that you get by understanding Johansen procedure is the way to do hypothesis testing on a set on the cointegrating vector. The good thing about the chapter and this hold for the entire book is that each concept is immediately followed by a relevant case study from econometrics. This makes the reader motivated in understanding stuff from the book. It's not some abstract math/stats that is being discussed but a budding econometrician can actually use these concepts in his work and develop his/her own theories. The last chapter of the book is on Nonlinear time series, which I have skipped reading for now. May be I will go over it some other day, some other time.

# Takeaway

This book falls some where between a cookbook for econometric techniques and a full fledged encyclopedia that covers all the important aspects of econometrics. Hence this book would be daunting for an absolute beginner but a light read for someone who understands the math behind the models. I think it is better to get the math straightened out in one's head before venturing in to this book. In that way, one can appreciate the intuition behind the various econometric models and there will be many "aha" moments along the way.