Dynamic Predication Model for Integrated Series and Application

Yuan Jia-hai[1], Zhao Zhi[2], Xiong Min-peng[1]
1 School of Business Administration North China Electric Power University
2 Graduate School North China Electric Power University
Changping District Beijing 102206 China

Abstract:- In the paper a predication model for integrated series is proposed. Granger causality analysis is deployed first for finding out the cointegrated series for the interested series. Then granger causality information is used for the identification of the prediction model and designing of prediction process. A case study for electricity consumption modeling for China is studied to validate the viability of the method proposed in the paper.
Key words: *Granger causality;Cointegration series; Dynamic prediction; Electricity demand*

1 Introduction

Co-integrated theory has been used extensively for the study of relationship among integrated series.According to Engle and Granger (1978), a linear combination of two or more non-stationary series (with the same order of integration) may be stationary. If such a stationary linear combination exists, the series are considered to be cointegrated and long run equilibrium relationships exist. Incorporating these cointegrated properties, an error correction (ECM) could be constructed to test for Granger causation of the series in at least one direction.

Since the use of ECM requires the series to be cointegrated with the same order, it is essential to first test the series for stationarity and cointegration. A series is said to be nonstationary if it has non-constant mean, variance, and autocovariance over time. If a nonstatioary series has to be differenced d times to become stationary, then it is said to be integrated of order *d*: i.e. *I(d)*.

If two series are cointegrated, how can we proceed to predict for the series? In cointegrated theory the problem is answered implicitly and the error correction model can be used to dynamically predict all the series in the system with further information except for the sample to establish the model. However, at least two questions arise for the treatment. First, for how many periods we can assuredly make the prediction for the series? In other way, when the series ceases to be cointegrated or when there is shift in the cointegration relationship, the prediction will not be credible anymore. However, in cointegration theory it does deal with this problem explicitly and it seem as if that the model can be applied for prediction for any long time span. Another flaw with cointegration theory is that it does not utilize the information of granger cause among series fully. In fact, the direction of granger cause is of vital importance for the identification of prediction model. However in error correction model, by simultaneous modeling the very information is negligent omitted.

2 Granger Causality Test in Error Correction Model

When both series are integrated of the same order, we can proceed to examine for the presence of cointegration. The Johansen Maximum likelihood procedures are used for the test (Johansen and Juselius, 1990). Any long-tem cointegrating relationship found between the series will contribute an additional error-correction term to the ECM. The Johansen procedure is a vector autogressive (VAR) based test on restriction imposed by cointegration in the unrestricted VAR. The null hypothesis in consideration is Ho, that there are a different number of cointegration relationship, against H1, that all series in the VAR are stationary. The ECM used in this paper is specified as follows:

$$\Delta y_t = \alpha_2 + \beta_2 ECM_{t-1} + \sum_i \alpha_{21}(i)\Delta y_{t-i} + \sum_i \alpha_{22}(i)\Delta x_{t-i} + \varepsilon_{2t} \quad (1)$$

$$\Delta x_t = \alpha_1 + \beta_1 ECM_{t-1} + \sum_i \alpha_{11}(i)\Delta y_{t-i} + \sum_i \alpha_{12}(i)\Delta x_{t-i} + \varepsilon_{1t} \quad (2)$$

Using for test the relationship between electricity consumption and economic growth, for example, where Yt and Xt represent natural logarithms of real GDP and electricity consumption, respectively, and (ΔY_t, ΔX_t) are the differences in these variables that capture their short-run disturbances, $\varepsilon_{1t}, \varepsilon_{2t}$ are the serially uncorrelated error terms, and ECT_{t-1} is the error-correction term (ECT), which is derived from the long-run cointegration relationship and measures the magnitude of the past disequilibrium.

In each equation, change in the endogenous variable is caused not only by their lags, but also by the previous period's disequilibrium in level, i.e. ECT_{t-1}. Given such a specification, the presence of short and long-run causality could be tested. Consider Eq.(1), if the estimated coefficients on lagged values of electricity consumption are statistically significant, then the implication is that the electricity consumption Granger causes real GDP in the short-run. On the other hand, long-run causality can be found by testing the significance of the past disequilibrium term.

The null hypothesis of the F test is:

H0 $\alpha_{22}(i) = 0, i = 1,2\cdots p$, and the statistics is

$$F = \frac{(RSS_R - RSS_0)/J}{RSS_0/(T-K)} \sim F(J, T-K) \quad (3)$$

3 Designing Principle of Integrated Series Prediction Model

When discussing with the prediction of integrated series in the framework of time series model, the first question to be asked is whether some other series should be added in the prediction model or only the series is used for predict itself. The existence of integration relationship is the foundation for adding other explanation series. Further more, the Granger causality provide information for the order among series and therefore provide important information on considering correlated factors.

To simplify and with losing generality, we start with two variable model to expatiate our idea. The relationship between two variables is fundamental in that the most widely studied relationship in economics is relationship between two variables. And of course, our analysis below is easily expanded to multi-variable model.

Step one: Establishment of the cointegration equation

Suppose that we are to consider the prediction of integrated series, X_t. Suppose also that we have already found another series to be cointegrated with it. Cointegration analysis is able to establish the long-term relationship between them.

Step two: Test of Granger causality relationship

We can establish error correction model for the two variables according to equation 1 and 2. Then equation 3 is used to test for the direction of Granger causality between them. According the Granger theorem, at least one direction Granger causality relationship exists for cointegrated series. So, in sum there may be the following three possibilities:

Possibility one: Xt is the unilateral Granger causality of Yt.

Possibility two: Yt is the unilateral Granger causality of Xt.

Possibility three: Xt and Yt is bilaterally Granger caused.

Step three: Dynamic prediction of Xt

As to possibility one, it is obvious that $P(Y_t|I_T)>P(Y_t|Y_{t-1},Y_{t-2},\ldots)$. That is, we can establish the one step ahead prediction model for Yt and get the prediction of $Y_{t\ 1}$. Then, though Yt is not the granger cause of Xt, however because of the cointegration relationship between them, it is obvious that $P(X_t|X_{t-1},X_{t-2},\ldots;Y_t)>P(X_t|X_{t-1},X_{t-2},\ldots)$ which implies that we can predict Xt+1 according to the prediction of $Y_{t\ 1}$, to enhance the prediction accuracy of $X_{t\ 1}$.

For possibility two, it is obvious that $P(X_t|I_T)>P(X_t|X_{t-1},X_{t-2},\ldots)$, we can predict Xt+1 directly. Then because of the cointegrated relationship,
$P(Y_t|Y_{t-1},Y_{t-2},\ldots;X_t)>P(Y_t|Y_{t-1},Y_{t-2},\ldots)$. We can use the prediction of $X_{t\ 1}$ to predict $Y_{t\ 1}$, and then proceed to predict $X_{t\ 2.}$

As for possibility three, it is obvious that

$P(X_t|I_T)>P(X_t|X_{t-1},X_{t-2},\ldots)$,

$P(Y_t|I_T)>P(Y_t|Y_{t-1},Y_{t-2},\ldots)$,

We can establish error correction model according the information set to simultaneously predict $(X_{t\ 1},Y_{t\ 1})$.

Step four: Updating the series with one step ahead prediction and proceeding with cointegration test

Test for the cointegration relationship with the updating series. If the cointegration relationship keep as before, proceed with dynamic prediction for the next period; else, establish new cointegration equation and go to step 2.

Step five: cease.

4 Case Study for Electricity Consumption Prediction in China

4.1 test of Granger causality

In this case study, we are interested in the prediction of electricity consumption for China and we use the electricity consumption data for China during period of 1978 to 2004. According our study in (Yuan Jiahai et al. 2006) we conclude that electricity and GDP are cointegrated and there exists unilateral Granger cause running from electricity consumption to GDP growth.

4.2 Establishment of prediction model

According to the granger cause analysis, there exist unilateral cause running from electricity consumption to GDP for China during 1978 to 2004. Based on our analysis in section 3, to predict electricity consumption in the future, we should first make prediction for GDP. To show the difference of adding electricity consumption in the model and without electricity consumption, we establish two GDP prediction model and compare them in terms of prediction accuracy. According to the generality of econometrics, we choose five indexes as RMSE, MAS, MAPE, theil coefficient and difference ratio. All the indexes should be near to zero when the model fits well. If can be seen from table 3 that model 1 (incorporating GDP) is systematically better than model 2. We can get the same judgment also from Figure 1 in that model 1 fits better than model 2 in sampling period.

GDP prediction Model 1

$$Lgdp = 1.5049 * Lgdp(-1) - 0.6426 * Lgdp(-2) \\ + 0.168 * Lelec(-1)$$

GDP prediction Model 2

$$Lgdp = 1.578 * Lgdp(-1) - 0.574 * Lgdp(-2)$$

Tab. 1 Electricity consumption prediction model incorporating GDP data

Variable	Regressing coefficient	Standard error	T-statistics	P-value
LGDP(-1)	1.504932	0.182404	8.250519	0
LGDP(-2)	-0.642664	0.182552	-3.520435	0.002
LELEC(-1)	0.168959	0.09412	1.795144496	0.16
R-2	0.998806	Mean dependent var	10.62574	

Adj. R-2	0.998693	S.D. dependent var	0.659188
Regressive SE	0.023833	AIC information	-4.519032
SRS	0.011928	SC information	-4.371775
Log likelihood ratio	57.22838	D-W statistics	1.554911

Tab. 2 Electricity consumption prediction model without considering GDP data

Variable	Regressing coefficient	Standard error	T-statistics	P-value
LGDP(-1)	1.578197	0.175176	9.009235	0
LGDP(-2)	-0.574537	0.176669	-3.252053	0.0037
R-2	0.998716	Mean dependent var	10.62574	
Adj. R-2	0.998658	S.D. dependent var	0.659188	
Regressive SE	0.024149	AIC information	-4.529513	
SRS	0.01283	SC information	-4.431342	
Log likelihood ratio	56.35416	D-W statistics	1.503025	

Tab. 3 Comparison of two GDP prediction model in fitting criterion

Item	GDP Model 1	GDP Model 2	Comparing criterion
RMSE	0.0568	0.0998	0
MAS	0.049	0.086	0
MAPE	0.555	0.8	0
Theil coefficient	0.003	0.0046	0
Difference ratio	0.244	0.6698	0

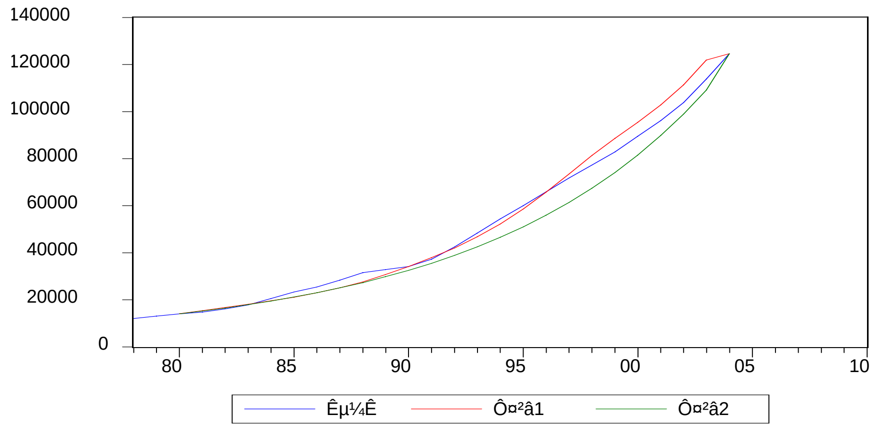

Fig 1 Comparison of two GDP prediction model in sampling accuracy

We proceed to establish the prediction model for electricity consumption. Accordingly, we have a model considering GDP and a model without GDP to compare the prediction accuracy of the two model.

Electricity consumption prediction model 1

$$Lelec = -0.117 + 1.48 * Lelec(-1) - 0.61 * Lelec(-2) + 0.127 * Lgdp$$

Electricity consumption prediction model 2

$$Lelec = 1.599 * Lelec(-1) - 0.595 * Lelec(-2)$$

According to the model, model 1 is systematically better than model 2, which implies that, we can make the prediction of GDP by the former two period of GDP and former one period of electricity consumption, and then proceed the prediction of electricity consumption by former two periods of electricity consumption and current period (predicted) GDP data. With the updating series, if the cointegration relationship does not changed, we can go on predicting the electricity consumption in that way.

Tab 4 Electricity consumption model considering GDP data

Variable	Regressing coefficient	Standard error	T-statistics	P-value
C	-0.11729	0.074629	-1.571633	0.1317
LELEC(-1)	1.480457	0.211371	7.004081	0
LELEC(-2)	-0.617604	0.190052	-3.249652	0.004
LGDP	0.127347	0.087978	1.447491	0.1633
R-2	0.998846	Mean dependent var	8.860186	
Adj.R-2	0.998673	S.D. dependent var	0.571051	
Regressive SE	0.020803	AIC information	-4.756467	
SRS	0.008655	SC information	-4.560124	
Log likelihood ratio	61.0776	F statistics	5770.586	
D-W Statistics	1.25001	P-value	0	

Tab 5 Electricity consumption model without considering GDP data

Variable	Regressing coefficient	Standard error	T-statistics	P-value
LELEC(-1)	1.599252	0.193177	8.278704	0
LELEC(-2)	-0.595409	0.194914	-3.054728	0.0058
R-2	0.998659	Mean dependent var	8.860186	
Adj. R-2	0.998598	S.D. dependent var	0.571051	
Regressive SE	0.021383	AIC information	-4.772787	
SRS	0.010059	SC information	-4.674616	
Log likelihood ratio	59.27345	D-W Statistics	1.200045	

Tab 6 Comparison of two electricity consumption prediction model in fitting criterion

Item	Electricity model 1	Electricity model 2	Comparing criterion
RMSE	0.0559	0.11	0
MAS	0.041	0.108	0
MAPE	0.45	1.12	0
Theil coefficient	0.003	0.006	0
Difference ratio	0.03	0.83	0

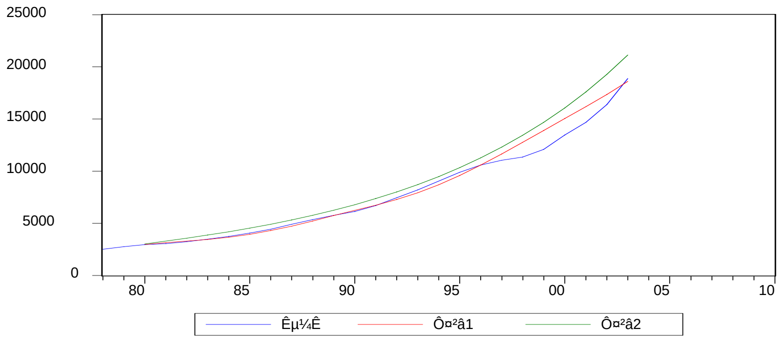

Fig 2 Comparison of two electricity consumption prediction models in sampling accuracy

5 Concluding Remarks

In the paper, we propose a novel dynamics prediction model for integrated series. Our model can utilize the full information of cointegration and Granger causality to establish the prediction model and make the dynamics evolution. A rather simple case study in the paper provides a preliminary proof for the feasibility of our method.

This is only a preliminary study on the road. Some questions still remain unanswered. For example, how to expand the model to multi-variable model? If more than one series are found to be cointegrated with the interested series, how can we choose among them? What is the judgment criterion for choosing among them? All these are our direction of future study.

References:

[1] National Bureau of Statistics of China. China Statistical Yearbook 2005 [M]. China Statistics Press, 2006.

[2] Kraft, J., Kraft, A.. On the relationship between energy and GNP [J]. Journal of Energy Development, 1978, 3, 401-403.

[3] Akarca, A.T., Long, T.V.,. On the relationship between energy and GNP: a re-examination [J]. Journal of Energy Development, 1980, 5, 326-331.

[4] Engle, R.F., and C.W.J. Granger,. Cointegration and Error Correction: Representation, Estimation and Testing [J]. Econometrica, 1987, 55, 251-276.

[5] R.F Engle. Cointegration, Causality and Forecasting [M]. Oxford University Press, 1999.

[6] R.F. Engler and C.W.J Granger. Cointegration and Error Correction: Representation, Estimation and Testing [J]. Econometrica, 1987, 55, 251-276.

[7] S. Johansen and K. Juselius. Maximum Likelihood Estimation and Inference on Cointegration with Applications to the demand for money [J]. Oxford Bulletion of Ecomomise and Statistics, 1990, 52, 169-210.

[8] Yuan Jiahai, Wang Jing ,Hu Zhaoguang Cointegration and Co-feature Analysis on Electricity Consumption and Economic Growth in China WSEAS Transaction on Systems, June 2006, 5(6):1396-1400

[9] Yuan Jiahai, Wang Jing, Hu Zhaoguang.Electricity consumption and economic growth in China: cointegration and co-feature analysis.5th WSEAS Int. Conf. on Applied Computer Science (ACOS '06). Hangzhou China, April 2006