Statistical Inference of Nonlinear Granger Causality: A Semiparametric Time Series Regression Analysis

Sooyoung Lee

Thesis submitted for the degree of
Master of Philosophy
in
Statistics
at
The University of Adelaide
(Faculty of Engineering, Computer and Mathematical Sciences)

School of Mathematical Sciences

October 2, 2012
Declaration

I, Sooyoung Lee certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University’s digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed by student:

Date:

Signed by supervisor:

Date:
Signed Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

SIGNED: ...................... DATE: ......................
Acknowledgements

First of all, I am extremely thankful to my supervisor, A/Professor Zudi Lu. This thesis would not have been possible without his encouragement, guidance and support during my study. I also thank my co-supervisor, A/Professor Andrew Metcalfe, and the postgraduate coordinators, Dr Trent Mattner and A/Professor Matthew Roughan whose supports were greatly needed and deeply appreciated. I sincerely thank University of Adelaide. Without the scholarship for 2 years, it could be hard to conduct my thesis. I am also grateful to all the students and staff in the School of Mathematical Sciences for their encouragement during my study. Most importantly, I would like to express my heart-felt gratitude to my husband and two children. I have to give a special mention for the support given by my husband, Dr. Seungmoon Choi.
Abstract

Since the seminal work of Granger (1969), Granger causality has become a useful concept and tool in the study of the dynamic linkages between economic variables and to explore whether or not an economic variable helps forecast another one. Researchers have suggested a variety of methods to test the existence of Granger causality in the literature. In particular, linear Granger causality testing has been remarkably developed; (see, for example, Toda & Philips (1993), Sims, Stock & Watson (1990), Geweke (1982), Hosoya (1991) and Hidalgo (2000)). However, in practice, the real economic relationship between different variables may often be nonlinear. Hiemstra & Jones (1994) and Nishiyama, Hitomi, Kawasaki & Jeong (2011) recently proposed different methods to test the existence of any non-linear Granger causality between a pair of economic variables under an $\alpha$-mixing framework of data generating process. Their methods are general with nonparametric features, which however suffer from curse of dimensionality when high lag orders need to be taken into consideration in applications.

In this thesis, the main objective is to develop a class of semiparametric time series regression models that are of partially linear structures, with statistical theory established under a more general framework of near epoch dependent (NED) data generating processes, which can be easily used in exploring nonlinear Granger causality. This general NED framework, which takes the $\alpha$-mixing one as a very special case, is popular in nonlinear econometric analysis. The reasons why we will adopt such a semi-parametric model structure for time series regression under the general NED framework are not only because it is a natural extension of the structure of lin-
ear Granger causality analysis but also because it enables us to estimate and hence understand the likely structure of the non-linear Granger-causality if it exists. To our best knowledge, in the literature, this is still an early effort that seeks the causality structure in nonlinear Granger-causality beyond the testing of its existence. Furthermore, semiparametric structures help to reduce the curse of dimensionality that purely nonparametric methods suffer from in Granger-causality testing. We study the semiparametric regression models under a more general framework of NED time series processes, which include, for example, the popular ARMA\((p,q)\)-GARCH\((r,m)\) model in financial econometrics which is hard to be shown to be \(\alpha\)-mixing except in some very special cases.

By using the idea of Robinson (1988) and the theory developed in Lu & Linton (2007) under NED, we can construct the estimators of both the parameters and the unknown function in our model. We have also established asymptotic theory for these estimators under NED. The estimated unknown functional part and its confidence interval can tell us useful information on whether or not there exists a nonlinear Granger causality between the variables, and moreover, we can find the functional form of the nonlinear Granger causality.

In order to examine the finite-sample performance of the proposed estimators for our model under study, besides the developed large-sample theory, we have also conducted Monte Carlo simulation studies. The simulation results clearly demonstrate that we can estimate both the parameters and the unknown function rather accurately under moderate sample sizes.

Finally, we have empirically applied the proposed methodology to examining the existence or effects of nonlinear Granger-causality between each pair of the financial markets involving Australia and the USA, UK and China which are closely related to Australia in economy. Weekly return data are used to avoid the possible market
Interestingly, we have found that there is a strongly nonlinear Granger causality from the UK FTSE100 to the Australian ASX200, and some linear or nonlinear Granger causality from the USA S&P500 to the Australian ASX200, as well as some negatively linear Granger causality from the China SSE index to the Australian ASX200. From these results, we can fairly say that the USA, UK and Chinese stock markets have close linkages with and impacts on the Australian market, with the stock prices from these foreign countries 1 week ago can help to predict current Australian stock market behavior. We have also found that there is some linear Granger causality from the UK FTSE100 and the USA S&P500 to the China SSE index, respectively, but we cannot see the Granger-causality from the Australian ASX200 to the China SSE. We either can not find apparent evidence of the Granger causalities from the other countries to the UK or the USA. In addition, we have examined the lag 2 Granger-causality effect between each pair of these markets, and only find that the Chinese stock price 2 weeks ago appears to help predict this week’s Australian stock price.

In summary, the main contributions of this thesis are as follows:

- We have suggested semiparametric time series regression models of partially linear structures to examine possibly nonlinear Granger causality, with methods to estimate both the parameters and the unknown nonparametric function proposed.

- We have established the consistency and asymptotic normality for the estimators under a more general, popular data generating framework of near epoch dependence.

- Monte Carlo simulation studies reveal that the proposed methodology works well for both linear and nonlinear functional forms of Granger causality in
finite samples.

- Interesting empirical applications find that there are clear dynamic linkages of the Australian stock market impacted by the USA, UK and Chinese markets.
Contents

Declaration i

Signed Statement ii

Acknowledgements iii

Abstract iv

1 Background and Literature Review 2
  1.1 Introduction ................................ 2
  1.2 Literature Review ................................ 3
  1.3 Objective of This Thesis .......................... 9

2 Methodology: Model and Estimation 12
  2.1 Introduction ................................ 12
  2.2 Model ....................................... 12
  2.3 Data Generating Process: Near Epoch Dependence ................. 13
  2.4 Estimation ................................... 16

3 Theoretical studies: Large sample property under NED 19
  3.1 Introduction ................................ 19
  3.2 Notation and Assumptions .......................... 19
  3.3 Theorems ................................... 21
  3.4 Proofs ....................................... 23
4 Simulation studies: Finite sample performance

4.1 Introduction ................................ 29
4.2 Case of Non-linear Granger-causality ................. 29
4.3 Case of Linear Granger-causality ....................... 33

5 Application to real data: Cross-country empirical evidence of Granger-causality

5.1 Introduction ................................ 36
5.2 Data .................................... 37
5.3 Semiparametric regression analysis of Granger-causality .... 38
5.4 Empirical evidence of Granger-causality ............... 43
  5.4.1 FTSE 100 and ASX 200 ........................ 43
  5.4.2 S&P 500 and ASX 200 ........................ 45
  5.4.3 FTSE 100 and S&P 500 ...................... 47
  5.4.4 FTSE 100 and SSE ........................ 49
  5.4.5 S&P 500 and SSE .......................... 51
  5.4.6 ASX 200 and SSE .......................... 53
5.5 Summary .................................... 55

6 Conclusion .................................. 57

Bibliography .................................. 59
List of Figures

4.2.1 Monte carlo simulation: sample size $T = 100$ (Top) and $T = 500$ (Bottom), with the theoretical function $f(Y_{t-1}) = \sin(Y_{t-1})$ .... 31

4.2.2 Boxplot of $|\hat{\alpha} - \alpha|$ for the nonlinear case .................. 32

4.3.1 Monte carlo simulation: sample size $T = 100$ (Top) and $T = 500$ (Bottom), with the theoretical function $f(Y_{t-1}) = 1 - 2 \times (Y_{t-1})$ .... 34

4.3.2 Boxplot of $|\hat{\alpha} - \alpha|$ for the linear case ................... 35

5.2.1 Weekly Data and Changes in log of S&P 500 ......................... 39

5.2.2 Weekly Data and Changes in log of FTSE 100 .......................... 40

5.2.3 Weekly Data and Changes in log of ASX 200 .......................... 41

5.2.4 Weekly Data and Changes in log of SSE .............................. 42

5.4.1 $f(Y_{t-1})$ and $g(X_{t-1})$ for UK vs AUS ............................... 44

5.4.2 $f(Y_{t-2})$ and $g(X_{t-2})$ for UK vs AUS ............................... 45

5.4.3 $f(Y_{t-1})$ and $g(X_{t-1})$ for USA vs AUS ............................. 46

5.4.4 $f(Y_{t-2})$ and $g(X_{t-2})$ for USA vs AUS ............................. 47

5.4.5 $f(Y_{t-1})$ and $g(X_{t-1})$ for USA vs UK .............................. 48

5.4.6 $f(Y_{t-2})$ and $g(X_{t-2})$ for USA vs UK .............................. 49

5.4.7 $f(Y_{t-1})$ and $g(X_{t-1})$ for UK vs China ........................... 50

5.4.8 $f(Y_{t-2})$ and $g(X_{t-2})$ for UK vs China ........................... 51

5.4.9 $f(Y_{t-1})$ and $g(X_{t-1})$ for USA vs China ........................... 52

5.4.10 $f(Y_{t-2})$ and $g(X_{t-2})$ for USA vs China ............................ 53

5.4.11 $f(Y_{t-1})$ and $g(X_{t-1})$ for AUS vs China ............................ 54
5.4.12 $f(Y_{t-2})$ and $g(X_{t-2})$ for AUS vs China
Chapter 1

Background and Literature Review

1.1 Introduction

It is of wide interest to understand how the financial shocks are transmitted between the foreign and domestic countries so that steps can be taken to reduce financial instabilities it causes in domestic markets. This interest is basically due to the massive welfare losses and economic dislocations that are caused by successive devaluations of currencies, collapse of the stock markets, and the ensuing economic depression in the countries that experience financial crisis.

As for much of the world, the key turning point for the Australian economy was the change that swept through the global economy in mid-September 2008, with the collapse of Lehman Brothers. The Australian Stock Exchange (ASX) fell by more than 40% and businesses exposed to the global economy or reliant on credit began to suffer. Job losses followed and consumer confidence began to wane, further impacting businesses and jobs. In the December quarter 2008, businesses ran down their stocks by $3.4 billion (in real terms), the largest fall on record. On all benchmarks,
Australia entered a recession for the first time in nearly 20 years.

Understanding of such transmission mechanism of the shocks among financial markets in different countries is obviously important and interesting both for academic researchers and for investment professionals (Brooks & Henry 2000). One challenging question that faces academic researchers is how to investigate the complex transmission of the shocks from one market to another from the perspective of prediction. Linear Granger causality testing has been found to be a useful tool to study the transmission mechanism among different financial markets in the literature.

1.2 Literature Review

In 1969, Granger proposed a causality test to describe the interdependence relations between economic time series. The concept of Granger causality starts with the premise that the future cannot cause the past. Granger applies this concept to economic time series to determine whether one time series causes in the sense of precedes another. The recent empirical evidence is invariably based on the linear Granger causality test (Granger 1969). The conventional approach of testing for Granger causality is to assume a parametric, linear time series model for the conditional mean. This approach is appealing, since the test reduces to determining whether the lags of one variable enter into the equation for another variable, although it requires the linearity assumption.

If event A occurs after event B, then A cannot cause B. Granger applies this concept to economic time series to determine whether one time series causes in the sense of precedes another. However, merely because event A occurs before B does not mean that A causes B. It is related to the question of how useful one variable
(or set of variables) $Y$ is for forecasting another variable (or set of variables) $X$. If $Y$ cannot help forecast $X$, then we say that $Y$ does not Granger-cause $X$. More formally, for a two-dimensional time series $(X_t, Y_t)$, $t = 1, ..., T$, we are concerned of the existence of any causality between $X_t$ and $Y_t$. Granger defined that $Y_t$ is said to cause $X_t$ in mean if $\beta(L) \neq 0$ in the following linear model

$$X_t = \alpha(L) X_t + \beta(L) Y_t + U_{X,t},$$  \hspace{1cm} (1.2.1)$$

where $\alpha(L)$ and $\beta(L)$ are respectively lag polynomials of order $p$ and $q$ without constant terms (see Granger (1969, p.429)), and denoted it as $Y_t \rightarrow X_t$. Otherwise, $Y_t \nrightarrow X_t$. We say that $Y_t$ causes $X_t$ if we have better linear prediction of $X_t$ using the past information $Y_{t-1}, ..., Y_1$. Granger remarked this definition as ”linear causality in mean”. With the linearity assumption, the null hypothesis $H_0 : Y_t \nrightarrow X_t$ against $H_1 : Y_t \rightarrow X_t$ can be tested as in Sims (1972) or Hosoya (1977). Many researchers have used this method since Sims (1972), Toda & Philips (1993), Hosoya (1991) and Lutkepohl & Poskitt (1996).

Baek & Brock (1992) noted that parametric linear Granger causality tests have low power against certain nonlinear alternatives. They pointed out that the rejection of linear causality does not imply there is no non-linear causality because the conventional vector autoregressive (VAR) model only focuses on linear relations. In view of this, nonparametric techniques are appealing because they place direct emphasis on prediction without imposing a linear functional form. Various nonparametric Granger causality tests have been proposed in the literature (Hiemstra & Jones 1994, Diks & Panchenko 2006, Hidalgo 2000, Nishiyama et al. 2011).

The recent literature, due to the availability of ever cheaper computational power, has shown an increasing interest in nonparametric versions of the Granger non-causality hypothesis against general (linear as well as nonlinear) Granger causality. Among the various nonparametric tests for the Granger non-causality hypoth-
thesis, the Hiemstra & Jones (1994) test is the most frequently used one among practitioners in economics and finance. Hiemstra & Jones (1994) used both linear and nonlinear causality tests to study daily Dow Jones stock returns and percentage changes in New York Stock Exchange trading volume over the 1915 to 1946 and 1947 to 1990 periods. Using the traditional Granger test, they investigated the presence of linear predictive power between stock returns and trading volume change. The nonlinear Granger causality test used in their article is based on nonparametric estimators of temporal relations within and across time series. It is a modified version of Baek & Brock (1992) nonlinear Granger causality test for conditional independence, with critical values based on asymptotic theory.

Both Baek & Brock (1992) and Hiemstra & Jones (1994) assume \( \alpha \)-mixing conditions for \( X_t \) and \( Y_t \) in model (1.2.2) and (1.2.3). In order to test for nonlinear Granger causality between \( X_t \) and \( Y_t \), Baek & Brock (1992) applied two estimated residuals series from the following VAR model, \( \hat{U}_{X,t} \) and \( \hat{U}_{Y,t} \), to a test statistic. Their test statistic is constructed by using the correlation-integral estimators of the joint probabilities under the assumption that the errors of the VAR model are mutually independent and individually i.i.d.

\[
X_t = A(L)X_t + B(L)Y_t + U_{X,t}, \quad (1.2.2)
\]
\[
Y_t = C(L)X_t + D(L)Y_t + U_{Y,t}, \quad (1.2.3)
\]

where \( A(L), B(L), C(L), \) and \( D(L) \) are one-sided lag polynomials of orders \( a, b, c \) and \( d \) in the lag operator \( L \) with roots outside the unit circle and no roots in common.

Hiemstra & Jones (1994) modified Baek & Brock (1992)’s test statistic so that it can be applied to more general case where the errors are allowed to be \( \alpha \)-mixing. This test can detect the nonlinear Granger-causal relationship between variables by testing whether the past values influence present and future values.
To our best knowledge, these are the first approaches to detect the existence of nonlinear Granger-causality. They suggested a test statistic that can be applied to test nonlinear Granger-causality between the variables. Even if we can tell whether or not there exist nonlinear Granger-causality by using the test, the form of Granger-causality cannot be identified. Moreover, their theory is confined to the data-generating process that is $\alpha$-mixing, which is not easy to verify (even theoretically) and may be violated in practice.

Besides Hiemstra-Jones test, other forms of nonlinear causality test have also been developed. Diks and Panchenko (2005, 2006) demonstrated that the Hiemstra and Jones test can severely over-reject the null hypothesis that non-causality is true, i.e., the Hiemstra and Jones test has serious size distortion problems. Alternatively Diks & Panchenko (2006) developed a new test statistic that overcomes these limitations. Their empirical results suggest that some of the rejections of the Granger non-causality hypothesis, using the Hiemstra and Jones test, may be spurious.

Hidalgo (2000) and Nishiyama et al. (2011) also introduced a nonparametric Granger-causality test. Hidalgo’s test for covariance stationary linear processes was under long range dependence while tests for causality are analyzed when the data are short-range dependent in the literature (Granger 1969, Geweke 1982). However, Hidalgo’s approach may fail if there are some nonlinear causal relationships because they constructed test statistics based on linear projections of the series of interest. Many aforementioned researchers applied frequency domain analysis where they capture causality through cross-spectra or covariances of $y$ and $x$ (Sims 1972, Geweke 1982, Sims et al. 1990, Toda & Philips 1993, Hosoya 1991). However, if there are nonlinear relationships then covariances can easily be zero even if the two variables are dependent. So, tests based on the covariances do not seem to have good power property against certain alternatives.
Nishiyama et al. (2011) developed a test statistics allowing nonlinear dependence under $\beta$-mixing condition for $X_t$ and $Y_t$, which are special $\alpha$-mixing processes. They replaced the linear projections by the optimum predictor, or conditional expectations and rewrote (1.2.1) as

$$E[X_t - E(X_t|X_{t-1},...,X_1)]^2 > E[X_t - E(X_t|X_{t-1},...,X_1,Y_{t-1},...,Y_1)]^2.$$ \hspace{1cm} (1.2.4)

They defined “$Y_t$ (possibly nonlinearly) causes $X_t$ in mean” if

$$E[E(X_t|X_{t-1},...,X_1,Y_{t-1},...,Y_1) - E(X_t|X_{t-1},...,X_1)]^2 > 0,$$ \hspace{1cm} (1.2.5)

and test the null hypothesis

$$H_0 : E (X_t|X_{t-1},...,X_1,Y_{t-1},...,Y_1) = E (X_t|X_{t-1},...,X_1) \text{ w.p. } 1$$

against the alternative hypothesis (1.2.4). The null and alternative hypotheses can be rewritten as

$$H_0 : P [E (U_t|Z_{t-1}) = 0] = 1$$

and

$$H_1 : P [E (U_t|Z_{t-1}) = 0] < 1,$$

where $U_t = X_t - E (X_t|X_{t-1})$ with $X_{t-1} = (X_{t-1},..,X_{t-p})$ and $Z_{t-1} = (X_{t-1},..,X_{t-p},Y_{t-1},..,Y_{t-q})$.

Note that

$$E (U_t|Z_{t-1}) = 0 \Leftrightarrow E [U_t p^*(Z_{t-1})] = 0 \text{ for } \forall p^*(z) \in s_X^\perp,$$

where $s_X^\perp$ is a Hilbert space orthogonal to the Hilbert $L_2$ space $s_X = \{ s(\cdot) | E [s(X_{t-1})^2] < \infty \}$. Rewriting the hypothesis as “density-weighted” version by using the probability density function of $X_{t-1}$, the null and alternative hypotheses can be represented as

$$H_0 : E [U_t f (X_{t-1}) p^*(Z_{t-1})] = 0 \text{ for } \forall p^*(z) \in s_X^\perp.$$
CHAPTER 1. BACKGROUND AND LITERATURE REVIEW

and

\[ H_1 : E[U_t f(X_{t-1}) p^*(Z_{t-1})] \neq 0 \text{ for some } p^*(z) \in s_X^\perp. \]

Using a complete basis \( H(z) = \{h_i(z)\}_{i=1}^\infty \) over \( s_X^\perp \) satisfying

\[ \text{Var}[U_t f(X_{t-1}) h_i(Z_{t-1})] = 1 \]

and

\[ \text{Cov}[U_t f(X_{t-1}) h_i(Z_{t-1}), U_t f(X_{t-1}) h_j(Z_{t-1})] = 0 \]

for all \( i \neq j \) and given \( \beta \)-mixing processes \( X_t \) and \( Y_t \), they define

\[ a_i = \frac{1}{\sqrt{T}} \sum_{t=r}^{T} U_t f(X_{t-1}) h_i(Z_{t-1}) \quad (1.2.6) \]

where \( r = \max(p,q) + 1 \) and \( i = 1, \ldots, k_T \) with \( k_T < T \). By the central limit theorem for martingale difference sequence

\[ a_i \xrightarrow{D} N(0,1) \]

under the null, while \( |a_i| \to \infty \) under the alternative for some \( i \) as \( T \to \infty \). Using this they constructed a test statistic

\[ S_T = \sum_{i=1}^{k_T} w_i a_i^2 \xrightarrow{D} \sum_{i=1}^{\infty} w_i \varepsilon_i^2 \]

where \( \varepsilon_i \) are i.i.d. \( N(0,1) \). Given a user-determined basis of \( s_Z \), \( \{q_i(Z)\}_{i=1}^\infty \), \( \{[q_i(Z) - r_i(Z)] f(X)\}_{i=1}^\infty \) forms a basis of \( s_X^\perp \), where

\[ r_i(X) = E[q_i(Z_{t-1}) | X_{t-1} = X]. \]

Normalizing this basis of \( s_X^\perp \) yields the complete basis \( H(z) \). The unknown elements \( U_t, f(X_{t-1}) \) and \( h_i(Z_{t-1}) \) in (1.2.6) are replaced by their estimates \( \hat{U}_t = X_t - \hat{E}(X_t | X_{t-1}), \hat{f}(X_{t-1}), \hat{h}_i(Z_{t-1}) \) and \( \hat{r}_i(X) \). They used Nadaraya-Watson kernel estimator for these.

In summary, Nishiyama et. al. (2011) look at the conditional moments to test for Granger causality while Hiemstra and Jones (1994) consider the joint cumulative distribution. The causality definition of the former is slightly stronger than the
latter but both use a framework of data generating process that is $\beta$-mixing, a special case of $\alpha$-mixing. Moreover, these mixing conditions are stronger than near epoch dependence we will employ in this study. Also, the above testing procedures are nonparametrically based. In practice, when the lag order $p$ in $X_t$ is large, say $p > 3$, the above methods in general suffer from “curse of dimensionality” in the nonparametric, such as Nadaraya-Watson kernel, estimation. In addition, none of these approaches provides us with the form or structure of nonlinear Granger-causality if it exists. Compared to Hidalgo (2000), Nishiyama et al. (2011)’s Monte Carlo studies show that their testing method works better than the method of Hidalgo (2000) in the case of nonlinear dependence but worse when the series is of linear dependence in mean.

1.3 Objective of This Thesis

As we have seen so far, some different parametric and nonparametric Granger-causality tests have been developed. However, all of them can only be used to investigate if there exists any linear or nonlinear Granger-causality or not. Furthermore, those tests have been developed under the framework of data generating process that is $\alpha$-mixing or $\beta$-mixing, which is not easy to verify (even theoretically). In fact, in many situations, $\alpha$-mixing or $\beta$-mixing may be violated under the near epoch dependence assumption we adopt in this paper. For example, Andrews (1984) showed that a simple AR(1) model: $X_t = 0.5X_{t-1} + e_t$, with $e_t$ being $i.i.d$ Bernoulli variables satisfying $P(e_t = -1) = P(e_t = 1) = 0.5$, is not $\alpha$-mixing. Researchers have applied the Granger causality test statistics to financial data such as stock return. However it is well known that the dynamics of stock return are often well described by an $ARMA(p,q)-GARCH(r,m)$ process. For most of $ARMA(p,q)-GARCH(r,m)$ processes, it is very difficult to verify their $\alpha$-mixing properties except in some very special cases (Lu (1996a, 1996b) and Carrasco and Chen (2002)). The near epoch dependence is more general than $\alpha$-mixing and it
covers the compounded processes and many nonlinear/non-$\alpha$-mixing processes such as ARMA($p,q$)-GARCH($r,m$). Therefore it is more appropriate to assume near epoch dependence of the variables to consider if there is any Granger-causality. In addition, if we use a semiparametric regression model (1.3.1), we can check if there is any linear or nonlinear Granger-causality from $Y_t$ to $X_t$ and find the form or structure of Granger-causality by estimating the unknown function $f$. Hence, we introduce a semiparametric model to find the form of the relations and to check the existence of linear or nonlinear Granger-causality.

In this thesis, the main objective is to develop a class of semiparametric time series regression models that are of partially linear structures, with statistical theory established under a more general framework of near epoch dependent (NED) data generating processes, which are easily applicable in exploring nonlinear Granger causality. Lu & Linton (2007) developed an asymptotic theory of local linear fitting for near epoch dependent processes, which provides a foundation for the development of theory in this paper. We applied Robinson (1988)’s method to estimate the nonlinear functions in the following semiparametric model (1.3.1):

$$X_t = \alpha(L)X_t + f(Y_{t-1}) + U_{X,t}, \quad (1.3.1)$$

and to investigate the complex transmission behaviour. We assume that variables in model (1.3.1) follow near epoch dependent process that is more general than $\alpha$-mixing process. Therefore we can use an ARMA($p,q$)-GARCH($r,m$) process which is a NED process. In doing so, Lu & Linton (2007)’s results are employed to estimate conditional mean function nonparametrically. In regard to nonlinear Granger causality test, if nonlinear Granger causality really exists from $\{Y_t\}$ to $\{X_t\}$, or vice versa, we can find not only the existence of the Granger causality but also the form of the relationship. We have shown that estimators of both parametric and nonparametric parts are consistent and have asymptotic normal distribution.
CHAPTER 1. BACKGROUND AND LITERATURE REVIEW

After developing theories, some Monte Carlo simulation studies are conducted to check how the theories work for a finite sample. We can estimate the parameters and unknown function very accurately for both cases where the unknown function is linear and nonlinear.

Finally, we have applied our proposed model and methodology to the stock market indices, including ASX200, S&P500, FTSE100 and SSE from Australia, the USA, the UK and China, respectively, to examine if there is any Granger-causality between each pair of these countries’ stock markets. Weekly data are used to avoid market micro structure problems. Interestingly, we could see that there is a strong nonlinear Granger causality from FTSE100 to ASX200, some linear or nonlinear Granger causality from S&P500 to ASX200 and some negative linear Granger causality from SSE to ASX200. From these results, we can say that the USA, UK and Chinese stock prices 1 week ago help to predict current Australian stock market behavior. There is also some linear Granger causality from FTSE100 and S&P500 to SSE but not from ASX200 to SSE. We could not find any evidence of Granger causalities from other countries to the UK or the USA. We could not find any evidence of Granger causality between the countries either when we use a two lagged variable except the (Australia, China) case. Chinese stock price 2 weeks ago appears to help to predict this week’s Australia stock price but not the other way around.

The remainder of this paper is organized as follows. In Chapter 2, we suggest semiparametric time series regression models and their estimation method. Then, we study the consistency and asymptotic normality of our estimators in Chapter 3. Chapter 4 reports the finite-sample performance of the theory by using Monte Carlo simulation studies. Empirical results are given in Chapter 5. Finally, Chapter 6 concludes.
Chapter 2

Methodology: Model and Estimation

2.1 Introduction

We are adopting a semiparametric time series regression model structure not only because it is a natural extension of the linear Granger causality testing but also because it enables us to better understand the likely form or structure of the nonlinear Granger causality if it exists. In this chapter, we first suggest a semiparametric time series regression model for near epoch dependent processes. Then we will discuss how to estimate the parameters and unknown function in the model.

2.2 Model

We suggest a semiparametric model to find, if any, the form or structure of Granger causality from $Y_t$ to $X_t$ as follows

$$X_t = a^T X_{t-1} + f(Y_{t-1}) + \epsilon_{X,t}, \text{where } E(\epsilon_{X,t}|X_{t-1}, Y_{t-1}) = 0,$$

where $X_{t-1} = (X_{t-1}, ..., X_{t-p})$ and $Y_{t-1} = (Y_{t-1}, ..., Y_{t-d})$. Here $\{(X_t, X_{t-1}, Y_{t-1})\}$ is a stationary near epoch dependent (NED) process with respect to an $\alpha$-mixing
process \( \{ \varepsilon_t \} \). Let us discuss the data structure that we will assume in our model before we consider estimation.

### 2.3 Data Generating Process: Near Epoch Dependence

We consider how to establish a model in a time series context under near epoch dependence. The near epoch dependence is more general than the most popular \( \alpha \)-mixing condition in the literature. Thus it can cover richer sets of processes for which \( \alpha \)-mixing condition is violated or harder to be proven. One of such important examples is an ARMA \((p,q) - GARCH (r,m)\) which has been used to explain the dynamics of financial data such as stock return in the literature. Uniform convergence with rates for nonparametric density and regression estimators based on the local constant paradigm was done under near epoch dependence conditions by Andrews (1995). Lu & Linton (2007) and Li, Lu and Linton (2012) established an asymptotic theory and uniform convergence rates of a more desired local linear fitting for NED processes.

Let us first provide some basic definitions. In this thesis, all the random variables are defined on some probability space \((\Omega, F, P)\). Among the mixing conditions such as \(\phi\)-, \(\rho\)-, \(\beta\)- and \(\alpha\)-mixing, \(\alpha\)-mixing is the weakest and most popular in the econometric literature. The stationary solutions of many time series linear or nonlinear econometric models are \(\alpha\)-mixing under suitable conditions: see e.g. (Pham 1986, Pham & Tran 1985, Tong 1990, Lu 1998). For references, the definition of \(\alpha\)-mixing is stated as follows.
CHAPTER 2. METHODOLOGY: MODEL AND ESTIMATION

Definition 1. A stationary sequence \( \{X_t, t = 0, \pm 1, \ldots\} \) is said to be \( \alpha \)-mixing if

\[
\alpha(k) = \sup_{A \in F_{-\infty}^n, B \in F_{n+k}^\infty} |P(AB) - P(A)P(B)| \to 0 \tag{2.3.1}
\]
as \( k \to \infty \), where \( F_{-\infty}^n \) and \( F_{n+k}^\infty \) are two \( \sigma \)-fields generated by \( \{X_t, t \leq n\} \) and \( \{X_t, t \geq n + k\} \), respectively. We call \( \alpha(\cdot) \) the mixing coefficient.

However, Davidson (1994) pointed out the followings (see also (Lu 2001)): (i) even simple autoregressive processes might not be \( \alpha \)-mixing and (ii) \( \alpha \)-mixing is hard to verify in practice, especially for the compound processes. See Andrews (1984) for the former case. Let us give an example for the latter case. The autoregressive moving average(ARMA) process with ARCH/GARCH errors, discussed in Engle (1982) and Weiss (1984) and also Ling & Li (1997), is applied well in financial econometrics, where the model is composed of two time series(ARMA and ARCH/GARCH) models; say

\[
X_t = a_0 + a_1 X_{t-1} + \epsilon_t + a_3 \epsilon_{t-1},
\]
\[
\epsilon_t = e_t h_t^{1/2}, h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}, \tag{2.3.2}
\]
where \( a_i \) are the coefficients in the ARMA model, \( \alpha_i \) and \( \beta_i \) are the coefficients in the GARCH model, with \( e_t \) being identically distributed innovation with mean 0 and variance 1. Under some mild conditions ARCH or GARCH models have been known as \( \alpha \)-mixing but there are no general results that the compound processes like (2.3.2) are \( \alpha \)-mixing and it may be even harder to see the \( \alpha \)-mixing of data generating processes. Therefore, we suggest a generalized version of mixing processes, called stable or near epoch dependent processes, which can cover the compounded processes and many nonlinear non-\( \alpha \)-mixing processes. Ibragimov (1962) introduced this concept and Billingsley (1968) and McLeish(1975a, 1975b, 1977) made further development, and many other researchers used this concept in the literature(Bierens 1981, Gallant 1987, Gallant & White 1988, Andrews 1995, Lu 2001).
We consider a pair of random processes in Granger causality. Let \( Y_t \) and \( X_t \) be both \( \mathbb{R}^1 \)-valued stationary processes, defined based on a stationary process \( \{\epsilon_t\} \) by

\[
Y_t = \Psi_Y(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots), \quad (2.3.3)
\]

\[
X_t = \Psi_X(\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots), \quad (2.3.4)
\]

where \( \Psi_Y : \mathbb{R}^\infty \to \mathbb{R}^1 \) and \( \Psi_X : \mathbb{R}^\infty \to \mathbb{R}^1 \) are two Borel measurable functions, respectively, and \( \{\epsilon_t\} \) may be vector-valued. Let \( \nu > 0 \) be a positive real number.

**Definition 2.** The stationary process \( \{(Y_t, X_t)\} \) is said to be near epoch dependent in \( L_\nu \) norm (NED in \( L_\nu \) for simplicity) with respect to a stationary \( \alpha \)-mixing process \( \{\epsilon_t\} \) if

\[
v_\nu(m) = E|Y_t - Y_t^{(m)}|\nu + E|X_t - X_t^{(m)}|\nu \to 0 \quad (2.3.5)
\]

as \( m \to \infty \), where \( |\cdot| \) is the absolute value, \( Y_t^{(m)} = \Psi_{Y,m}(\epsilon_t, \ldots, \epsilon_{t+m}) \), \( X_t^{(m)} = \Psi_{X,m}(\epsilon_t, \ldots, \epsilon_{t+m}) \), and \( \Psi_{Y,m} \) and \( \Psi_{X,m} \) are \( \mathbb{R}^1 \)-valued Borel measurable functions with \( m \) arguments, respectively. We will call \( v_\nu(m) \) the stability coefficients of order \( \nu \) of the process \( \{(Y_t, X_t)\} \).

Definitely, \( \{(Y_t^{(m)}, X_t^{(m)})\} \) is an \( \alpha \)-mixing process with mixing coefficient

\[
\alpha_m(k) = \begin{dcases}
\alpha(k - m) & k \geq m + 1, \\
1 & k \leq m.
\end{dcases} \quad (2.3.6)
\]

This kind of setting is good for models with complicated dynamics in both mean and variance for which the usual mixing conditions are not applied or difficult to check. In finance and economics, these kinds of models are very common and near epoch dependence is easier to see in these cases. This is the reason why we consider establishing a model under near epoch dependence in this thesis.
2.4 Estimation

We are interested in how \( \{Y_t\} \) Granger-cause \( \{X_t\} \) in this model. If we can find the form of \( f \), then we can know what kind of Granger-causality exists from \( \{Y_t\} \) to \( \{X_t\} \). For example, if \( f(Y_t) = 0 \), it implies non-Granger-causality. By using the idea of Robinson (1988) and the theory developed in Lu & Linton (2007) for near epoch dependent processes, we can estimate both the parameters and the unknown function in our semiparametric regression model.

Let us discuss how to estimate the unknown function \( f \) and the parameter vector \( a \) by using the method in Robinson (1988).

Since \( E(\epsilon_{X,t}|X_{t-1}, Y_{t-1}) = 0 \), taking conditional expectation of both sides of equation (2.2.1) given \( Y_{t-1} \), we get

\[
E(X_t|Y_{t-1}) = a^T E(X_{t-1}|Y_{t-1}) + f(Y_{t-1}).
\] (2.4.1)

If we subtract equation (2.4.1) from (2.2.1),

\[
X_t - E(X_t|Y_{t-1}) = a^T (X_{t-1} - E(X_{t-1}|Y_{t-1})) + \epsilon_{X,t}.
\] (2.4.2)

From the equation (2.4.2), the parameter vector \( a \) can be estimated by the least squares estimation as in (2.4.3) if we could know \( f_1(Y_{t-1}) = E(X_t|Y_{t-1}) \) and \( f_2(Y_{t-1}) = E(X_{t-1}|Y_{t-1}) \). We will replace \( f_1 \) and \( f_2 \) by their suitable estimators \( \hat{f}_1 \) and \( \hat{f}_2 \) below. Then,

\[
\hat{a} = \left[ \frac{1}{T} \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) \right]^{-1} 
\times \left[ \frac{1}{T} \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t} - \hat{f}_1(Y_{t-1})) W_T(Y_{t-1}) \right],
\] (2.4.3)

where \( W_T(Y_{t-1}) = I(A_T \leq Y_{t-1} \leq B_T) \) with suitable chosen \( A_T \) and \( B_T \), which is used to exclude extreme values on both sides.
Although \( f_1 \) and \( f_2 \) are unknown, we can estimate the unknown functions \( f_1 \) and \( f_2 \) by Lu & Linton (2007)'s results.

Let’s consider the function \( f_1 \) first. We can approximate the unknown function \( f_1 \) in the neighborhood of \( y \) using the local linear fitting method. We have

\[
\hat{f}_1(Y_{t-1}) \approx f_1(y) + (f'_1(y))' (Y_{t-1} - y) := \alpha_0 + \alpha_1 (Y_{t-1} - y)
\]

(2.4.4)

in a neighborhood of \( y \) under Assumption (A3) in the Section 3.1. Locally, we can estimate \( f_1 \) from

\[
\begin{pmatrix}
\hat{f}_1(y) \\
\hat{f}'_1(y)
\end{pmatrix} = \left( \begin{array}{c}
\hat{\alpha}_0 \\
\hat{\alpha}_1
\end{array} \right) = \arg\min_{\{\alpha_0, \alpha_1\}} \sum_{t=d+1}^{T} (X_t - \alpha_0 - \alpha_1 (Y_{t-1} - y))^2 K \left( \frac{Y_{t-1} - y}{\varrho_T} \right),
\]

where \( \varrho_T \) is a sequence of bandwidths tending to zero at appropriate rate as \( T \) tends to infinity and \( K(\cdot) \) is a (bounded) kernel with values in \( \mathbb{R}^1 \).

Then, we have

\[
\begin{pmatrix}
\hat{f}_1(y) \\
\hat{f}'_1(y) \varrho_T
\end{pmatrix} = \begin{pmatrix}
\hat{\alpha}_0 \\
\hat{\alpha}_1 \varrho_T
\end{pmatrix} = U_T^{-1} V_T,
\]

(2.4.5)

where

\[
V_T := \begin{pmatrix}
v_{T0} \\
v_{T1}
\end{pmatrix} \quad \text{and} \quad U_T := \begin{pmatrix}
u_{T00} & u_{T01} \\
u_{T10} & u_{T11}
\end{pmatrix}
\]

with

\[
(V_T)_i := (T \varrho_T^d)^{-1} \sum_{t=d+1}^{T} X_t \left( \frac{Y_{t-1} - y}{\varrho_T} \right) K \left( \frac{Y_{t-1} - y}{\varrho_T} \right)
\]

and

\[
(U_T)_{il} := (T \varrho_T^d)^{-1} \sum_{t=d+1}^{T} \left( \frac{Y_{t-1} - y}{\varrho_T} \right)_i \left( \frac{Y_{t-1} - y}{\varrho_T} \right)_l K \left( \frac{Y_{t-1} - y}{\varrho_T} \right).
\]
Similarly, we can estimate \( f_2(Y_{t-1}) = E(X_{t-1}|Y_{t-1}) \) in a neighborhood of \( y \) since
\[
f_2(Y_{t-1}) = E(X_{t-1}|Y_{t-1}) = (E(X_{t-1}|Y_{t-1}), ..., E(X_{t-p}|Y_{t-1})) = (f_{21}(Y_{t-1}), ..., f_{2p}(Y_{t-1})) \approx (b_{01} + b_{1}^T(Y_{t-1} - y), ..., b_{0p} + b_{p}^T(Y_{t-1} - y)). \tag{2.4.6}
\]
\[
\left( \hat{f}_{2j}(y), \hat{f}'_{2j}(y) \right) = \left( \hat{b}_0, \hat{b}_1 \right) = \arg\min_{(b_0, b_j)} \left\{ \sum_{t=d+1}^{n} (X_{t-j} - b_{0j} - b_j(Y_{t-1} - y))^2 K\left( \frac{Y_{t-1} - y}{\varrho_T} \right) \right\},
\]
where \( \varrho_T \) is a sequence of bandwidths tending to zero at appropriate rate as \( T \) tends to infinity and \( K(\cdot) \) is a (bounded) kernel with values in \( \mathbb{R}^+ \). Then we have
\[
\left( \hat{f}_{2j}(y), \hat{f}'_{2j}(y) \varrho_T \right) = \left( \hat{b}_0, \hat{b}_j \right) = U_T^{-1} \tilde{V}_{Tj}, \tag{2.4.7}
\]
where \( U_T := \begin{pmatrix} u_{T00} & u_{T01} \\ u_{T10} & u_{T11} \end{pmatrix} \) is defined in (2.4.5) and \( \tilde{V}_{Tj} := \begin{pmatrix} \tilde{v}_{T0} \\ \tilde{v}_{T1} \end{pmatrix} \) with
\[
(\tilde{V}_{Tj})_i := (T \varrho_T^{d-1})^{-1} \sum_{t=d+1}^{T} X_{t-j} \left( \frac{Y_{t-1} - y}{\varrho_T} \right)_i K\left( \frac{Y_{t-1} - y}{\varrho_T} \right)_i, \quad i = 0, ..., d, \text{ and } j = 1, ..., p.
\]
Finally, we can estimate \( f \) by using (2.4.1), as follows
\[
\hat{f}(y) = \hat{f}_1(y) - \hat{a}^T \hat{f}_2(y). \tag{2.4.8}
\]
We will show the consistency and asymptotic normality of \( \hat{a} \) and \( \hat{f} \) in Chapter 3.
Chapter 3

Theoretical studies: Large sample property under NED

3.1 Introduction

In this chapter, we establish the consistency and asymptotic normality for the estimators we constructed in chapter 2 under a general, popular data generating framework of near epoch dependence. After presenting notations and assumptions that will be used, we discuss the main asymptotic results and proofs are provided in the last section.

3.2 Notation and Assumptions

The main assumptions that we require for the data generating process \((X_t, X_{t-1}, Y_{t-1})\) in our model (2.2.1) are as follows. We follow the main assumptions (A1)-(A4) from Lu & Linton (2007) for our data generating process and assumption (A5) for the kernel function \(K\) in the estimation. In addition to the assumptions (A1)-(A5), we assume that \(E(\epsilon_{X,t}|X_{t-1}, Y_{t-1}) = 0\). For the consistency and asymptotic normality, we follow the assumptions (B1)-(B4) from Lu & Linton (2007).
CHAPTER 3. THEORETICAL STUDIES: LARGE SAMPLE PROPERTY

UNDER NED

\( (A0) \) \( E(\epsilon_{X,t}|X_{t-1}, Y_{t-1}) = 0. \)

\( (A1) \) The data generating process \( \{(X_t, Y_t)\} \) is a strictly stationary NED process, with order \( \nu = 2 + \delta/2 \), with respect to the \( \alpha \)-mixing process \( \{\epsilon_t\} \) defined in Definition 2 (Chapter 2), where the constant \( \delta > 0 \) is specified in Assumption \( (A2) \), which follows. For all \( i \) and \( j \) in \( \mathbb{Z} \), the vectors \( Y_i \) and \( Y_j \) admit a joint density \( h \); moreover, \( h_{ij}(y', y'') \leq C \) for all \( i, j \in \mathbb{Z} \), all \( y', y'' \in \mathbb{R}^d \), where \( C > 0 \) is some constant, and \( h \) denotes the marginal density of \( Y_i \).

\( (A2) \) The random variable \( X_i \) has finite absolute moment of order \( (2 + \delta) \), that is, \( E[|X_i|^{2+\delta}] < \infty \) for some \( \delta > 0 \).

\( (A3) \) (i) The regression functions \( f_1(y) = E(X_t|Y_{t-1} = y) \) and \( f_2(y) = E(X_{t-1}|Y_{t-1} = y) \) are twice differentiable. Denoting by \( f'_1(y) \), \( f'_2(y) \) and \( f''_1(y) \), \( f''_2(y) \) their gradients and the matrices of their second derivatives (at \( y \)), respectively, \( y \mapsto f''_1(y) \) and \( y \mapsto f''_2(y) \) are continuous at \( y \). (ii) The density function \( h(y) \) is continuous at \( y \). (iii) The conditional variance function \( \text{Var}(X_t|Y_{t-1} = y) \) and \( \text{Var}(X_{t-1}|Y_{t-1} = y) \) are continuous at \( y \).

\( (A4) \) For the mixing coefficient of \( \epsilon_t \) in Definition 2, the function \( \alpha \) is such that

\[
\lim_{k \to \infty} k^\alpha \sum_{j=k}^{\infty} \alpha(j) \delta/(4+\delta) = 0
\]

for some constant \( a > \delta/(4 + \delta) \).

\( (A5) \) For any \( c := (c_0, c_1^\tau) \in \mathbb{R}^{d+1} \), define \( K_c(u) := (c_0 + c_1^\tau u)K(u) \).

(i) For any \( c \in \mathbb{R}^{d+1} \), \( |K_c(u)| \) is uniformly bounded by some constant \( K_c^+ \) and is integrable: \( \int_{\mathbb{R}^d} |K_c(x)| dx < \infty \).

(ii) For any \( c \in \mathbb{R}^{d+1} \), \( |K_c| \) has an integrable second-order radial majorant; that is, \( Q_c^K(x) := \sup_{\|y\| \geq s} ||y||^2 K_c(y) \) is integrable.

(iii) For any \( c \in \mathbb{R}^{d+1} \), \( K_c \) is Lipschitz continuous of order 1; that is, for some constant \( C > 0 \), \( |K_c(u) - K_c(v)| \leq C\|u - v\| \) for any \( u, v \in \mathbb{R}^d \).
There exist two sequences of positive integer vectors, $\rho$.

(B1) The bandwidth $\rho$ tends to zero in such a way that $T\rho^d \to \infty$ as $T \to \infty$.

(B2) There is a positive integer $m_T \to \infty$ such that the stability coefficients, defined in (2.3.5) with $\nu = 2$ and $\nu = 2 + \delta/2$, satisfy $T^{2+4/\delta}\rho T^{-(2+2d+4d/\delta)}v_2(m_T) \to 0$ and $\rho (1+2d+2d/\delta)v_2(m_T) = O(1)$ and $\rho T^{-(2+2d+4d/\delta)}v_2(m_T) = O(1)$.

(B3) There exist two sequences of positive integer vectors, $P_T \in \mathbb{Z}$ and $Q_T = 2m_T \in \mathbb{Z}$, with $m_T \to \infty$ such that $P_T = o((T\rho^d)^{1/2})$, $Q_T/P_T \to 0$ and $T/P_T \to \infty$, and $T\rho^{-1}\alpha(m_T) \to 0$.

(B4) $\rho$ tends to zero in such a manner that $Q_T\rho^d = O(1)$ such that $\rho^{-d/(4+\delta)}\sum_{t=Q_T}^{\infty} \{\alpha_{m_T}(t)\}^{\delta/(4+\delta)} \to 0$ as $T \to \infty$.

The above assumptions are standard for this kind of problem; see Lu and Linton (2007).

3.3 Theorems

We first state a proposition, which is needed in establishing the consistency and asymptotic normality for the estimators constructed in the above.

For any constants $k_i$'s, define

$g_T(y) = k_1f_1(y) + k_2f_{21}(y) + \ldots + k_2pf_{2p}(y)$,

$g_T'(y) = k_1f_1(y) + k_2f_{21}(y) + \ldots + k_2pf_{2p}(y)$,

$g(y) = k_1f_1(y) + k_2f_{21}(y) + \ldots + k_2p f_{2p}(y) = k_1\alpha_0 + k_2b_1 + \ldots + k_2b_p$ and

$g'(y) = k_1f_1(y) + k_2f_{21}(y) + \ldots + k_2p f_{2p}(y) = k_1\alpha_1 + k_2b_1 + \ldots + k_2b_p$.

We denote $U := \begin{pmatrix} h(y) \int K(u)du & h(y) \int u^rK(u)du \\ h(y) \int uK(u)du & h(y) \int uu^rK(u)du \end{pmatrix}$,

$\Sigma := Var(k_1X_t + k_2X_{t-1} + \ldots + k_2pX_{t-p} | Y_{t-1} = y)h(y) \begin{pmatrix} \int K^2(u)du & \int u^rK^2(u)du \\ \int uK^2(u)du & \int uu^rK^2(u)du \end{pmatrix}$. 
CHAPTER 3. THEORETICAL STUDIES: LARGE SAMPLE PROPERTY

UNDER NED


d_0(y) := \frac{1}{2} h(y) \sum_{i=1}^{d} \sum_{j=1}^{d} g_{ij}(y) \int u_j u_i K(u) du,

B_1(y) := \frac{1}{2} h(y) \sum_{i=1}^{d} \sum_{j=1}^{d} g_{ij}(y) \int u_j u_i u_i K(u) du,

g_{ij}(y) = \frac{\partial^2 g(y)}{\partial y_i \partial y_j}, i, j = 1, ..., d, and u := (u_1, ..., u_d)^\tau \in \mathbb{R}^d.

Proposition 1. Let Assumptions (A0),(A1),(A2),(A3),(A4) and (A5) hold with
\( h(y) > 0, \quad v_{2+\delta/2}(x) = O(x^{-\mu}), \quad \text{and} \quad \alpha(x) = O(x^{-\lambda}) \) for some \( \mu \geq \max\{4(\kappa_1 - 1), \kappa_3/(1 + \delta/4)\} \kappa_2 \) and some \( \lambda > (a + 1)(1 + 4/\delta) \) with \( a > \delta/(4 + \delta) \), such that
\[ T^{2+4/\delta} \varrho_T^{\mu/\kappa_2-\kappa_1} \to 0, \quad T \varrho_T^{1+2/[a(1+4/\delta)]} \log T \to \infty, \] where \( \kappa_1 = 2 + d + 2d/\delta, \kappa_2 = a(1 + 4\delta)(1 + \delta/4)/d, \) and \( \kappa_3 = 2 + 2d + \delta/2 + 4d/\delta \) and suppose that the bandwidth \( \varrho_T \) satisfies (B1)-(B4).

Then
\[ \sqrt{T \varrho_T^d} \left( \begin{pmatrix} g_T(y) - g(y) \\ \varrho_T(g_T(y) - g'(y)) \end{pmatrix} - U^{-1} \begin{pmatrix} B_0(y) \\ B_1(y) \end{pmatrix} \right) \xrightarrow{L} N(0, U^{-1} \Sigma(U^{-1})^\tau) \]
as \( T \to \infty \).

Now we state the asymptotic normality of the parameter estimator \( \hat{a} \).

Theorem 1. Under Assumptions as specified in Proposition 1, suppose further that
\[ E(\Psi_t \Psi_t^T) = D_t > 0 \] with \( \Psi_t = (X_{t-1} - f_2(Y_{t-1}))W_T(Y_{t-1})\xi_{x,t} \),
\[ \frac{1}{T} \sum_{d+1}^T D_t \to D^*, \quad E|\psi_{it}\psi_{jt}\psi_{lt}\psi_{mt}| < \infty \] for all \( t, i, j, l, \) and \( m \) where \( \psi_{it} \) is the \( i \)-th element of the vector \( \Psi_t \) and \( \frac{1}{T} \sum_{d+1}^T \Psi_t \Psi_t^T \to D^* \) as \( T \to \infty \). Then if the bandwidth \( \varrho_T \) further satisfies \( T \varrho_T^4 \to 0 \) as \( T \to \infty \), we have
\[ \sqrt{T}(\hat{a} - a) \to N(0, D^{-1}D^*D^{-1}) \] as \( T \to \infty \),
CHAPTER 3. THEORETICAL STUDIES: LARGE SAMPLE PROPERTY UNDER NED

where $D$ is the probability limit of $\frac{1}{T} \sum_{d+1}^{T} (X_{t-1} - f_2(Y_{t-1})) (X_{t-1} - f_2(Y_{t-1}))^{T} W_{T}(Y_{t-1})$.

Now we can discuss the asymptotic normality of $\hat{f}(y) = \hat{f}_1(y) - \hat{a}^{T} \hat{f}_2(y)$ from equation (2.4.8) in chapter 2.

**Theorem 2.** Suppose the assumptions same as in Proposition 1 together with the additional conditions in Theorem 1 except with a new bandwidth $\varrho_T$ only satisfying (B1)–(B4) hold true. Then

$$
\sqrt{T\varrho_T} \left( \begin{pmatrix} \hat{f}(y) - f(y) \\ \varrho_T (\hat{f}_T'(y) - f'(y)) \end{pmatrix} - U^{-1} \begin{pmatrix} B_0^*(y) \\ B_1^*(y) \end{pmatrix} \right) \xrightarrow{L} N(0, U^{-1} \Sigma^*(U^{-1})^T)
$$

as $T \to \infty$, where

- $U$ was defined in the above, and
- $\Sigma^* := \text{Var}(X_t - a_1 X_{t-1} - ... - a_p X_{t-p} | Y_{t-1} = y) h(y) \left( \begin{array}{c} \int K^2(u) du \\ \int uu^T K^2(u) du \\ \int uu^T K^2(u) du \end{array} \right)$
- $B_0^*(y) := \frac{1}{2} h(y) \sum_{i=1}^{d} \sum_{j=1}^{d} f_{ij}(y) \int u_j u_i K(u) du$,
- $B_1^*(y) := \frac{1}{2} h(y) \sum_{i=1}^{d} \sum_{j=1}^{d} f_{ij}(y) \int u_j u_i u K(u) du$,
- $f_{ij}(y) = \frac{\partial^2 f(y)}{\partial y_i \partial y_j}, i, j = 1, ..., d$, and $u := (u_1, ..., u_d)^T \in \mathbb{R}^d$.

### 3.4 Proofs

We have approximated the unknown functions $f_1(Y_{t-1}) = E(X_t | Y_{t-1})$ and $f_2(Y_{t-1}) = E(X_{t-1} | Y_{t-1})$ using the local linear fitting idea in the equations (2.4.4) and (2.4.6) and their solutions are in (2.4.5) and (2.4.7) respectively. Followed by Lu & Linton (2007)'s results, it can be shown that the estimator of $f_1$ and $f_2$ has asymptotically normal distribution respectively.
Using Lemma 3.1-3.4 of Lu & Linton (2007), we have the following results (C1)-(C3).

Further, notice that we have

\[ \hat{\alpha}_0 + k_2 \hat{b}_0 = k_1 \hat{\alpha}_0 + k_2 \hat{b}_0, \]

as follows.

Proof of Proposition 1: For any constants \( k_1, k_2, \ldots, k_{2p} \), note that

\[ k_1 \hat{f}_1(y) + k_2 \hat{f}_2(y) + \ldots + k_{2p} \hat{f}_p(y) = k_1 \hat{\alpha}_0 + k_2 \hat{b}_0 + \ldots + k_{2p} \hat{b}_{0p}. \]

Let \((\hat{\alpha}_1, \ldots, \hat{\alpha}_p)\) follow the solutions in (2.4.5) and (2.4.7). Then we have

\[
\begin{pmatrix}
\hat{\alpha}_0 + k_2 \hat{b}_0 \\
k_2 \alpha_1 + k_2 \hat{b}_1 + \ldots + k_{2p} b_{p, j}
\end{pmatrix}
= k_1 U_T^{-1} V_T + k_2 U_T^{-1} \tilde{V}_T + \ldots + k_{2p} U_T^{-1} \tilde{V}_{T_p}
= U_T^{-1} (k_1 V_T + k_2 \tilde{V}_T + \ldots + k_{2p} \tilde{V}_{T_p})
= U_T^{-1} V_T^*,
\]

where \( V_T \) is a \((d + 1)\)-dimensional vector with its \( i \)-th element

\[
(V_T)_i = \left( T \varphi \right)^{-1} \sum_{t=d+1}^T (k_1 X_t + k_2 X_{t-1} + \ldots + k_{2p} X_{t-p}) \left( Y_{t-1} - \bar{y} \right)_i K \left( Y_{t-1} - \bar{y} \right).
\]

Let \((W_T)_i := (T \varphi)^{-1} \sum_{t=d+1}^T Z_t \left( Y_{t-1} - \bar{y} \right)_i K \left( Y_{t-1} - \bar{y} \right), i = 0, \ldots, d, \) with

\[ Z_t = (k_1 X_t + k_2 X_{t-1} + \ldots + k_{2p} X_{t-p}) - (k_1 \alpha_0 + k_2 \alpha_1 + \ldots + k_{2p} b_{0p}) - (k_1 \alpha_1 + k_2 \alpha_2 + \ldots + k_{2p} b_{p}) \bar{y}. \]

Using Lemma 3.1-3.4 of Lu & Linton (2007), we have the following results (C1)-(C3).

(C1) \( (T \varphi)^{1/2} (W_T - EW_T) \) is asymptotically normal.

(C2) \( (T \varphi)^{1/2} (EW_T - \mu_T) \to 0 \), where \( \mu_T = h(y)(g(y), \varphi_T g'(y))^\tau \), and

\[ Var((T \varphi)^{1/2} W_T) \to \Sigma \]

for some positive definite matrix \( \Sigma \), and

(C3) \( U_T \to U \) for some nonsingular matrix \( U \).
CHAPTER 3. THEORETICAL STUDIES: LARGE SAMPLE PROPERTY UNDER NED

Then, by Theorem 3.1 of Lu & Linton (2007)’s we have

\[
\sqrt{T} g^D_T \left( \begin{pmatrix} g_T(y) - g(y) \\ g_T(g_T(y) - g'(y)) \end{pmatrix} - U^{-1} \begin{pmatrix} B_0(y) \\ B_1(y) \end{pmatrix} \right) \xrightarrow{L} N(0, U^{-1} \Sigma (U^{-1})^T)
\]
as \(T \to \infty\).

Proof of theorem 1.

From the equation (2.4.2), it follows that the least squares estimator \(\hat{a}\) can be expressed as

\[
\hat{a} = \left[ \frac{1}{T} \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) \right]^{-1} \\
\times \left[ \frac{1}{T} \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_t - \hat{f}_1(Y_{t-1})) W_T(Y_{t-1}) \right], \tag{3.4.1}
\]

where \(W_T(Y_{t-1}) = I(A_T \leq Y_{t-1} \leq B_T)\), with \(A_T \to \infty\) and \(B_T \to \infty\) suitably chosen as in Li, Lu & Linton (2012).

Noting that \(f_1(Y_{t-1}) = E(X_t|Y_{t-1})\) and \(f_2(Y_{t-1}) = E(X_{t-1}|Y_{t-1})\), we then have
\[ \hat{a} - a = \left[ \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) \right]^{-1} \]

\[ \times \left[ \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_t - \hat{f}_1(Y_{t-1})) W_T(Y_{t-1}) \right] \]

\[ - \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) a \]

\[ = \left[ \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) \right]^{-1} \]

\[ \times \left[ \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) a \right] \]

\[ = \left[ \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) \right]^{-1} \]

\[ \times \left[ \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) \right] \]

\[ = \left[ \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) \right]^{-1} \]

\[ \times \left[ \sum_{t=d+1}^{T} (X_{t-1} - \hat{f}_2(Y_{t-1}))(X_{t-1} - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1}) \right] \]

\[ \times (f_1(Y_{t-1}) - \hat{f}_1(Y_{t-1})) - (f_2(Y_{t-1}) - \hat{f}_2(Y_{t-1}))^T a + \epsilon_{x,t} \]

\[ \times W_T(Y_{t-1}) \]

\[ = \left[ \sum_{t=d+1}^{T} (X_{t-1} - f_2(Y_{t-1}) + f_2(Y_{t-1}) - \hat{f}_2(Y_{t-1})) \right] \]

\[ \times (X_{t-1} - f_2(Y_{t-1}) + f_2(Y_{t-1}) - \hat{f}_2(Y_{t-1}))^T W_T(Y_{t-1})^{-1} \]

\[ \times \left[ \sum_{t=d+1}^{T} (X_{t-1} - f_2(Y_{t-1}) + f_2(Y_{t-1}) - \hat{f}_2(Y_{t-1})) \right] \]

\[ \times (f_1(Y_{t-1}) - \hat{f}_1(Y_{t-1}) + (f_2(Y_{t-1}) - f_2(Y_{t-1}))^T a + \epsilon_{x,t}) W_T(Y_{t-1}) \]  \quad (3.4.2)

Due to the uniform consistency of \( \hat{f}_1 \) and \( \hat{f}_2 \) (c.f., Li, Lu & Linton (2012)) and the condition \( T \hat{\varrho}_1 \to 0 \) that controls the bias coming from estimating \( f_1 \) and \( f_2 \), it can be easily showed that

\[ \sqrt{T}(\hat{a} - a) = \left[ \frac{1}{T} \sum_{t=d+1}^{T} (X_{t-1} - f_2(Y_{t-1}))(X_{t-1} - f_2(Y_{t-1}))^T W_T(Y_{t-1}) \right]^{-1} \]

\[ \times \left[ \frac{\sqrt{T}}{T} \sum_{t=d+1}^{T} (X_{t-1} - f_2(Y_{t-1})) W_T(Y_{t-1}) \epsilon_{x,t} \right] + o_p(1) \]  \quad (3.4.3)
CHAPTER 3. THEORETICAL STUDIES: LARGE SAMPLE PROPERTY

as $T \to \infty$.

We know that $\frac{1}{T} \sum_{t=d+1}^{T} (X_{t-1} - f_2(Y_{t-1}))(X_{t-1} - f_2(Y_{t-1}))^{T}W_{T}(Y_{t-1})$ goes to $E[(X_{t-1} - f_2(Y_{t-1}))(X_{t-1} - f_2(Y_{t-1}))^{T}W_{T}(Y_{t-1})] \to D = E[(X_{t-1} - f_2(Y_{t-1}))(X_{t-1} - f_2(Y_{t-1}))^{T}]$ as $T \to \infty$ by the law of large numbers as well as $A_T \to \infty$ and $B_T \to \infty$.

Let $\Psi_t = (X_{t-1} - f_2(Y_{t-1}))W_T(Y_{t-1})\epsilon_{x,t}$. Since we have supposed $E(\epsilon_{x,t}|X_{t-1}, Y_{t-1}) = 0$, we have

$$E(\Psi_t|X_{t-1}, Y_{t-1}) = (X_{t-1} - f_2(Y_{t-1}))W_T(Y_{t-1})E(\epsilon_{x,t}|X_{t-1}, Y_{t-1}) = 0. \quad (3.4.4)$$

Therefore, $\Psi_t$ is a martingale difference sequence. Because $E(\Psi_t, \Psi_t^T) = D_t > 0$,

$$\frac{1}{T} \sum_{d+1}^{T} D_t \to D^* > 0, E|\psi_{it}\psi_{jt}\psi_{lt}\psi_{mt}| < \infty \text{ for all } t, \text{ and } i, j, l, \text{ and } m$$

where $\psi_{it}$ is the $i$-th element of the vector $\Psi_t$ and $\frac{1}{T} \sum_{d+1}^{T} \Psi_t \Psi_t^T \to D^*$,

$$\frac{1}{\sqrt{T}} \sum_{d+1}^{T} \Psi_t \to N(0, D^*) \quad (3.4.5)$$

by the central limit theorem.

Combining the above, we finally obtain

$$\sqrt{T}(\hat{a} - a) \longrightarrow N(0, D^{-1}D^*D^{-1}) \quad \text{as} \quad T \to \infty.$$

The proof is completed.

Proof of theorem 2.

Note that if we take $k_1, k_{21}, ..., k_{2p}$ in proposition 1 equal to 1, $a_1, ..., a_p$, then we get $g(y) = f(y) = f_1(y) - a^T f_2(y)$, $g_T(y) = \hat{f}_1(y) - a^T \hat{f}_2(y)$. In addition note that by Theorem 1, $\hat{a} - a = O_P(T^{-1/2}) = o(T g_2^d)^{-1/2}$, and hence $\hat{f}(y) = \hat{f}_1(y) - \hat{a}^T \hat{f}_2(y) = g_T(y) + o(T g_2^d)^{-1/2}$. Thus

$$\hat{f}(y) - f(y) = g_T(y) - g(y) + o(T g_2^d)^{-1/2}.$$
Then Theorem 2 follows from Proposition 1.
Chapter 4

Simulation studies: Finite sample performance

4.1 Introduction

In this chapter, we are investigating the finite-sample performance of the estimation method described in the previous chapter when estimating the unknown function and parameters in a semiparametric model under near epoch dependence by using Monte Carlo simulation studies. To examine how well we can estimate both the parameters and the unknown function regardless of the form of the actual unknown function, we will consider both cases where the unknown function $f$ is linear and nonlinear. The data generating processes satisfy near epoch dependence. We take the sample sizes equal to 100 and 500, respectively.

4.2 Case of Non-linear Granger-causality

We first examine the case where the actual function $f$ is nonlinear.
(Simulation model) 1. Consider the semiparametric model

\[
\begin{align*}
X_t &= a_1 X_{t-1} + \sin(Y_{t-1}) + \epsilon_t \\
Y_t &= a Y_{t-1} + \epsilon_t + b \epsilon_{t-1},
\end{align*}
\]

where \( \epsilon_t \sim N(0, \sigma^2_\epsilon) \), \( e_t = \tilde{\epsilon}_t h_t^{1/2} \), \( \tilde{\epsilon}_t \sim N(0, 1) \) and \( h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \).

\( X_t \) and \( Y_t \) can be thought of as weekly returns in week \( t \) for two different stocks and \( h_t \) is the conditional variance of \( Y_t \), given the past information up to week \( t-1 \). The coefficient of \( X_{t-1} \), \( a_1 \), is an unknown parameter and \( f(Y_{t-1}) \) is an unknown function. Parameters \( a \) and \( b \) are the coefficients in the ARMA(1,1) model, and \( \alpha_0, \alpha_1 \) and \( \beta_1 \) are the coefficients in the GARCH(1,1) model, with \( \tilde{\epsilon}_t \) being i.i.d. random sequence with mean 0 and variance 1 and \( \alpha_0 > 0, \alpha_1 \geq 0, \beta_1 \geq 0 \). This is a special case of the general ARMA(p,q)-GARCH(r,m) model so it satisfies near epoch dependence.

We set the parameters \( a_1 = 0.6, a = -0.1, b = 0.04, \alpha_0 = 0.13, \alpha_1 = 0.09 \) and \( \beta_1 = 0.72 \). \( \epsilon_t \) is generated from \( N(0, 0.01) \). In this model, \( \sin(Y_{t-1}) \) is used for the unknown function \( f(Y_{t-1}) \). The process \( Y_t \) is generated by (4.2.2) and \( X_t \) is generated by (4.2.1) using the function \( f(Y_{t-1}) = \sin(Y_{t-1}) \). We generate the simulated time series of size \( T = 100 \) and \( T = 500 \), respectively. The bandwidth \( \varrho_T \) in this simulations study was chosen by the conventional generalized cross-validation rule. We repeat the simulation 100 times.

The Figure (4.2.1) shows the estimation results for nonlinear function \( f \). The top graph is for the case of sample size \( T = 100 \) and the bottom one for \( T = 500 \), respectively. The solid line is the true values for the function \( f(Y_{t-1}) = \sin(Y_{t-1}) \). The long-dashed line stands for the mean of the estimated values of 100 replications for each case. Two dotted lines are 95% confidence band. From these graphs, we can see that our asymptotic theory applied to Simulation model 1 works very well. It is
Figure 4.2.1: Monte carlo simulation: sample size $T = 100$ (Top) and $T = 500$ (Bottom), with the theoretical function $f(Y_{t-1}) = \sin(Y_{t-1})$. 
clear that the mean of the estimated values of the unknown function $f$ are close to the true function for both sample size cases. As expected, the 95% confidence bands get narrower as the sample size increases from 100 to 500, which shows consistency of $\hat{f}$.

We can get an evidence on the consistency of $\hat{a}$ from the mean of the absolute error $|\hat{a} - a|$ for the 100 replications. We obtain the mean of $|\hat{a} - a|$ equal to 0.011 for the sample size $T = 100$ and 0.0043 for the sample size $T = 500$. Figure (4.2.2) shows box-plots of $|\hat{a} - a|$ of $T = 100$(Left) and $T = 500$(Right) cases. It shows the consistency of $\hat{a}$ for the nonlinear case.

Figure 4.2.2: Boxplot of $|\hat{a} - a|$ for the nonlinear case.
4.3 Case of Linear Granger-causality

Nishiyama et al. (2011) found their Granger-causality testing method works well in the case of nonlinear dependence but was worse, when the theoretical series has linear dependence in mean, in comparison with Hidalgo (2000)’s test.

In this section we are examining how about the performance of our semiparametric method when the actual unknown function is linear. We conduct similar Monte Carlo studies by the following model.

(Simulation model) 2. Consider the linear model

\[ X_t = a_1 X_{t-1} + (1 - 2Y_{t-1}) + \epsilon_t \]  
\[ Y_t = aY_{t-1} + \epsilon_t + be_{t-1}, \]  

(4.3.1) 
(4.3.2)

where \( \epsilon_t \sim N(0, \sigma^2) \), \( \epsilon_t = \hat{\epsilon}_t h_t^{1/2} \), \( \hat{\epsilon}_t \sim N(0, 1) \) and \( h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \).

The simulation model 2 is also a special case of ARMA(p,q)-GARCH(r,m) model, so the generated processes \( X_t \) and \( Y_t \) are near epoch dependent. The simulation is done as in the last section, with the outcomes reported in Figure (4.3.1). The sample sizes are \( T = 100 \) (Top) and \( T = 500 \) (Bottom), respectively. Clearly we can see that our theory also works very well for the linear function case. Here the solid line is the true values for the function \( f(Y_{t-1}) = 1 - 2 \times (Y_{t-1}) \). The mean of the estimated values of 100 replications is represented by the long-dashed line for each case and two dotted lines are 95% confidence band. It seems that our asymptotic theory applied to simulation model 2 works very well like in the nonlinear case. In linear case, the mean of the estimated values of the unknown function \( f \) is also close to the true linear function for both sample size cases. The 95% confidence bands are narrower for the sample size 500 than 100. The results show the consistency of \( \hat{f} \).
Figure 4.3.1: Monte carlo simulation: sample size $T = 100$ (Top) and $T = 500$ (Bottom), with the theoretical function $f(Y_{t-1}) = 1 - 2 \times (Y_{t-1})$. 
Let us check the mean of $|\hat{a} - a|$ for the 100 replications. The mean of $|\hat{a} - a|$ is equal to 0.00040 for the sample size 100 and 0.00020 for the sample size 500. Figure (4.3.2) shows box-plots of $|\hat{a} - a|$ of $T = 100$(Left) and $T = 500$(Right) cases. It shows the consistency of $\hat{a}$ for the linear case.
Chapter 5

Application to real data: Cross-country empirical evidence of Granger-causality

5.1 Introduction

In this chapter, we are empirically demonstrating the application of our proposed semiparametric time series regression method to the analysis of Granger-causality among the financial markets of Australia and its economically close foreign countries including the USA, the UK and China. We use the typical stock price indices data from the 4 countries of Australia, USA, UK and China to investigate if there exist any Granger causalities between the stock markets of each two countries. The partially linear regression model and related theories we developed in this paper are applied to these data sets to capture the existence or structure of nonlinear Granger causality. Six different pairs of the four countries are considered.
5.2 Data

The typical stock market indices from Australia, the USA, the UK and China have been obtained from finance.yahoo.com. The S&P500 index is a market capitalisation-weighted index that tracks the daily total return performance of 500 leading companies in leading industries of the U.S. The S&P500 accounts for about 75% of the market value of the shares of the U.S. equities. For the UK we use the Financial Times-Stock Exchange (FTSE) 100 index which is a share index of the 100 most highly capitalised UK companies listed on the London Stock Exchange. FTSE100 companies represent about 81% of the market capitalisation of the whole London Stock Exchange. We take ASX200 index for the Australian market, which covers approximately 78% of Australian equity market capitalization. The Shanghai Stock Exchange (SSE) composite index is an index of the stock exchange that is based in the city of Shanghai, which is the most typical index in China.

We are therefore using FTSE100, S&P500, ASX200 and SSE indices. Weekly data are used to avoid market micro structure problems. The period is from 15th of October 2001 to 19th of May 2008 and the sample size is 345. Figures (5.2.1), (5.2.2), (5.2.3) and (5.2.4) are the plots of the raw data (top panel) and the returns (bottom panel) on stock indices for S&P500, FTSE100, ASX200 and SSE, respectively, where for a stock index $x_t$, its return series is defined as $X_t = \log(x_t/x_{t-1}) \times 100$.

We are concerned with the relationship of the 4 markets in terms of the Granger causality of each pair of the return series. The descriptive statistics of the data sets are provided in Table 5.2.1. From the weekly return perspective, it looks that the Australian stock market is less volatile, with the standard deviation of its return series equal to 1.61, than the USA and the UK stock markets, whose return standard deviations equal 1.99 and 1.95, respectively, while the Chinese stock market is the most volatile with the return standard deviation as large as 3.45.
Table 5.2.1: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>FTSE 100 Returns</th>
<th>S&amp;P 500 Returns</th>
<th>ASX 200 Returns</th>
<th>SSE Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>345</td>
<td>345</td>
<td>345</td>
<td>345</td>
</tr>
<tr>
<td>Minimum</td>
<td>3491.6</td>
<td>-8.86</td>
<td>800.58</td>
<td>-8.33</td>
</tr>
<tr>
<td>Maximum</td>
<td>6732.4</td>
<td>6.95</td>
<td>1561.8</td>
<td>7.23</td>
</tr>
<tr>
<td>Mean</td>
<td>5166.54</td>
<td>0.067</td>
<td>1190.15</td>
<td>0.083</td>
</tr>
<tr>
<td>Std.Dev</td>
<td>860.47</td>
<td>1.95</td>
<td>185.89</td>
<td>1.99</td>
</tr>
</tbody>
</table>

5.3 Semiparametric regression analysis of Granger-causality

We will apply our semiparametric model method to examine if there exists any kind of Granger-causality between AUS & UK, AUS & USA, AUS & China, UK & USA, UK & China and USA & China. As we discussed before, in practice the dynamics of stock returns are better explained by ARMA(p,q)-GARCH(r,m) processes, which satisfy the near epoch dependence properties.

For each pair of the 4 countries, we consider the following semiparametric time series regression model:

\[ X_t = a^T X_{t-1} + f(Y_{t-1}) + \epsilon_{X,t}, \]

where \( X_t \) and \( Y_t \) are the return series for the index series \( x_t \) and \( y_t \) of each pair of the 4 countries, respectively; that is, \( X_t = \log(x_t / x_{t-1}) \times 100 \) and \( Y_t = \log(y_t / y_{t-1}) \times 100 \). Here \( X_{t-1} = (X_{t-1}, \cdots, X_{t-p}) \), and \( Y_{t-1} \) consists of some lag variables of \( Y_t \).

In order to have a reasonable order \( p \) of lags for the linear part of the model (5.3.1), we simply applied the well-known AIC to each rerun series. It was found that for the Chinese data, lag 3 was preferred by the AIC and we therefore used \( p = 3 \) lagged terms for the SSE return series, while for other countries their stock markets appear quite efficient with lag zero chosen by the AIC and we therefore used only \( p = 1 \) lag term for the return series of FTSE100, S&P500 and ASX200.
CHAPTER 5. APPLICATION TO REAL DATA: CROSS-COUNTRY
EMPIRICAL EVIDENCE OF GRANGER-CAUSALITY

For possible non-linear Granger-causality from $Y_t$ to $X_t$ in (5.3.1), we look into two cases of $Y_{t-1}$, where we use $f(Y_{t-1})$ and $f(Y_{t-2})$ for the function $f$ in equation (5.3.1) to see if there is any Granger causality from $Y_{t-1}$ or $Y_{t-2}$.

Similarly, for $Y_{t-1} = (Y_{t-1}, \ldots, Y_{t-p})$, we can estimate

$$Y_t = b^T Y_{t-1} + g(X_{t-1}) + \epsilon_{Y,t},$$

(5.3.2)

to look into the possible non-linear Granger-causality from $X_t$ to $Y_t$, where two cases

![S&P 500](image1.png)

15 Oct 2001~19 May 2008

![Changes in log(S&P 500)](image2.png)

15 Oct 2001 -19 May 2008

Figure 5.2.1: Weekly Data and Changes in log of S&P 500
Figure 5.2.2: Weekly Data and Changes in log of FTSE 100
CHAPTER 5. APPLICATION TO REAL DATA: CROSS-COUNTRY
EMPIRICAL EVIDENCE OF GRANGER-CAUSALITY

Figure 5.2.3: Weekly Data and Changes in log of ASX 200

ASX 200

Changes in log(ASX 200)
Figure 5.2.4: Weekly Data and Changes in log of SSE
of $X_{t-1}$ are considered, as before, with $g(X_{t-1})$ and $g(X_{t-2})$ used for the function $g$ in equation (5.3.2) to see if there is any Granger causality from $X_{t-1}$ or $X_{t-2}$.

### 5.4 Empirical evidence of Granger-causality

Using these data, we have estimated unknown functions $f$ and $g$ and parameters $a$ and $b$ in the semiparametric model (5.3.1) and (5.3.2) for each pair of countries.

#### 5.4.1 FTSE 100 and ASX 200

For the case of FTSE 100 and ASX 200, $X_t$ is percentage changes in FTSE 100 and $Y_t$ is percentage changes in ASX 200.

Firstly, let us look at the case where we use $Y_{t-1}$ for $f$ and $X_{t-1}$ for $g$. Figure (5.4.1) shows the estimation results of $f(Y_{t-1})$ and $g(X_{t-1})$. In the top plot, the solid line is estimator of function $f$ and the long-dashed lines are 95% confidence band. Here the 95% confidence band is constructed by applying bootstrap method using GARCH(1,1) process for the variance of error.

Since the estimate of $f$ is close to zero and the 95% confidence band includes x-axis for all values of $Y_{t-1}$, $Y_t$(ASX 200) does not seem to Granger-cause $X_t$(FTSE 100). On the contrary, the bottom plot shows a strong evidence for the existence of highly nonlinear Granger-causality from FTSE 100 to ASX 200 because the estimate of function $g$ is highly nonlinear and significantly very different from zero for most of the values of $X_{t-1}$.

Interestingly, when the FTSE 100 return is in the range of (-4%, -3%), if the FTSE 100 return increases ASX200 return decreases. But if the FTSE 100 return is between -1% and 2%, changes in the FTSE 100 return have constant and relatively small effect on ASX 200. For other values of FTSE 100, we can see there is positive and nonlinear Granger causality from UK stock market to Australia stock
CHAPTER 5. APPLICATION TO REAL DATA: CROSS-COUNTRY
EMPIRICAL EVIDENCE OF GRANGER-CAUSALITY

Figure 5.4.1: \( f(Y_{t-1}) \) and \( g(X_{t-1}) \) for UK vs AUS market. The estimates of parameters (95% confidence interval) \( a \) and \( b \) are respectively 0.074([-0.012, 0.17]) and -0.12([-0.22, -0.012]).

Next, when we used \( Y_{t-2} \) for \( f \) and \( X_{t-2} \) for \( g \), we found little evidence of Granger causality from either direction. This can be seen from Figure (5.4.2). The estimates of parameters (95% confidence interval) \( a \) and \( b \) are respectively 0.0099([-0.093, 0.13]) and 0.055([-0.060, 0.14]).
CHAPTER 5. APPLICATION TO REAL DATA: CROSS-COUNTRY
EMPIRICAL EVIDENCE OF GRANGER-CAUSALITY

5.4.2 S&P 500 and ASX 200

Let us see the estimation results of S&P 500 and ASX 200 case. Here $X_t$ is percentage changes in S&P 500 and $Y_t$ is percentage changes in ASX 200.

If we estimate $f(Y_{t-1})$ and $g(X_{t-1})$, then we can get the results in Figure (5.4.3). Similarly to the case of FTSE 100 and ASX 200, we have found no evidence for the existence of Granger causality from ASX 200 to S&P 500 because the 95% confidence band includes the $x$-axis for all values of $Y_{t-1}$ as can be seen from the top panel of Figure (5.4.3). The bottom panel of Figure (5.4.3) depicts estimation results of function $g$ in (5.3.2). Again, there is a strong evidence of nonlinear Granger causality from S&P 500 to ASX 200. While there is linear relationship when the
S&P 500 return is around zero, for other values of S&P 500, the S&P 500 nonlinearly Granger cause the ASX 200. The estimates of parameters (95% confidence interval) $a$ and $b$ are 0.048([-0.064, 0.16]) and -0.097([-0.20, 0.020]), respectively.

If we use $Y_{t-2}$ for $f$ and $X_{t-2}$ for $g$, it is hard to find an evidence of Granger causality from either direction. This can be seen from Figure (5.4.4). The estimates of parameters (95% confidence interval) $a$ and $b$ are respectively -0.0077([-0.13, 0.088]) and 0.0051([-0.10, 0.11]).
CHAPTER 5. APPLICATION TO REAL DATA: CROSS-COUNTRY
EMPIRICAL EVIDENCE OF GRANGER-CAUSALITY

Figure 5.4.4: \( f(Y_{t-2}) \) and \( g(X_{t-2}) \) for USA vs AUS

5.4.3 FTSE 100 and S&P 500

We estimated the pair of FTSE 100 and S&P 500. Let \( X_t \) be percentage changes in FTSE 100 and \( Y_t \) be percentage changes in S&P 500.

Let us see the case of \( f(Y_{t-1}) \) and \( g(X_{t-1}) \) first. The top and bottom panels of Figure (5.4.5) show estimation results of function \( f \) and \( g \), respectively. As we can see from these, it appears to be no Granger-causality between UK and USA stock markets in either direction. The estimates of parameters (95% confidence interval) \( a \) and \( b \) are 0.038([-0.069, 0.13]) and -0.11([-0.22, 0.0025]), respectively.
For the case of \( f(Y_{t-2}) \) and \( g(X_{t-2}) \), there seems to be no Granger causality between USA and UK. Figure (5.4.6) shows the results. The estimates of parameters (95% confidence interval) \( a \) and \( b \) are respectively 0.034([-0.076, 0.15]) and -0.0090([-0.084, 0.099]).
CHAPTER 5. APPLICATION TO REAL DATA: CROSS-COUNTRY

EMPIRICAL EVIDENCE OF GRANGER-CAUSALITY

Figure 5.4.6: \( f(Y_{t-2}) \) and \( g(X_{t-2}) \) for USA vs UK

5.4.4 FTSE 100 and SSE

We have estimated the pair of FTSE 100 and SSE. Let \( X_t \) be percentage changes in SSE and \( Y_t \) be percentage changes in FTSE 100.

Figure (5.4.7) represents the estimation results of \( f(Y_{t-1}) \) and \( g(X_{t-1}) \). When we look at the top graph, if the FTSE 100 return is between 1% and 2.5%, there is an evidence of positive linear Granger causality from FTSE 100 to SSE. There seems to be no impact from FTSE 100 to SSE for other values of FTSE 100. From the bottom panel, it is hard to say there is Granger causality from SSE to FTSE 100. The estimates of parameters (95% confidence interval) are \( a = (0.037, 0.083, 0.10) ([-0.079, 0.15], [-0.03, 0.17], [-0.032, 0.17]) \) and \( b = 0.033 \) \( ([-0.057, 0.14]) \), respectively.
CHAPTER 5. APPLICATION TO REAL DATA: CROSS-COUNTRY
EMPIRICAL EVIDENCE OF GRANGER-CAUSALITY

Figure 5.4.7: $f(Y_{t-1})$ and $g(X_{t-1})$ for UK vs China
Figure 5.4.8: $f(Y_{t-2})$ and $g(X_{t-2})$ for UK vs China

Let us see the case of $f(Y_{t-2})$ and $g(X_{t-2})$ from Figure (5.4.8). Since the 95% confidence band clearly includes x-axis for all values of $Y_{t-2}$ (top panel) and $X_{t-2}$ (bottom panel), there seems to be no relationship between China and UK. The estimates of parameters (95% confidence interval) are $a = (0.036, 0.062, 0.10)$ ($[-0.074, 0.12], [-0.050, 0.14], [-0.036, 0.21]$) and $b = -0.0090([−0.084, 0.099])$.

### 5.4.5 S&P 500 and SSE

Next, let us investigate the relationship between S&P 500 and SSE. Let $X_t$ be percentage changes in SSE and $Y_t$ be percentage changes in S&P 500.
The results of the estimation of $f(Y_{t-1})$ and $g(X_{t-1})$ are similar to those for the case of FTSE 100 and SSE. From the top panel of Figure (5.4.9), we can also see the positive linear Granger causality from S&P 500 to SSE when the S&P 500 return is between 1% and 3% and no Granger causality from S&P 500 to SSE for other values of S&P 500. From the bottom panel, there seems to be no Granger causality from SSE to S&P 500. The estimates of parameters (95% confidence interval) are $a = (0.058, 0.096, 0.11)([-0.061, 0.178], [0.0053, 0.19], [-0.0092, 0.20])$ and $b = -0.001131623([-0.11, 0.092])$, respectively.

Figure (5.4.10) shows the estimation results for $f(Y_{t-2})$ and $g(X_{t-2})$. It is difficult
to say that there is any relationship between China and USA. The estimates of parameters (95% confidence interval) are $a = (0.038, 0.063, 0.11)$
($[-0.071, 0.15], [-0.057, 0.17], [0.0061, 0.21]$) and $b = -0.01267341([-0.13, 0.090])$.

### 5.4.6 ASX 200 and SSE

Finally, we can have a look at the relationship between ASX 200 and SSE. Let $X_t$ be percentage changes in SSE and $Y_t$ be percentage changes in ASX 200.

The results of the estimation of $f(Y_{t-1})$ and $g(X_{t-1})$ can be seen in Figure (5.4.11). It is hard to say that Granger causality from ASX 200 to SSE exists
when we see the top panel, since the 95% confidence band clearly includes x-axis for all values of $Y_{t-1}$. However, we can see some negative linear Granger causality when SSE return is between -6% and 0%. The estimates of parameters (95% confidence interval) are $a = (0.030, 0.062, 0.11)([-0.094, 0.14], [-0.051, 0.17], [0.024, 0.20])$ and $b = 0.060([-0.063, 0.15])$, respectively.

Figure (5.4.12) shows the estimation results for $f(Y_{t-2})$ and $g(X_{t-2})$. From the bottom graph, we can see some nonlinear Granger causality from SSE to ASX 200 when SSE return is around 0. However, there seems to be no Granger causality from ASX 200 to SSE. The estimates of parameters (95% confidence interval) are $a =$
5.5 Summary

Interestingly, we could see that there is a strong nonlinear Granger causality from FTSE 100 to ASX 200, some linear or nonlinear Granger causality from S&P 500 to ASX 200 and some negative linear Granger causality from SSE to ASX 200. From these results, we can see that the USA, the UK and the Chinese stock prices 1 week ago help to predict, and hence have important impacts on, current Australian stock market behavior. There is also some linear Granger causality from the FTSE 100
and the S&P 500 to the SSE, respectively, but not from the ASX200 to the SSE. We could not find any evidence of Granger causalities from the other country to either the UK or the USA in the above analysis.

We also could not find any evidence of Granger causality between the countries when we look at the effect of lag 2, $Y_{t-2}$, except the (Australia, China) case. It appears that the Chinese stock price 2 weeks ago is helpful to predict this week’s Australia stock price but not the vice versa.
Chapter 6

Conclusion

In this thesis, we have suggested a semiparametric regression model under the framework of near epoch dependent process to explore the existence and structure of nonlinear Granger-causality. Applying the idea of Robinson (1988) and Lu & Linton (2007) to our model, we can estimate the parameters and unknown function. We have also showed the consistency and asymptotic normality of the estimators. With the estimate of the unknown function, we can know if there is any Granger-causality between the variables or not. Also, the form or structure of nonlinear Granger-causality can be found.

The Monte Carlo simulation studies that were conducted have showed that our method works well whenever the actual Granger-causality is linear or nonlinear in finite sample. The estimates of the parameters and unknown function are accurate for both linear and nonlinear cases when the sample size is greater than 100.

We have applied our proposed method to the analysis of the stock market indices for Australia, the USA, the UK and China. The estimation results obtained have demonstrated that there is a strong evidence of nonlinear Granger causality from the FTSE 100 to the ASX 200, some linear or nonlinear Granger causality from the S&P 500 to the ASX 200 and some negative linear Granger causality from the
SSE to the ASX 200. From these results, we can see that the USA, the UK and the Chinese stock prices have important impacts on the Australian economy, where the stock prices from these countries 1 week ago can help to predict the current Australian stock market behavior. There is also some evidence of linear Granger causality from the FTSE 100 and the S&P 500 to the SSE but not from the ASX 200 to the SSE. In addition, we have interestingly found that the Chinese stock price 2 weeks ago appears to help to predict this week’s Australia stock price but not the vice versa.

Differently from the nonparametric methods used in the literature for testing of nonlinear Granger causality, our proposed semiparametric time series regression analysis reduces the curse of dimensionality that a purely nonparametric method suffers from. For example, in our model (2.2.1), we allow a large lag order \( p \) for \( X_t \) if only the dimension \( d \) in the lag vector of \( Y_t \) is suitably small as done in Chapter 5. When the dimension \( d \) is also large (say, larger than 3), we may need further semiparametric structures (c.f., Gao 2007), such as an additive one, for the unknown function \( f \) in (2.2.1), under the framework of near epoch dependence. We leave this for future work.
Bibliography


Lu, Z. (1996a), ‘A note on geometric ergodicity of autoregressive conditional het-


