

# Essays in Econometrics and Robust Control

By

**Hon Ho Kwok**

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2012

Date of final oral examination: May 24, 2012

The dissertation is approved by the following members of the Final Oral  
Committee:

Steven N. Durlauf, Vilas Professor, Laurents R. Christensen Professor of  
Economics, Kenneth J. Arrow Professor of Economics, Department of  
Economics

William A. Brock, Vilas Research Professor of Economics, Department of  
Economics

Xiaoxia Shi, Assistant Professor, Department of Economics

Chao Fu, Assistant Professor, Department of Economics

Michael Gofman, Assistant Professor, School of Business

© Copyright by Hon Ho Kwok 2012  
All Rights Reserved

To my parents

# Contents

<b>Acknowledgements</b>	<b>iv</b>
<b>Abstract</b>	<b>vi</b>
<b>1 Identification Problems of Linear Social Interaction Models: A General Analysis Based on Matrix Spectral Decompositions</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Disentangling Exogenous and Endogenous Interactions . . . . .	6
1.2.1 Network Terminology . . . . .	6
1.2.2 The Problem of Disentangling Social Interactions . . . . .	6
1.2.3 Spectral Decompositions . . . . .	8
1.2.4 Decompositions of Social Interactions . . . . .	9
1.2.5 Non-identification and Network Structure . . . . .	14
1.3 Differencing . . . . .	18
1.3.1 Information Loss from Differencing . . . . .	18
1.3.2 Local Differencing . . . . .	20
1.3.3 Information Loss-minimizing Differencing . . . . .	21
1.3.4 An Example: Interactions in Circulant Networks . . . . .	22
1.4 Disentangling the Effects of Correlated Unobservables and Endogenous Interactions . . . . .	24
1.4.1 Covariance Matrix Decompositions . . . . .	24
1.4.2 An Example: Graham's Linear-in-means Model . . . . .	25
1.5 Conclusions . . . . .	29
1.6 Appendix: Proofs of Theorems . . . . .	30
1.7 Appendix: Covariance Matrix Decompositions with Differencing	35
1.8 Appendix: Linear-in-means models . . . . .	38
<b>2 Estimating Covariance Matrices with Correlated Disturbances</b>	<b>45</b>
2.1 Introduction . . . . .	45
2.2 Estimating Covariance Matrices . . . . .	46
2.3 Appendix: Corrected Lemma 6.19 of White (1984) . . . . .	50

<b>3 Robust Ramsey Taxation under Implementation Lag Uncertainty</b>	<b>53</b>
3.1 Introduction . . . . .	53
3.2 Ramsey Taxation . . . . .	55
3.3 Time-consistency . . . . .	56
3.4 Linear-quadratic Approximation . . . . .	57
3.5 Robust Ramsey Taxation . . . . .	59
3.6 Equilibrium . . . . .	62
3.7 Purely Stochastic Exogenous Variables . . . . .	63
3.8 General Exogenous Variables . . . . .	66
3.9 Dynamics of Allocations . . . . .	68
3.10 Remarks . . . . .	71
3.11 Appendix: Steady States . . . . .	72
3.12 Appendix: Steady States with Log Utility . . . . .	74

# Acknowledgements

First, I would like to express my gratitude to Professor Steven N. Durlauf, my main advisor, for his encouragement, support, and help on research. I thank for his effort and time on promoting me on the job market. Professor Durlauf's series of study groups, his ECON 836 (Applied macroeconomic analysis) in fall 2008, and his series of voluntary lectures on social interactions in fall 2011 gave me solid knowledge in macroeconomics, model uncertainty, robust control, social interactions, networks, econometrics, and statistics. He organized a series of study groups on various fields: decision theory and robust control in fall 2009 and spring 2010, social interactions and networks in fall 2010, and econometrics and statistics in spring 2011. He has diverse knowledge in philosophy and sciences; his series of study groups gave me an excellent opportunity to learn his ways of thinking; in particular, I learnt to focus on ideas instead of technical details from him. The training in the series of study groups was an important intellectual development process in my life; that two-year period was a transition period from a student to a researcher. Without the study group on statistics and econometrics in spring 2011, my job market paper (chapter 1) would not have been possible. One of the important papers that inspired the ideas in my job market paper: Speed (1987) is one of the suggested readings of that study group. In November 2011, he suggested me to work on random graphs and asymptotic theory and to read White (1984, 2001); chapter 2 is a preliminary work on these topics. The robust Ramsey taxation project (chapter 3) is a consequence of his suggestion to consider policy lags in spring 2008.

Second, I would like to thank Professor William A. Brock for his encouragement, support, and help on research and the job market. Professor Brock's ECON 606 (Mathematical economics II) in spring 2008 and spring 2009, and ECON 899 (Topics in dynamic economic theory and policy) in fall 2009 and fall 2010 gave me solid training on various fields in applied mathematics and economics, such as control theories, dynamical systems, macroeconomics, model uncertainty, new economic geography, social interactions, networks, and econometrics. His courses are featured by his thought-provoking ideas and insights. Attending his courses was an important training process in my life. His teaching on random graphs and suggested readings are very helpful for the project of random graphs and asymptotic theory (chapter 2). I thank him for writing detailed advices to me through e-mails and notes on several topics, for example,

control theories, policy lag problems, Ramsey policies, social interactions, and urban economics. In particular, the notes on policy lags and Ramsey policies are very helpful for the project of robust Ramsey taxation (chapter 3).

Things that I learnt from Professor Durlauf and Professor Brock make me a scholar.

I would like to thank Professor Xiaoxia Shi for her support and help on the job market; she is always very friendly. Professor Chao Fu and Professor Michael Gofman are very helpful and supportive. I thank for their encouragement.

I would like to take this opportunity to thank some people in the University of Hong Kong. First, I thank Professor Chung Yi Tse, Professor Paul S. H. Lau, and Professor James P. Vere for helping me to pursue my doctoral study in the United States. Special thanks to Professor Tse for his concern, patience, and continuous support. He introduced urban economics to me. He is not only the advisor of my master thesis but also a good friend of mine. I thank Professor Lau for his concern and support. Second, I thank the mathematicians in the University of Hong Kong who taught me mathematics in their courses when I was an undergraduate and master student. I deeply appreciate their enthusiasm in teaching; I have never found a course from the Department of Mathematics that the lecturer is not serious in teaching. Their courses gave me a solid foundation of mathematics so that I am able to work on theoretical economics and econometrics. I started appreciating mathematics as an art of proofs when I was taking Professor Man Keung Siu's YSCN 0002 in spring 2002 (Mathematics: a cultural heritage) and Professor Nam Kiu Tsing's MATH 1101 in fall 2002 (Linear algebra I).

Lastly, I have to thank my parents for giving me a warm and supportive growing environment, and for giving me enough rationality to do mathematics and enough sensibility to understand life and humans. My mother, Sau Ying Wong, is a fortitudinous woman, who contributes her life to the family; all my good personalities are from her (the bad personalities are developed by myself); I have been proud to hear that I am similar to my mother in personalities. My enthusiasm in knowledge, especially mathematical sciences, should be inherited from my father, Sit Kwan Kwok, who was fascinated with mathematics and science when he was young. I will be a son that they can be proud of.

# Abstract

## **Chapter 1: Identification Problems of Linear Social Interaction Models: A General Analysis Based on Matrix Spectral Decompositions**

This chapter develops new social interactions identification methods and develops a framework that includes some important existing results as special cases, such as the identification results in Bramoulle, Djebbari, and Fortin (2009) (except their proposition 3), the section 4.ii of Blume, Brock, Durlauf, and Ioannides (2011) (except their theorems 3 and 5), and Graham (2008). This chapter discovers that diameter is a key network property closely related to identification. The proposed methods are based on the matrix spectral decompositions; they address three canonical identification problems. First, this chapter offers a method of disentangling endogenous and exogenous interactions by the matrix spectral decompositions. Second, this chapter offers a detailed analysis of differencing methods, which solve the endogeneity problem arising from the presence of unobservable group-level heterogeneity (or fixed effects), and provides a method of minimizing the information loss from differencing. Third, this chapter develops an identification method based on the spectral decompositions of covariance matrices for the problem arising from the absence of observable individual-level heterogeneity; Graham's (2008) variance contrast method is a special case of this method.

## **Chapter 2: Estimating Covariance Matrices with Correlated Disturbances**

This chapter considers linear regression models where individuals interact in a social network so that the disturbances are correlated. A sufficient condition under which the covariance matrices can be consistently estimated is derived. The theory is based on the ideas in White (1984, 2001).

## **Chapter 3: Robust Ramsey Taxation under Implementation Lag Uncertainty**

This chapter develops a robust control theory under implementation lag uncertainty. The theory is applied to a Ramsey taxation problem. Implementation lags refer to the lag polynomials of the control variables. In the Ramsey taxation model, the control variable is the tax rate. The Ramsey planner has an

estimate of the implementation lag polynomial. He does not totally trust the estimate and thinks that any lag polynomial close to the estimate could be the true implementation lag polynomial. The closeness is measured by the  $l^2$ -norm. The Ramsey planner applies the robust control concept of Hansen and Sargent (2008). Robust control refers to maximin decision making. The Ramsey planner assumes that there is a hypothetical adversarial agent minimizes the welfare by choosing an implementation lag polynomial that is close to the estimate in the  $l^2$  sense, for a given policy choice. The inter-temporal correlations between the control variable and the exogenous state variables are the sources for the adversarial agent to minimize welfare.

## Chapter 1

# Identification Problems of Linear Social Interaction Models: A General Analysis Based on Matrix Spectral Decompositions

**JEL classification:** C31

**Keywords:** change of basis, diagonalization, diameter, eigenvalue, identification, social interaction, network, spectral decomposition

### 1.1 Introduction

Social interactions abound in theories in social sciences. But the development of the econometrics of social interactions is still in an initial stage.<sup>1</sup> Externalities, neighborhood effects, peer effects, and spillovers are examples of social interactions. Urban economic models such as Ogawa and Fujita (1980), Fujita and Ogawa (1982), Lucas (2001), and Lucas and Rossi-Hansberg (2002) rely on positive externalities to derive clusterings of human activities.

This chapter develops new social interactions identification methods and develops a framework which unifies some important existing results. For example, all theorems except the proposition 3 in Bramoulle, Djebbari, and Fortin (2009), all theorems except the “only if” parts of the theorem 3 and the theorem 5 in the section 4.ii of Blume, Brock, Durlauf, and Ioannides (2011), and Graham’s

---

<sup>1</sup>See Brock and Durlauf (2001), Blume, Brock, Durlauf, and Ioannides (2011), Blume, Brock, Durlauf, and Jayaraman (2011) for surveys.

(2008) variance contrast method are special cases in the proposed framework. Also, counterexamples to the statement ii of the theorem 3 of Blume, Brock, Durlauf, and Ioannides (2011, p.892) are discovered. Precisely, the proposed framework is based on the matrix spectral decompositions, which rely on the diagonalizability of the matrices that represent social interactions or the network structure. If we assume this diagonalizability, the above claims hold.<sup>2</sup>

General forms of linear social interaction models with indirect endogenous and exogenous interactions are first considered by this chapter. For example, the models capture the friends' friends' effects on academic achievements and the influences of indirectly connected households' political opinions on the voting behaviors. In this introduction, simple examples without these indirect interactions are used to illustrate the ideas.

There are two types of social interactions: endogenous interaction, which is the effect of group members' endogenous behaviors, and exogenous (contextual) interaction, which is the effect of group members' exogenous characteristics. Correlations between behaviors may be due to these two types of social interactions or due to correlations between unobservable characteristics.<sup>3</sup> Endogenous interactions are typically difficult to be disentangled from exogenous interactions, the effects of correlated unobservable characteristics, or both.

The methods address three canonical identification problems. Each problem is related to a form of linear social interaction model. The first identification problem is the difficulties in disentangling endogenous and exogenous interactions.<sup>4</sup> This chapter develops an effective method of decomposing endogenous and exogenous interactions by the spectral decompositions of social interaction matrices.<sup>5</sup> Suppose there is a classroom of  $N$  students. These students' academic results are determined by the social interaction model

$$\omega = k1_{N \times 1} + JA\omega + c_0x + c_1Ax + \epsilon. \quad (1.1)$$

$\omega$  is an  $N \times 1$  column of endogenous outcomes, academic results.  $x$  is an  $N \times 1$  column of observable exogenous characteristics, such as intelligence quotients.  $k$  is a constant and  $1_{N \times 1}$  is an  $N \times 1$  column of 1's.  $\epsilon$  is an  $N \times 1$  column of

<sup>2</sup>The diagonalizability assumption is not as restrictive as it may first seem. All undirected networks satisfy this condition (the formal definitions of directed and undirected networks are in section 1.2.1). The counterexamples to the statement ii of the theorem 3 of Blume, Brock, Durlauf, and Ioannides (2011, p.892) do not impose the diagonalizability; the matrices are diagonalizable per se. For Graham's (2008) variance contrast method, covariance matrices are diagonalizable because they are symmetric.

<sup>3</sup>For example, in a classroom, there are endogenous interactions if a student's academic achievement depends on the academic achievements of other students. There are exogenous interactions if a student's academic achievement depends on other students' characteristics, such as intelligence quotients and family backgrounds. Also, the correlation between the students' academic achievements may be due to the correlations between students' unobservable characteristics, such as unobservable cognitive abilities, or unobservable class characteristics, such as teacher quality.

<sup>4</sup>This is Manski's (1993) reflection problem.

<sup>5</sup>This chapter considers linear social interaction models, which are systems of linear equations; social interactions are represented by matrices. Terms such as "social interaction matrix" and "endogenous interaction matrix" will be used to address these matrices.

unobservable exogenous characteristics, such as personality traits.  $A$  is an  $N \times N$  matrix representing how the group members are connected or interacting. The  $(i, j)$ -entry of  $A$  is non-zero if the  $i^{\text{th}}$  student is connected to the  $j^{\text{th}}$  student.  $JA$  and  $c_1A$  are respectively the endogenous and exogenous interaction matrices. The magnitude of these interactions are represented by the scalars  $J$  and  $c_1$ . A reduced form of the model (1.1) is

$$\omega = (I_N - JA)^{-1}k1_{N \times 1} + (I_N - JA)^{-1}(c_0I_N + c_1A)x + (I_N - JA)^{-1}\epsilon, \quad (1.2)$$

where  $I_N$  is the  $N \times N$  identity matrix. The endogenous and exogenous interactions are mixed together in the “total social interaction matrix”  $(I - JA)^{-1}(c_0I_N + c_1A)$ . Separating these interactions or identifying  $J$ ,  $c_0$ , and  $c_1$  is typically not straightforward. Applying matrix spectral decomposition to the total social interaction matrix can decouple the system of  $N$  linear equations (1.2) into  $N$  independent linear equations. The social interactions are nicely summarized in the eigenvalues of the total social interaction matrix. The number of distinct eigenvalues determines how many equations are available for identification. The parameters  $J$ ,  $c_0$ , and  $c_1$  are identified if the number of distinct eigenvalues is larger than or equal to the number of parameters, which is three in this example. The number of distinct eigenvalues is determined by the network structure or the structure of the matrix  $A$ .<sup>6</sup>

The matrix  $A$  having sufficient number of distinct eigenvalues is the key identification condition. The number of distinct eigenvalues of  $A$  is closely related to the network diameter, which is the longest distance between individuals in the network. This observation offers a new insight into the identification problem. Precisely, the non-identification of parameters impose an necessary condition, an upper bound, on the network diameter. Some important network properties that are related to identification or non-identification in the literature, such as the properties of transitivity and interacting in completely connected components in the section 4.ii of Blume, Brock, Durlauf, and Ioannides (2011), can be derived from the upper bound on the diameter. Thus, network diameter is of more primary importance to identification than other network properties. This new insight into the relation between identification and diameter help us to find counterexamples to the statement ii of the theorem 3 of Blume, Brock, Durlauf, and Ioannides (2011, p.892).

The second identification problem arises from the presence of unobservable group-level heterogeneity (or fixed effects). The presence of group-level unobservable variables causes an endogeneity problem because the observable exogenous characteristics and the group-level unobservable characteristics are

---

<sup>6</sup>The case where all entries of  $A$  in the model (1.1) are  $1/N$  is a version of linear-in-means model. Parameters in Datcher (1982) and Manski’s (1993) linear-in-means models are not identified (if there is no valid instrument) because the social interaction matrices do not have enough distinct eigenvalues. Bramouille, Djebbari, and Fortin (2009) considered models with richer structure of social interactions in social networks, or with more general  $A$  than Datcher (1982) and Manski’s (1993). They derived the necessary and sufficient conditions on the social interaction matrices for identification. Their conditions can be equivalently stated in terms of the number of distinct eigenvalues of  $A$  if  $A$  is diagonalizable.

potentially correlated. This endogeneity problem is a major impediment to identifying social interactions because we can easily envision the existence of group-level unobservable variables in various contexts: teacher quality in the context of classroom peer effects (Cooley 2007 and Graham 2008), legal and political factors in the context of Appalachian poverty (Durlauf 2010), and geographical characteristics in spatial econometric models. The problem can be illustrated by adding an unobservable group-level variable  $\alpha$ , teacher quality, to the model (1.1):

$$\omega = \alpha 1_{N \times 1} + JA\omega + c_0x + c_1Ax + \epsilon. \quad (1.3)$$

$\alpha$  is correlated with  $x$  because it is reasonable to assume that teaching performance is affected by student characteristics.

Bramouille, Djebbari, and Fortin (2009) and the section 4.iii of Blume, Brock, Durlauf, and Ioannides (2011) proposed various differencing methods to solve the endogeneity problem.<sup>7</sup> Differencing refers to subtracting a subgroup or subclass average of the dependent variables from each dependent variable to cancel out the group-level unobservable variables. For example,  $\alpha 1_{N \times 1}$  can be cancelled away by subtracting the classroom average academic result from each student's academic result:

$$(I_N - 1_{N \times N})\omega = (I_N - 1_{N \times N})(JA\omega + c_0x + c_1Ax + \epsilon), \quad (1.4)$$

where  $1_{N \times N}$  is an  $N \times N$  matrix of 1's. Note that any subclass average works also.

But differencing causes information loss. A detailed analysis of various differencing methods is absent in the literature. This chapter provides a detailed analysis of differencing methods and a method to minimize information loss. Precisely, by applying spectral decomposition to the differencing matrix, which is the matrix taking total average  $1_{N \times N}$  in the example, we are able to identify which vector subspace's information is lost from differencing. Hence we are able to choose a differencing method to minimize information loss. This chapter offers a sufficient condition under which the lost information from differencing is unnecessary for identification, a sufficient condition under which the presence of group-level unobservable variables does not require more restrictive identification conditions.

The third identification problem arises from the absence of observable heterogeneity so that there is no instrument for the endogenous interaction terms in social interaction models. The absence of instrument forces us to rely on the information from higher-order moments to identify parameters. For example, this problem arises if there is no  $x$  in the model (1.3):

$$\omega = \alpha 1_{N \times 1} + JA\omega + \epsilon. \quad (1.5)$$

There are correlations between unobservable characteristics because the students' performances are correlated with each other and are correlated with

---

<sup>7</sup>The instrumental variable method is another way to address the endogeneity problem; see Cooley (2007) for example.

the teacher's performance; the off-diagonal entries of the covariance matrix of  $\alpha\mathbf{1}_{N \times 1} + \epsilon$  are non-zero. Endogenous interactions and the effects of correlated unobservables are mixed together in a complicated matrix multiplication:

$$\text{Var}(\omega) = (I_N - JA)^{-1} \text{Var}(\alpha\mathbf{1}_{N \times 1} + \epsilon) ((I_N - JA)^{-1})'. \quad (1.6)$$

This chapter develops a systematic method of identifying parameters by covariance matrix decompositions. A spectral decomposition of a covariance matrix decomposes the covariance matrix into a linear combination of a certain number of components. The information of social interactions is summarized in the coefficients of these components, which are eigenvalues of the covariance matrix. The parameters are identified if the number of distinct eigenvalues is larger than or equal to the number of parameters.

The variance contrast method introduced by Graham (2008) is a special case of the proposed covariance matrix decomposition method because the contrasted variances are scaled eigenvalues of  $\text{Var}(\omega)$ . While the variance contrast method identifies only the endogenous interaction parameter  $J$ , the covariance matrix decomposition method identifies the endogenous interaction parameter as well as all second moments of the group-level and individual-level unobservable variables  $\text{Var}(\alpha)$ ,  $\text{Var}(\epsilon_i)$ , and  $\text{Cov}(\epsilon_i, \epsilon_j)$ , which are of economic interest.

The covariance matrix decomposition method requires less restrictive assumptions than that in Graham (2008). Graham assumed no within-class correlations (or no within-group correlations) between unobservable variables and homoskedasticity across class types (or group types). No within-class correlations means that the individual-level unobservables (student productivities) are uncorrelated with each other, and are uncorrelated with the group-level unobservable (teacher effectiveness). Homoskedasticity across class types means that the variances and covariances of the unobservable variables are invariant in class type. The covariance matrix decomposition method relaxes both assumptions: it allows within-group correlations between unobservable variables and heteroskedasticity across class types. Furthermore, the covariance matrix decomposition method is applicable to more general network structure than the linear-in-means model in Graham (2008). Allowing within-group correlations is desirable because it is reasonable to expect that the students' productivities are correlated with each other and correlated with the teacher's effectiveness (even if students and teachers are randomly assigned to classes). Homoskedasticity across class types limits the applicability of the variance contrast method; one of the purposes of the Project STAR, the data in Graham (2008), is to investigate the effects of class types on academic achievements.

The chapter is organized as follows. Sections 1.2, 1.3, and 1.4 consider the three identification problems one by one. Section 1.2 develops the central ideas of identifying social interactions by the spectral decompositions of matrices. Section 1.2.5 is important section on the relation between non-identification and diameter. Sections 1.3 and 1.4 can be read in any order. The last subsections of sections 1.3 and 1.4 may be skipped without affecting the understanding of other sections.

## 1.2 Disentangling Exogenous and Endogenous Interactions

### 1.2.1 Network Terminology

This chapter considers social interactions in social networks. An individual has social influences on those who have a link from him. Consider a network of  $N$  individuals. The network structure is described by an  $N \times N$  matrix  $A$ .  $A$  is called the adjacency matrix of the network. The  $(i, j)$ -entry of  $A$  is non-zero,  $A_{ij} \neq 0$ , if there is a link from individual  $i$  to individual  $j$ , and  $A_{ij} = 0$  otherwise. Note that the direction of links matters; the network is directed. For undirected networks,  $A_{ij} = A_{ji}$ . Undirected networks' adjacency matrices are symmetric. Some authors normalize the row sums of adjacency matrices to be 1. When a theorem applies to normalized adjacency matrices only, this chapter will state so clearly.

If  $A_{ij} \in \{0, 1\}$ , then simple algebra shows that the  $(i, j)$ -entry of  $A^2$  is the number of walks from  $i$  to  $j$  of length 2.<sup>8</sup> <sup>9</sup> Similarly, the  $(i, j)$ -entry of  $A^d$  is the number of walks from  $i$  to  $j$  of length  $d$ . If  $A_{ij} \in \mathbf{R}$ , then the  $(i, j)$ -entry of  $A^d$  is the number of walks from  $i$  to  $j$  of length  $d$  weighted by the strengths of connections. This property allows us to express social interactions of indirectly linked individual easily. For example, in the context of classroom peer effects, we may use  $J_2 A^2$  to represent the peer effects of friends' friends on academic achievement, where  $J_2$  is a scalar to represent the magnitude of the indirect interactions.

### 1.2.2 The Problem of Disentangling Social Interactions

The model in this section has endogenous interactions between individuals that are indirectly connected up to distance  $D$  and exogenous interactions up to distance  $L$ . For example, in the context of voting behavior, the model captures the effects of indirectly connected households' political opinions. This type of

---

<sup>8</sup>A (directed) walk from individual  $i$  to individual  $j$  is a sequence of links from  $i$  to  $j$ . A (directed) path from individual  $i$  to individual  $j$  is a sequence of links from  $i$  to  $j$  where all individuals in the sequence of links from  $i$  to  $j$  are distinct.

<sup>9</sup>For example, consider a network consisting of three individuals: 1, 2 and 3. These three persons are located on a circle and connected clockwise. There are three links: from 1 to 2, 2 to 3, and 3 to 1. The adjacency matrix for this network is

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \quad (1.7)$$

Obviously, there are three walks of length 2: from 1 to 3, 2 to 1, and 3 to 2. The matrix  $A^2$  represents this "distance-2" structure:

$$A^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \quad (1.8)$$

more general social interaction models are first considered by this chapter. The model is

$$\omega = k1_{N \times 1} + \sum_{d=1}^D J_d A^d \omega + c_0 x + \sum_{l=1}^L c_l A^l x + \epsilon, \quad (1.9)$$

$$E(\epsilon|x) = 0. \quad (1.10)$$

$\omega$  is a  $N \times 1$  column of individual behaviors or outcomes, such as academic achievements.  $x$  and  $\epsilon$  are respectively  $N \times 1$  columns of observable individual characteristics, such as students' demographic characteristics, and unobservable individual characteristics, such as students' unobservable cognitive skills.  $k$  is a constant.  $1_{N \times 1}$  is a  $N \times 1$  column of 1's. The matrix polynomial  $\sum_{d=1}^D J_d A^d$  is the endogenous interaction matrix, where  $J_d$  is the endogenous interaction of individuals at distance  $d$ . The matrix polynomial  $\sum_{l=1}^L c_l A^l$  is the exogenous interaction matrix, where  $c_l$  is the exogenous (contextual) effects of individuals at distance  $l$ . The network structure, or the adjacency matrix  $A$ , is observable to the econometrician. The set of parameters to be identified is  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$ . The analysis in this section can be easily extended to models with more than one column of individual characteristics.

Endogenous interactions are mixed with exogenous interactions. This is the well-known reflection problem in Manski (1993). Rearranging terms in the model (1.9) and using the fact that  $(I_N - \sum_{d=1}^D J_d A^d)^{-1} = \sum_{t=0}^{\infty} (\sum_{d=1}^D J_d A^d)^t$  give a reduced form

$$\begin{aligned} \omega = k \sum_{t=0}^{\infty} \left( \sum_{d=1}^D J_d A^d \right)^t 1_{N \times 1} + \sum_{t=0}^{\infty} \left( \sum_{d=1}^D J_d A^d \right)^t \left( c_0 I_N + \sum_{l=1}^L c_l A^l \right) x \\ + \sum_{t=0}^{\infty} \left( \sum_{d=1}^D J_d A^d \right)^t \epsilon. \end{aligned} \quad (1.11)$$

The endogenous and exogenous interactions are mixed together in the infinite degree matrix polynomial in second term. Note that  $(I_N - \sum_{d=1}^D J_d A^d)^{-1}$  is assumed to be invertible. This is a self-consistency assumption in the sense that the choices of individuals,  $\omega$  in the left hand side of (1.9), are consistent with the perceived behaviors of individuals,  $\omega$  in the right hand side of (1.9).

A projection of  $\omega$  onto the set of (or a subset of) regressors  $\{A^t 1_{N \times 1}, A^t x | t = 0, 1, \dots\}$  does not necessarily identify the set of parameters. Such a projection identifies only a certain number of composite coefficients which are functions of  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$ . Since  $\{1_{N \times 1}, x, Ax, \dots, A^L x\}$  is already in the model (1.9), identification requires enough regressors in  $\{A 1_{N \times 1}, A^2 1_{N \times 1}, \dots\}$  and  $\{A^{L+1} x, A^{L+2} x, \dots\}$  that are linearly independent of  $\{1_{N \times 1}, x, Ax, \dots, A^L x\}$  so that the endogenous terms,  $\{A\omega, \dots, A^D \omega\}$ , are linearly independent of the exogenous terms,  $\{1_{N \times 1}, x, Ax, \dots, A^L x\}$ .

In other words, identification requires all (endogenous and exogenous) regressors in the model (1.9),  $\{1_{N \times 1}, A\omega, \dots, A^D \omega, Ax, \dots, A^L x\}$ , to be linearly independent. Among this regressors, (1.11) tells us that the endogenous terms

$\{A\omega, \dots, A^D\omega\}$  are infinite sums of  $A^t \mathbf{1}_{N \times 1}$ 's and  $A^t x$ 's. Since the lower degree terms,  $\{\mathbf{1}_{N \times 1}, Ax, \dots, A^L x\}$ , are in the model (1.9), identification requires the higher degree terms to be linearly independent of the lower degree terms so that the endogenous terms are linearly independent of the exogenous terms. We will see that the linear independence of the lower and higher degree terms can be stated in terms of the number of distinct eigenvalues of  $A$ .

### 1.2.3 Spectral Decompositions

This subsection introduces a version of spectral decomposition that is used in this chapter. A  $N \times N$  matrix  $A$  is said to be diagonalizable if there exist an invertible matrix  $P$  so that

$$A = P \text{diag}(\xi_1, \dots, \xi_N) P^{-1}, \quad (1.12)$$

where  $\xi_i$ 's are eigenvalues of  $A$ , and  $\text{diag}(\xi_1, \dots, \xi_N)$  is a diagonal matrix with  $\xi_1, \xi_2, \dots$ , and  $\xi_N$  as the diagonal entries. A spectral decomposition of  $A$  can be constructed as follows. Denote the  $i^{\text{th}}$  column of  $P$  which is an eigenvector of  $A$  as  $P(i)$  and the  $j^{\text{th}}$  row of  $P^{-1}$  as  $P^{-1}(j)$ . Then (1.12) can be written as

$$A = \sum_{i=1}^N \xi_i P(i) P^{-1}(i). \quad (1.13)$$

The right hand side of (1.13) is a spectral decomposition of  $A$ .

The  $N \times N$  matrices  $P(i)P^{-1}(i)$ 's are projection matrices.  $P(i)P^{-1}(i)$  projects vectors onto a vector subspace  $V_i$  which is spanned by  $\{P(i)\}$ .<sup>10</sup> These projection matrices are idempotent:  $P(i)P^{-1}(i)P(i)P^{-1}(i) = P(i)P^{-1}(i)$  for all  $i$ ,<sup>11</sup> pairwise orthogonal:  $P(i)P^{-1}(i)P(j)P^{-1}(j) = 0$  for all  $i \neq j$ , and summing to the identity matrix:  $\sum_{i=1}^N P(i)P^{-1}(i) = I_N$ .<sup>12</sup>

We will see that these three nice properties help us to reduce the infinite degree matrix polynomial in the reduced form (1.11) into a finite degree matrix polynomial. The spectral decompositions of social interaction matrices separate the social interactions into  $N$  pieces. The actions of the social interaction matrices on a vector subspace are nicely summarized in the eigenvalue corresponding to that vector subspace.

<sup>10</sup>Although all social interaction matrices in this chapter are real matrices, their eigenvalues may be complex numbers and the projections  $P(i)P^{-1}(i)$ 's may be complex matrices. We need to consider the vector space of  $N \times 1$  complex vectors  $\mathbf{C}^N$ . Note that, first,  $V_i$ 's are disjoint since  $\{P(1), \dots, P(N)\}$  is a basis of  $\mathbf{C}^N$ . Second, for all  $v \in V_i$ ,  $Av = \xi_i v$ , so the vector  $Av$  is in  $V_i$ .  $V_i$ 's are  $A$ -invariant subspaces.

<sup>11</sup>A matrix is a projection if and only if it is idempotent. This is the theorem 1 of Halmos (1958, p.73).

<sup>12</sup> $P^{-1}P = I_N$  implies that  $P^{-1}(i)P(i) = 1$  for all  $i$  and that  $P^{-1}(i)P(j) = 0$  for all  $i \neq j$ , so the matrices  $P(i)P^{-1}(i)$ 's are idempotent and pairwise orthogonal. Since  $PP^{-1} = \sum_{i=1}^N P(i)P^{-1}(i) = I_N$ , the matrices  $P(i)P^{-1}(i)$ 's sum to the identity matrix.

The above construction of a spectral decomposition is summarized in the following theorem.<sup>13</sup>

**Theorem 1 (Spectral Decomposition Theorem)**<sup>14</sup> *A square matrix  $A$  is diagonalizable if and only if  $A$  can be expressed as a linear combination of projection matrices as in (1.13) where the projection matrices are idempotent, pairwise orthogonal and summing to the identity matrix.*

All adjacency matrices in this chapter are assumed to be diagonalizable so that we can apply the spectral decomposition theorem. For undirected networks,  $A$  is symmetric, so  $A$  is orthogonally diagonalizable.<sup>15</sup> For directed networks,  $A$  may not be diagonalizable. For example, the adjacency matrices of directed networks without cycles are not diagonalizable, where cycles are closed loops of links.<sup>16</sup> To see this, recall that the  $(i, j)$ -entry of  $A^d$  is the number of walks from  $i$  to  $j$  of length  $d$ . If a network has no cycles, then  $A^d = 0$  for some integer  $d$ . That is,  $A$  is nilpotent. All non-zero nilpotent matrices are not diagonalizable.

## 1.2.4 Decompositions of Social Interactions

Before introducing the identification methods, we need a definition of identification. Suppose  $\theta$  is a set of parameters to be identified. For model (1.9),  $\theta = \{k, J_1, \dots, J_D, c_0, \dots, c_L\}$ . Denote the probability distribution of  $\omega$  given  $x$  and  $\theta$  as  $P(\omega|x, \theta)$ .

**Definition 1** *Suppose  $\Theta$  is the set of  $\theta$  satisfying the self-consistency assumption. The set of parameters  $\theta$  is identified if  $P(\omega|x, \theta') = P(\omega|x, \theta'')$  implies  $\theta' = \theta''$  for all  $\theta', \theta'' \in \Theta$ .*

<sup>13</sup>We may understand spectral decomposition as follows. Interpret  $A$  as a diagonalizable linear transformation  $A : V \rightarrow V$  where  $V$  is a vector space, such as  $\mathbf{C}^N$ . First, a diagonalizable linear transformation can be represented by a diagonal matrix  $P^{-1}AP$  after changing the basis of  $V$  from the standard basis to the basis consisting of the eigenvectors of  $A$ . Second, after changing the basis, the coordinate of a vector  $v$  with respect to the basis vector  $P(i)$  is  $P^{-1}(i)v$ . Third, a diagonalizable linear transformation  $A$  can be written as a linear combination of projections  $P(i)P^{-1}(i)$ 's as in (1.13). Fourth, according to the spectral decomposition of  $A$ , the vector space  $V$  can be written as a direct sum:

$$V = V_1 \oplus V_2 \oplus \dots \oplus V_N. \quad (1.14)$$

The action of  $A$  on  $V_i$  is represented by the  $i^{\text{th}}$  eigenvalue  $\xi_i$ .

<sup>14</sup>The spectral decomposition (1.13) is not a standard version of the spectral decomposition in linear algebra textbooks. In a standard linear algebra textbook, the projection matrices corresponding to the same eigenvalues are grouped into one term. Suppose  $A$  has  $R+1$  distinct eigenvalues, and the  $r^{\text{th}}$  distinct eigenvalue is denoted as  $\xi_r$  where  $r = 0, \dots, R$ . Then we may write  $A = \sum_{r=0}^R \xi_r S_r$ , where  $S_r$ 's are projection matrices that are idempotent, pairwise orthogonal, and summing to the identity matrix. See the theorems D1 and D2 of Shaw (1982, p.80-81).

<sup>15</sup>A square matrix  $P$  is orthogonal if the inverse of  $P$  is equal to the transpose of  $P$ :  $P^{-1} = P'$ . A square matrix  $A$  is orthogonally diagonalizable if there exists an orthogonal matrix  $P$  so that  $P'AP$  is diagonal. The theorem 2 of Nicholson (2002, p.280) states that a matrix  $A$  is orthogonally diagonalizable if and only if  $A$  is symmetric.

<sup>16</sup>A (directed) cycle is a sequence of links where the beginning node (agent) and the ending node (agent) are the same.

By an argument similar to that in the section 3.2 of Hsiao (1983), two sets of parameters implying the same reduced form are observationally equivalent.  $\theta$  is identified if this condition holds: for all  $\theta', \theta'' \in \Theta$ , if  $\theta'$  and  $\theta''$  implies the same reduced form, then  $\theta' = \theta''$ . In the subsequent discussion, we will see that the reduced form of the social interaction model is identified. The parameters are identified whenever the identified coefficients or matrix entries of the reduced form uniquely determine the parameters of interest.<sup>17</sup>

Assume the following two conditions.

**Assumption 1 (Diagonalizability)** *The adjacency matrix  $A$  is diagonalizable.*

**Assumption 2 (Self-consistency)** *The adjacency matrix's eigenvalues,  $\xi_1, \xi_2, \dots$ , and  $\xi_N$ , and the endogenous interaction parameters,  $J_1, J_2, \dots$ , and  $J_D$ , are of values so that  $\max_i |\sum_{d=1}^D J_d \xi_i^d| < 1$ .*

Diagonalizability allows us to use spectral decomposition. Self-consistency condition guarantees the invertibility of  $(I_N - \sum_{d=1}^D J_d A^d)$ .

After substituting the spectral decomposition (1.13) into the reduced form (1.11) and using the idempotence and pairwise orthogonality of  $P(i)P^{-1}(i)$ 's,  $\omega$  is a linear combination of column vectors  $P(i)$ 's, the eigenvectors of  $A$  (recall that  $P^{-1}(i)$ 's are row vectors):

$$\omega = \sum_{i=1}^N \Lambda_i P(i) P^{-1}(i) 1_{N \times 1} + \sum_{i=1}^N \Delta_i P(i) P^{-1}(i) x + \hat{\epsilon}, \quad (1.15)$$

where

$$\Lambda_i \equiv \frac{k}{1 - \sum_{d=1}^D J_d \xi_i^d}, \quad (1.16)$$

$$\Delta_i \equiv \frac{c_0 + \sum_{l=1}^L c_l \xi_i^l}{1 - \sum_{d=1}^D J_d \xi_i^d}, \quad (1.17)$$

and

$$\hat{\epsilon} \equiv \sum_{i=1}^N \frac{1}{1 - \sum_{d=1}^D J_d \xi_i^d} P(i) P^{-1}(i) \epsilon. \quad (1.18)$$

The scalars  $\Lambda_i$ 's are eigenvalues of  $k(I_N - \sum_{d=1}^D J_d A^d)^{-1}$ . The scalars  $\Delta_i$ 's are eigenvalues of the "total social interaction matrix"  $(I_N - \sum_{d=1}^D J_d A^d)^{-1}(c_0 I_N + \sum_{l=1}^L c_l A^l)$ .

The manipulation from (1.11) to (1.15) is essentially a change of basis process from writing the vectors,  $\omega$ ,  $1_{N \times 1}$ ,  $x$ , and  $\epsilon$ , with respect to the standard basis

<sup>17</sup>See Hsiao (1983) and Blume, Brock, Durlauf, and Jayaraman (2011) for more discussions on the concepts of identification.

of  $\mathbf{C}^N$  to writing vectors with respect to the basis  $\{P(1), \dots, P(N)\}$ . The coordinate of a vector  $v$  with respect to the basis vector  $P(i)$  is the scalar  $P^{-1}(i)v$ . The coordinate of  $\omega$  with respect to the eigenvector  $P(i)$  is the scalar

$$P^{-1}(i)\omega = \Lambda_i P^{-1}(i)1_{N \times 1} + \Delta_i P^{-1}(i)x + P^{-1}(i)\hat{\epsilon}. \quad (1.19)$$

(1.19) can be obtained by multiplying (1.15) by the row vector  $P^{-1}(i)$  because  $P^{-1}(i)P(j) = 0$  for all  $i \neq j$ .

The endogenous and exogenous interactions are now neatly represented in a finite degree matrix polynomial in the reduced form (1.15), while they are mixed in an infinite degree matrix polynomial in the reduced form (1.11). In (1.19), instead of matrix multiplications, the social interactions are nicely represented by the scalar multiplications of the coordinates of  $1_{N \times 1}$  and  $x$  with respect to the basis vector  $P(i)$  (which are  $P^{-1}(i)1_{N \times 1}$  and  $P^{-1}(i)x$ ) by the eigenvalues of the social interaction matrices (which are  $\Lambda_i$  and  $\Delta_i$ ).

The system of  $N$  interdependent equations in the reduced form (1.11) is decoupled into  $N$  equations (1.19). These equations are independent of each other. We now have a system of  $N$  equations that are independent of each other.

Assume that the support of  $x$  is not in a proper subspace of  $\mathbf{R}^N$ . For each equation of the decoupled system, the coefficients  $\Lambda_i$  and  $\Delta_i$  are identified.<sup>18</sup> That is, the reduced form coefficients or the matrix entries are identified. They give a set of equations of the forms in (1.16) and (1.17). If the number of distinct eigenvalues,  $\Lambda_i$  and  $\Delta_i$ , are large enough so that the number of equations is large enough, then the set of parameters  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified. The number of distinct  $\Lambda_i$  and  $\Delta_i$  depends on the number of distinct eigenvalues of  $A$ ,  $\xi_i$ .

In the literature, identification conditions are typically stated in terms of the linear independence of a subset of  $\{I_N, A, A^2, \dots\}$ . The following theorem connects the number of distinct eigenvalues to the linear independence of a subset of  $\{I_N, A, A^2, \dots\}$ .

**Theorem 2**<sup>19 20</sup> *Suppose  $A$  is a diagonalizable matrix.  $A$  has  $R + 1$  distinct eigenvalues if and only if  $R$  is the minimal degree so that  $\{I_N, A, \dots, A^R\}$  is linearly independent.*

There are two cases:  $A$  is normalized so that the row sums of the non-zero rows are 1 and  $A$  is not normalized. Normalization of row sums imposes an assumption that each individual is getting the same amount of social effects.

<sup>18</sup>Indeed, for each  $i$ ,  $\Lambda_i$  and  $\Delta_i$  are identified if  $P^{-1}(i)x$  can take at least two different values.

<sup>19</sup>This theorem is implied by the theorems 23.11 and 23.12 of Curtis (1984, p.199-200). See the theorems D1 and D2 of Shaw (1982, p.80-81) for reference also. The theorem 22.3 of Curtis (1984, p.185) states that, for any  $N \times N$  matrix  $M$ , there exists a unique integer  $r$  so that  $\{I_N, M, \dots, M^r\}$  is linearly independent and  $\{I_N, M, \dots, M^{r+1}\}$  is linearly dependent.

<sup>20</sup>The number of distinct eigenvalues of a matrix is closely related to that the minimal polynomial of that matrix; see the theorem 24.1 of Curtis (1984, p.202) and the theorem J3 of Shaw (1982, p.89).

We consider the case of normalized  $A$  first. There are two cases also:  $A$  does not have zero rows and  $A$  has zero rows. Having a zero row means that there is an individual who does not link to others; that is, this individual is not affected by anyone in the network (but he may affect others if there are links from others to him).

If  $A$  does not have zero row and the row sums are 1, then  $1_{N \times 1}$  is an eigenvector of  $A$  (the corresponding eigenvalue is 1). Assume without loss of generality that  $P(1) = 1_{N \times 1}$ . Then  $P^{-1}(i)1_{N \times 1} = 0$  for  $i = 2, \dots, N$  because  $P^{-1}P = I_N$ . The first term in (1.19),  $\Lambda_i P^{-1}(i)1_{N \times 1}$ , is zero for  $i = 2, 3, \dots, N$ .<sup>21</sup> Only  $\Lambda_1, \Delta_1, \dots$ , and  $\Delta_N$  are identified. If  $A$  has  $D + L + 1$  distinct eigenvalues so that there are  $D + L + 1$  distinct  $\Delta_i$ 's, then  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified by these  $D + L + 1$  equations from  $\Delta_i$ 's. Hence,  $k$  is identified by the equation from  $\Lambda_1$ . If  $A$  has less than  $D + L + 1$  distinct eigenvalues, the parameters are not identified. Note that the arguments are valid even if  $k = 0$  because the equation from  $\Lambda_1$  is for identifying  $k$ .

**Theorem 3** *Consider the model defined by (1.9) and (1.10). Suppose  $A$  is diagonalizable,  $A$  has no zero rows,  $A$ 's row sums are 1, and the self-consistency assumption holds.  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has at least  $D + L + 1$  distinct eigenvalues, or equivalently,  $\{I_N, A, \dots, A^{D+L}\}$  is linearly independent.*

A formal proof of theorem 3 is in appendix 1.6. The reduced form (1.11) tells us that the endogenous terms in  $\{A\omega, \dots, A^D\omega\}$  are linear combinations of the regressors in  $\{A^t 1_{N \times 1}, A^t x | t = 0, 1, \dots\}$ . Identification requires enough regressors in  $\{A 1_{N \times 1}, A^2 1_{N \times 1}, \dots\}$  and  $\{A^{L+1}x, A^{L+2}x, \dots\}$  that are linearly independent of the exogenous terms  $\{1_{N \times 1}, x, Ax, \dots, A^L x\}$  so that the endogenous terms are linearly independent of the exogenous terms. But  $A^t 1_{N \times 1} = 1_{N \times 1}$  for all  $t$  when  $A$  does not have zero rows and the row sums are 1. The regressors in  $\{A 1_{N \times 1}, A^2 1_{N \times 1}, \dots\}$  cannot help making the endogenous terms to be linearly independent of the exogenous terms. In theorem 3, the linear independence of  $\{I_N, A, \dots, A^L\}$  ensures that the exogenous terms are linearly independent so that we can identify  $\{c_0, \dots, c_L\}$ . The linear independence of  $\{I_N, A, \dots, A^L, A^{L+1}, \dots, A^{D+L}\}$  ensures that the endogenous terms are linearly independent of the exogenous terms so that we can identify  $\{J_1, \dots, J_D\}$ .

If  $A$  has zero rows and the row sums of the non-zero rows are 1,  $1_{N \times 1}$  cannot be an eigenvector corresponding to any eigenvalue of  $A$ .  $1_{N \times 1}$  must be in a space spanned by the eigenvectors corresponding to at least two distinct eigenvalues. We must have  $P^{-1}(s)1_{N \times 1} \neq 0$  and  $P^{-1}(t)1_{N \times 1} \neq 0$  for some  $s$  and  $t$  so that  $\xi_s \neq \xi_t$ ; for these  $s$  and  $t$ ,  $\Lambda_s$  and  $\Lambda_t$  are identified. We at least identify  $\Lambda_s, \Lambda_t, \Delta_1, \dots$ , and  $\Delta_N$ . If  $A$  has  $D + L$  distinct eigenvalues, then

<sup>21</sup>Here is a more precise argument. Suppose the multiplicity of 1 is  $m_A$  and the first  $m_A$  eigenvalues are 1. Since  $1_{N \times 1}$  is an eigenvector corresponding to the eigenvalue 1,  $1_{N \times 1}$  is a linear combination of  $P(1), P(2), \dots$ , and  $P(m_A)$ . There may be more than one  $i \in \{1, \dots, m_A\}$  so that  $P^{-1}(i)1_{N \times 1} \neq 0$ . But the conclusion will not change since  $\Lambda_1 = \dots = \Lambda_{m_A}$ . Changing the basis of the eigenspace will not change the final conclusion.

$\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified by the  $D + L + 2$  distinct equations from these  $\Lambda_s, \Lambda_t, \Delta_1, \dots,$  and  $\Delta_N$ .

The arguments in the previous paragraph require  $k \neq 0$ . If  $k = 0$ ,  $\Lambda_i \equiv 0$  for all  $i$ ; the equations from  $\Lambda_s$  and  $\Lambda_t$  are not available.  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has  $D + L + 1$  distinct eigenvalues, the same necessary and sufficient condition as in theorem 3.

**Theorem 4** *Consider the model defined by (1.9) and (1.10). Suppose  $A$  is diagonalizable,  $A$  has zero rows, the row sums of the non-zero rows are 1, and the self-consistency assumption holds.*

(i.) *Suppose  $k \neq 0$ ,  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if  $A$  has at least  $D + L$  distinct eigenvalues.*

(ii.) *Suppose  $k = 0$ ,  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has at least  $D + L + 1$  distinct eigenvalues.*

A formal proof of theorem 4 is in appendix 1.6. When  $A$  has zero rows,  $A1_{N \times 1} \neq 1_{N \times 1}$ .  $A1_{N \times 1}$  and  $1_{N \times 1}$  are linearly independent. If  $k \neq 0$ ,  $A1_{N \times 1}$  is always an available regressor to make the endogenous terms to be linearly independent of the exogenous terms. Thus, when  $k \neq 0$ , the linear independence of  $\{I_N, A, \dots, A^{D+L-1}\}$  is sufficient for identification.

If  $A$  has zero rows, 0 must be an eigenvalue. If  $A \neq 0$ ,  $A$  must have at least one non-zero eigenvalue.  $A$  has at least two distinct eigenvalues. The statement (i.) of theorem 4 implies that the parameters must be identified if  $D = L = 1$  and  $k \neq 0$ .

The models in the section 2 of Bramoulle, Djebbari and Fortin (2009) and the section 4.ii of Blume, Brock, Durlauf, and Ioannides (2011) are the cases where  $L = D = 1$  and  $A$  is a normalized adjacency matrix. If  $A$  is diagonalizable, theorems 3 and 4 cover the proposition 1 of Bramoulle, Djebbari and Fortin (2009) and the theorem 2 of Blume, Brock, Durlauf, and Ioannides (2011). Theorem 4 goes beyond their results by offering a weaker sufficient condition when  $A$  has zero rows.

When the row sums of  $A$  are not normalized, individuals are allowed to receive different amounts of social effects. For example, some students are easier to be influenced by others due to personality differences. If  $A$ 's row sums are not normalized, there may be  $P^{-1}(i)1_{N \times 1} \neq 0$  for some  $i$ . Assume without loss of generality that  $P^{-1}(i)1_{N \times 1} \neq 0$  for  $i = 1, 2, \dots, j$ . Then  $\Lambda_1, \dots, \Lambda_j, \Delta_1, \dots, \Delta_N$  are identified. The following theorem gives a set of sufficient conditions so that there are enough equations to identify  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$ .

**Theorem 5** *Consider the model defined by (1.9) and (1.10). Suppose  $A$  is diagonalizable,  $k \neq 0$ , and the self-consistency assumption holds. If*

(i.)  *$A$  has at least  $D + 1$  distinct eigenvalues so that these distinct  $\xi_i$ 's corresponding eigenvectors satisfy  $P^{-1}(i)1_{N \times 1} \neq 0$ , and*

(ii.)  *$A$  has at least  $L + 1$  distinct eigenvalues,*  
*then  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified.*

A formal proof of theorem 5 is in appendix 1.6. Condition (i.) implies that  $\{I_N, A, \dots, A^D\}$  is linearly independent so that  $\{A1_{N \times 1}, \dots, A^D 1_{N \times 1}\}$  can

help making the endogenous terms  $\{A\omega, \dots, A^D\omega\}$  linearly independent of the exogenous terms  $\{1_{N \times 1}, x, Ax, \dots, A^Lx\}$ , thus the subset  $\{k, J_1, \dots, J_D\}$  is identified. The column  $1_{N \times 1}$  is in a proper subspace of  $\mathbf{R}^N$ ; only the information in the subspaces that are not orthogonal to  $1_{N \times 1}$  remains. Therefore, identification requires that the eigenvalues corresponding to the eigenvectors such that  $P^{-1}(i)1_{N \times 1} \neq 0$  are distinct. Condition (ii.) is obviously needed for the linear independence of the exogenous terms  $\{x, Ax, \dots, A^Lx\}$  in order to distinguish the effects of  $\{c_0, \dots, c_L\}$ .

Theorem 5 reveals that the existence of a constant column  $k1_{N \times 1}$  requires less restrictive conditions for identification if  $A$  is not a normalized adjacency matrix, while the row sums of  $A$  are normalized in some important works.<sup>22</sup>

### 1.2.5 Non-identification and Network Structure

This section discusses the relations between the non-identification of parameters and network structure. Theorems 3, 4, and 5 say that whether the parameters are identified depends critically on the number of the distinct eigenvalues of  $A$  or on the linear independence of a subset of  $\{I_N, A, A^2, \dots\}$ . We will see that the network diameter is closely related to the number of the distinct eigenvalues of  $A$  and the linear independence of  $\{I_N, A, A^2, \dots\}$ .

**Definition 2** *The diameter of a network is the maximal distance between two individuals in the network, where the distance between two individuals in the network is the length of the shortest walk between them.*

The section 4.ii of Blume, Brock, Durlauf, and Ioannides (2011) is a notable advance in the relations between non-identification and network structure. There are five theorems in that section; three of them: corollary 1, theorem 3, and theorem 4 concern network structure. The subsequent discussions show that non-identification imposes an upper bound on the diameter. The non-identification upper bound of the diameter implies the corollary 1, the “if” parts of the theorem 3, and the theorem 4 of Blume, Brock, Durlauf, and Ioannides (2011). The non-identification upper bound helps us to find counterexamples to the “only if” part of the statement ii of their theorem 3. Therefore, diameter is of more primary importance to non-identification than other network properties in the sense that the network properties that are related to non-identification can be derived from the non-identification upper bound of diameter.

The following theorem relates the diameter to the linear independence of a subset of  $\{I_N, A, A^2, \dots\}$  and the number of distinct eigenvalues  $A$ .

---

<sup>22</sup>If there are more than one column of exogenous characteristics in the model (1.9), such as a column of family background  $y$ , the analysis will be similar to that in previous two paragraphs. One major difference is that the support of  $y$  is not in a proper subspace of  $\mathbf{R}^N$ , so there is no condition like  $P^{-1}(i)y \neq 0$ . Additional exogenous regressors could make identification conditions less restrictive.

**Theorem 6**<sup>23</sup> Suppose the diameter of a network is  $d$ .

(i.) If  $A$  is the adjacency matrix of the network (which can be diagonalizable or non-diagonalizable), then  $\{I_N, A, \dots, A^d\}$  is linearly independent; the dimension of the matrix subspace spanned by  $\{I_N, A, A^2, \dots\}$  is at least  $d + 1$ .

(ii.) If  $A$  is the adjacency matrix of the network and  $A$  is diagonalizable, then the number of distinct eigenvalues of  $A$  is at least  $d + 1$ .

Theorem 6 tells us that the parameters are identified if the diameter is long enough. Non-identification imposes an upper bound on the diameter. Theorems 3 and 6 imply that the following theorem.<sup>24</sup>

**Theorem 7** Consider the model defined by (1.9) and (1.10). Suppose  $A$  is diagonalizable without zero rows,  $A$ 's row sums are 1. If  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$  is not identified, then the network diameter is less than or equal to  $D + L - 1$ .

From now on, we consider the model in the section 4.ii of Blume, Brock, Durlauf, and Ioannides (2011). Their model is the model defined by (1.9) and (1.10) where  $D = L = 1$  and the row sums of  $A$  are 1. They considered two types of normalized adjacency matrices: exclusive averaging and inclusive averaging. Denote the number of individuals that  $i$  is connected to as  $n(i)$ . Under exclusive averaging,  $A_{ii} = 0$ ,  $A_{ij} = 1/n(i)$  if there is link from  $i$  to  $j$ , and  $A_{ij} = 0$  otherwise. Under inclusive averaging,  $A_{ii} = 1/(n(i) + 1)$ ,  $A_{ij} = 1/(n(i) + 1)$  if there is link from  $i$  to  $j$ , and  $A_{ij} = 0$  otherwise. That is, self-links are allowed under inclusive averaging;  $i$  himself is included in taking average.

When  $D = L = 1$ ,  $A$  has no zero rows, and  $A$ 's row sums are 1, the diameter must be 1 if the parameters are not identified.<sup>25</sup> This upper bound helps us to determine what network structure leads to non-identification.

First, having diameter 1 immediately implies that the network is transitive. A network is transitive if there is a link from  $i$  to  $k$  whenever there are links from  $i$  to  $j$  and from  $j$  to  $k$ . Suppose there are links from  $i$  to  $j$  and  $j$  to  $k$  but there is no link from  $i$  to  $k$ , then the distance between  $i$  and  $k$  is 2; the diameter is at least 2. Thus non-identification implies transitivity. If the network is undirected, then a connected component<sup>26</sup> of the network is a completely connected network; otherwise, the diameter is larger than 1. These results are the theorem 4 of Blume, Brock, Durlauf, and Ioannides (2011).

<sup>23</sup>(i.) is proven by the argument in the first six lines of the proof of the theorem 5 of Van Mieghem (2011, p.41); the argument in the proof of the proposition 2.6 of Biggs (1974, p.12) works also. Then (i.) and the theorem 2 of this chapter imply (ii.); see corollary 2.7 of Biggs (1974, p.13) for reference also. Although Biggs and Van Mieghem focused on adjacency matrices with entries that are 0 or 1, the arguments in their proofs work for adjacency matrices with entries that are any real number.

<sup>24</sup>When there is no walk of any length connecting two individuals, some books, such as Jackson (2008, p.32), define the distance between them to be infinity. Under this definition, a network having no link has a diameter equal to infinity. But this definition is not suitable for the discussions in this section. When there is no walk of any length connecting two individuals, this chapter define their distance to be zero.

<sup>25</sup>The trivial case where  $A$  is a zero matrix is ignored.

<sup>26</sup>A network (or a component of a network) is connected if there is a walk from  $i$  to  $j$  for all  $i$  and  $j$  in the network (or in the component).

**Theorem 8** Consider the model defined by (1.9) and (1.10). Suppose  $D = L = 1$  and  $A$ 's row sums are 1. Suppose  $\{k, J_1, c_0, c_1\}$  is not identified, then

- (i.) the network is transitive;
- (ii.) the network consists of completely connected components if the network is undirected.

Second, we consider two types of networks with diameter 1: networks consisting of completely connected components and networks consisting of directed stars.

A model with a network consisting of completely connected components is a linear-in-means model. The adjacency matrix is a block diagonal matrix. Suppose the  $b^{\text{th}}$  components consists of  $n_b$  individuals and  $1_{n_b \times n_b}$  is a  $n_b \times n_b$  matrix of 1's. Under inclusive averaging, the  $b^{\text{th}}$  block is  $n_b^{-1}(1_{n_b \times n_b})$ , which has two distinct eigenvalues: 0 and 1. Under exclusive averaging, the  $b^{\text{th}}$  block is  $(n_b - 1)^{-1}(1_{n_b \times n_b} - I_{n_b})$ , which has two distinct eigenvalues: 1 and  $-1/(n_b - 1)$ . Note that the second eigenvalue  $-1/(n_b - 1)$  depends on the component size.

Under inclusive averaging, even if the completely connected components are of different sizes, there are only two distinct eigenvalues: 0 and 1;  $\{k, J_1, c_0, c_1\}$  is not identified. Under exclusive averaging, if the completely connected components are of different sizes, then the number of distinct eigenvalues is the number of components of different sizes plus 1;  $\{k, J_1, c_0, c_1\}$  is not identified if and only if all components are of the same size.

These non-identification results are the “if” parts of the statements i and ii of the theorem 3 of Blume, Brock, Durlauf, and Ioannides (2011). The above arguments also prove the corollary 1 of Blume, Brock, Durlauf, and Ioannides (2011), and the proposition 2 of Bramouille, Djebbari, Fortin (2009, p.45). Moffitt (2001) and Lee (2007) consider linear-in-means models with exclusive averaging. Moffitt considers a network of completely connected components of the same size, so the parameters are not identified. Lee considers a network of completely connected components of different sizes, so the parameters are identified.<sup>27</sup>

Consider networks consisting of directed stars. A directed star is a network in which all individuals are either connected to or connected from (but not both) an individual, who is the center of the star.

Under exclusive averaging, the adjacency matrices of a outward directed star and a inward directed star are respectively

$$A = \frac{1}{N-1} \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \text{ and } A = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}. \quad (1.20)$$

For both outward and inward directed stars,  $A$  is not diagonalizable, and  $A^2 = 0_{N \times N}$ , where  $0_{N \times N}$  is a  $N \times N$  matrix of 0's. But, if  $k \neq 0$ , the proposition 3

<sup>27</sup>Using variations in group size to identify parameters is common in the literature, for example Davezies, D'Haultfoeuille, and Fougere (2009) and Graham (2008)

of Bramoulle, Djebbari, Fortin (2009) shows that  $\{k, J_1, c_0, c_1\}$  is identified.<sup>28</sup>

Under inclusive averaging, the adjacency matrices of an outward directed star and an inward directed star are respectively

$$A = \begin{pmatrix} 1/N & 1/N & \cdots & 1/N \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \text{ and } A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1/2 & 1/2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1/2 & 0 & \cdots & 1/2 \end{pmatrix}. \quad (1.21)$$

For outward directed stars,  $A$  is diagonalizable with two distinct eigenvalues: 1 and  $1/N$ , and  $(-1/N)I_N + (1 + 1/N)A = A^2$ . For networks consists of outward directed stars,  $\{k, J_1, c_0, c_1\}$  is not identified if and only if the outward directed stars are of the same size. For inward directed stars,  $A$  is diagonalizable with two distinct eigenvalues: 1 and  $1/2$ , and  $(-1/2)I_N + (3/2)A = A^2$ . For networks consists of inward directed stars,  $\{k, J_1, c_0, c_1\}$  is not identified.

The results of completely connected components and directed stars are summarized in the following theorem.

**Theorem 9** *Consider the model defined by (1.9) and (1.10). Suppose  $D = L = 1$  and  $A$ 's row sums are 1.*

(i.) *Under exclusive averaging, if the network consists of completely connected components of the same size, then  $\{k, J_1, c_0, c_1\}$  is not identified.*

(ii.) *Under inclusive averaging, if the network consists of completely connected components (which can be of different sizes), outward directed stars of the same size, or inward directed stars (which can be of different sizes), then  $\{k, J_1, c_0, c_1\}$  is not identified.<sup>29</sup>*

The non-identification results of directed stars under inclusive averaging are counterexamples to the “only if” part of the statement ii of the theorem 3 of Blume, Brock, Durlauf, and Ioannides (2011). The parts in (i.) and (ii.) for completely connected components are the “if” parts of their theorem 3.

Bramoulle, Djebbari, and Fortin (2009) and Blume, Brock, Durlauf, and Ioannides (2011) attempted to search for the network properties that lead to non-identification. But a network structure leads to non-identification can be very fancy; for example, under inclusive averaging, if the network consists of outward directed stars of size 2 and inward directed stars, the parameters are not identified (because there are only two distinct eigenvalues: 1 and  $1/2$ ). Perhaps, the most general statement about network structure and non-identification that we can find is a necessary condition on the diameter, such as the condition that the diameter is less than or equal to  $D + L - 1$  in theorem 7.

When  $D$  or  $L$  is larger than one, non-identification allows the diameter to be larger than 1. A diameter larger than 1 allows much more network structures. This is the reason why it is hard to write down a general statement about the

<sup>28</sup>Bramoulle, Djebbari, Fortin (2009) consider social interactions under exclusive averaging.

<sup>29</sup>All “or” in the statement are exclusive.

network properties (besides diameter) when  $\{I_N, A, A^2, A^3\}$  is linearly dependent, such as in Bramoulle, Djebbari, and Fortin (2009, p.48).

The discussions in this section provide a new way to understand identification and network structure. Identification requires enough number of matrices in  $\{I_N, A, A^2, \dots\}$  to be linearly independent so that the endogenous terms of the model are linearly independent of the exogenous terms. Then non-identification requires the linear dependence of a subset of  $\{I_N, A, A^2, \dots\}$ ; such a linear dependence imposes an upper bound on the diameter. This upper bound implies other network properties related to non-identification, such as transitivity, and helps us to search for networks that lead to non-identification.

## 1.3 Differencing

### 1.3.1 Information Loss from Differencing

This section extends the method to models with group-level unobservable variables. Denote an unobservable group-level characteristic as  $\alpha$ . The model is

$$\omega = \alpha 1_{N \times 1} + \sum_{d=1}^D J_d A^d \omega + c_0 x + \sum_{l=1}^L c_l A^l x + \epsilon, \quad (1.22)$$

$$E(\epsilon | \alpha, x) = 0. \quad (1.23)$$

As in the section 3 of Bramoulle, Djebbari and Fortin (2009),  $A$  has no zero rows and the row sums of  $A$  are 1; the reasons for these assumptions will become clear in the subsequent discussions. Since the row sums of  $A$  are 1,  $A^t 1_{N \times 1} = 1_{N \times 1}$  for all  $t$ , and then  $(I_N - \sum_{d=1}^D J_d A^d)^{-1} 1_{N \times 1} = (1 - \sum_{d=1}^D J_d)^{-1} 1_{N \times 1}$ . A reduced form of (1.22) is

$$\omega = \frac{\alpha}{1 - \sum_{d=1}^D J_d} 1_{N \times 1} + \left( I_N - \sum_{d=1}^D J_d A^d \right)^{-1} \left( c_0 x + \sum_{l=1}^L c_l A^l x + \epsilon \right). \quad (1.24)$$

Besides the identification problem in the previous section, the endogeneity of  $\alpha$ ,  $E(\alpha | x) \neq 0$ , is an additional problem.

Bramoulle, Djebbari and Fortin (2009) suggested two methods to handle this endogeneity problem: local differencing and global differencing. Local differencing refers to subtracting from  $\omega_i$  the (weighted) average of  $\omega_j$ 's where  $j$  has a link from  $i$ . Local differencing is achieved by multiplying  $\omega$  by  $I_N - A$ .<sup>30</sup> Global differencing refers to subtracting from  $\omega_i$  the total average of  $\omega_j$ 's. Global differencing is achieved by multiplying  $\omega$  by  $I_N - (1/N)1_{N \times N}$ , where  $1_{N \times N}$  is a  $N \times N$  matrix of 1's. Blume, Brock, Durlauf, and Ioannides (2011, p.894) suggested a third type of differencing by subtracting from  $\omega_i$  the average of  $\omega_j$ 's where  $j$  has an indirect link from  $i$ .

<sup>30</sup>Note that local differencing does not work if  $A$  has zero rows.

These differencing methods are essentially subtracting a subgroup average of  $\omega_j$ 's from each  $\omega_i$  to cancel out the group-level unobservables. This cancellation of the group-level unobservables can be achieved by multiplying  $\omega$  by  $I_N - F$  where  $F$  is a matrix without zero rows and the row sums of  $F$  are 1.

If the row sums of  $A$  are not equal to 1 or if  $A$  has zero rows, then  $\alpha(I_N - \sum_{d=1}^D J_d A^d)^{-1} \mathbf{1}_{N \times 1}$  is not equal to  $\alpha(I_N - \sum_{d=1}^D J_d)^{-1} \mathbf{1}_{N \times 1}$ ; in other words,  $(I_N - \sum_{d=1}^D J_d A^d)^{-1} \mathbf{1}_{N \times 1}$  is not a constant column so that the fixed effects are not represented by a constant column in a reduced form of the model (1.22). Differencing may not work.<sup>31</sup>

Differencing causes a loss of information by projecting  $\omega$  onto a lower dimensional space. This can be easily seen by a change of basis which expresses vectors with respect to the basis consisting of the eigenvectors of the differencing matrix  $F$ . Multiplying  $\omega$  by  $I_N - F$  cancels out the group-level unobservable  $\alpha$ :

$$(I_N - F)\omega = (I_N - F) \left( I_N - \sum_{d=1}^D J_d A^d \right)^{-1} \left( c_0 x + \sum_{l=1}^L c_l A^l x + \epsilon \right). \quad (1.25)$$

Assume  $F$  is diagonalizable. There exists an invertible matrix  $U$  so that  $F = U \text{diag}(\lambda_1, \dots, \lambda_N) U^{-1}$ , where  $\lambda_j$ 's are eigenvalues of  $F$ . A spectral decomposition of  $F$  is  $\sum_{j=1}^N \lambda_j U(j) U^{-1}(j)$ , where  $U(j)$  is the  $j^{\text{th}}$  column of  $U$  and  $U^{-1}(j)$  is the  $j^{\text{th}}$  row of  $U^{-1}$ . Since  $F$  has no zero rows and the row sums of  $F$  are equal to 1, 1 is an eigenvalue of  $F$ .<sup>32</sup> Denote the multiplicity of 1 as  $m_F$ . Assume the first  $m_F$  eigenvalues are 1. Then (1.25) becomes<sup>33</sup>

$$(I_N - F)\omega = \sum_{j=m_F+1}^N (1 - \lambda_j) U(j) U^{-1}(j) \left( I_N - \sum_{d=1}^D J_d A^d \right)^{-1} \times \left( c_0 x + \sum_{l=1}^L c_l A^l x + \epsilon \right). \quad (1.26)$$

$(I_N - F)\omega$  is in the  $N - m_F$  dimensional subspace spanned by  $\{U(m_F + 1), \dots, U(N)\}$ . Differencing causes a “ $m_F$ -dimensional” information loss. The amount of information loss depends on the multiplicity of 1. This suggests that we can minimize the information loss by choosing a differencing matrix  $F$  with  $m_F = 1$ . The matrix  $(1/N)\mathbf{1}_{N \times N}$  (the matrix for global differencing) and the normalized adjacency matrices of connected regular undirected networks<sup>34</sup> are examples of differencing matrices having  $m_F = 1$ .

<sup>31</sup>If the differencing matrix and the adjacency matrix commute, the differencing method works even the row sums of  $A$  are not equal to 1. The model with circulant networks in section 1.3.4 is an example.

<sup>32</sup>1 is the largest eigenvalues  $F$  by the Perron-Frobenius theorem.  $I_N - F$  is not invertible.

<sup>33</sup>Since  $U U^{-1} = I_N$ ,  $\sum_{j=1}^N U(j) U^{-1}(j) = I_N$ , and hence  $I_N - F = \sum_{j=1}^N (1 - \lambda_j) U(j) U^{-1}(j)$ .

<sup>34</sup>Precisely, if we restrict attention to adjacency matrices whose entries are 0 or 1, and these adjacency matrices are subsequently normalized to have row sums equal to 1, then the multiplicity of eigenvalue 1 is equal to 1 for connected regular undirected networks. See the

### 1.3.2 Local Differencing

This subsection considers local differencing. Since  $A$  has no zero rows and  $A$ 's row sums 1, 1 is an eigenvalue of  $A$ . Denote the multiplicity of  $A$ 's eigenvalue 1 as  $m_A$ . Assume the first  $m_A$  eigenvalues are 1. Substituting  $F = A$  and a spectral decomposition of  $A$  into (1.26) give

$$(I_N - A)\omega = \sum_{i=m_A+1}^N (1 - \xi_i)\Delta_i P(i)P^{-1}(i)x + \hat{\epsilon}, \quad (1.27)$$

where

$$\Delta_i \equiv \frac{c_0 + \sum_{l=1}^L c_l \xi_i^l}{1 - \sum_{d=1}^D J_d \xi_i^d}, \quad (1.28)$$

and

$$\hat{\epsilon} \equiv \sum_{i=m_A+1}^N (1 - \xi_i) \frac{1}{1 - \sum_{d=1}^D J_d \xi_i^d} P(i)P^{-1}(i)\epsilon. \quad (1.29)$$

For  $i = 1, \dots, m_A$ , the  $(I_N - A)\omega$ 's coordinate with respect to  $P(i)$  is equal to zero because  $\xi_i = 1$ . For  $i = m_A + 1, \dots, N$ , the coordinate of  $(I_N - A)\omega$  with respect to  $P(i)$  is

$$P^{-1}(i)(I_N - A)\omega = (1 - \xi_i)\Delta_i P^{-1}(i)x + P^{-1}(i)\hat{\epsilon}. \quad (1.30)$$

Assume the support of  $x$  is not in a proper subspace of  $\mathbf{R}^N$ . The coefficient  $\Delta_i$ 's are identified. This gives  $N - m_A$  equations of the form in (1.28). The number of distinct equations is the number of distinct  $\xi_i$  for  $i = m_A + 1, \dots, N$ . The number of linearly independent equations is the number of distinct eigenvalues of  $A$  minus 1.  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has at least  $D + L + 2$  distinct eigenvalues. A formal proof is in appendix 1.6.

**Theorem 10** *Consider the model defined by (1.22) and (1.23). Suppose  $A$  is diagonalizable,  $A$  has no zero rows,  $A$ 's row sums are 1, and the self-consistency assumption holds. With local differencing,  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has at least  $D + L + 2$  distinct eigenvalues, or equivalently,  $\{I_N, A, \dots, A^{D+L+1}\}$  is linearly independent.*

In contrast to theorem 3 (which concerns the model without group-level unobservable), identification with local differencing requires  $A$  to have one more distinct eigenvalue due to the information loss from local differencing. The model in the section 3 of Bramoulle, Djebbari and Fortin (2009) is the case where  $L = D = 1$ . Theorem 10 is consistent with their proposition 4, which states that the linear independence of  $\{I_N, A, A^2, A^3\}$  is necessary and sufficient for identification with local differencing. Theorem 6 and 10 imply their corollary 1 because a diameter longer than or equal to 3 guarantees the linear independence of  $\{I_N, A, A^2, A^3\}$ .

---

proposition 3.1 of Biggs (1974, p.14). A network (or a graph) is regular if all individuals in the network are connected to the same number of individuals. A network is connected if there is a walk from  $i$  to  $j$  for all  $i$  and  $j$  in the network.

### 1.3.3 Information Loss-minimizing Differencing

This subsection considers the type of differencing with a differencing matrix  $F$  so that the multiplicity of 1 is 1:  $m_F = 1$ . Global differencing is a special case. Suppose the first eigenvalue of  $F$  is 1:  $\lambda_1 = 1$ . Then

$$(I_N - F)\omega = \sum_{j=2}^N (1 - \lambda_j)U(j)U^{-1}(j) \sum_{i=1}^N \Delta_i P(i)P^{-1}(i)x + \hat{\epsilon}, \quad (1.31)$$

where  $\Delta_i$  is defined in (1.28) and

$$\hat{\epsilon} \equiv \sum_{j=2}^N (1 - \lambda_j)U(j)U^{-1}(j) \sum_{i=1}^N \frac{1}{1 - \sum_{d=1}^D J_d \xi_i^d} P(i)P^{-1}(i)\epsilon. \quad (1.32)$$

The coordinate of  $(I_N - F)\omega$  with respect to  $U(j)$  is

$$U^{-1}(j)(I_N - F)\omega = (1 - \lambda_j)U^{-1}(j) \sum_{i=2}^N \Delta_i P(i)P^{-1}(i)x + U^{-1}(j)\hat{\epsilon}. \quad (1.33)$$

The summation in (1.33) starts with 2, instead of 1. Here is the reason. Since  $\lambda_1 = 1$ , the eigenvector corresponding to  $\lambda_1$  is  $U(1) = 1_{N \times 1}$ . Recall that  $A$  has no zero rows and  $A$ 's row sums are 1;  $1_{N \times 1}$  is an eigenvector of  $A$  corresponding to eigenvalue 1. Assume without loss of generality that  $P(1) = 1_{N \times 1}$ .<sup>35</sup> Since  $U^{-1}U = I_N$ ,  $U^{-1}(j)U(1) = U^{-1}(j)1_{N \times 1} = 0$  for all  $j = 2, \dots, N$ .

Consider the system of equations for  $j = 2, \dots, N$  of the form in (1.33). The information from  $i = 1$ ,  $\Delta_1$ , is lost from differencing. If  $m_A > 1$  and  $\xi_1 = \xi_2$ , then  $\Delta_1 = \Delta_2$ . The information from the first eigenvalue  $\Delta_1$ , which is lost from differencing, is also contained in the second eigenvalue  $\Delta_2$ . Therefore, if  $m_A > 1$ , the lost information from differencing is not necessary for identification; the parameters are identified under the same conditions for models without group-level unobservable variables. A formal proof is in appendix 1.6.

**Theorem 11** *Consider the model defined by (1.22) and (1.23). Suppose  $A$  is diagonalizable,  $A$  has no zero rows,  $A$ 's row sums are 1,  $m_A > 1$ , and the self-consistency assumption holds. By choosing a differencing matrix  $F$  with  $m_F = 1$ ,  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has at least  $D+L+1$  distinct eigenvalues, or equivalently,  $\{I_N, A, \dots, A^{D+L}\}$  is linearly independent.*

The information from the subspace spanned by  $A$ 's eigenvector  $1_{N \times 1}$ , whose corresponding eigenvalue is 1, is lost from differencing.  $m_A > 1$  means that the lost information is contained in somewhere else: the lost information is contained in the subspace spanned by other  $A$ 's eigenvector(s) corresponding to

<sup>35</sup>If the multiplicity of 1, as an eigenvalue of  $A$ , is larger than 1, we may choose eigenvectors other than  $1_{N \times 1}$  as the basis vectors of the eigenspace corresponding to 1. But changing the basis of the eigenspace corresponding to 1 so that  $1_{N \times 1}$  is not a basis vector will not change the final conclusion. Changing the basis just makes the arguments more complicated.

the eigenvalue 1. So if we choose a correct differencing method, identification does not require more restrictive condition. The proposition 5 of Bramoulle, Djebbari and Fortin (2009, p.48) is about global differencing, which is an example of information loss-minimizing differencing. Since the multiplicity of 1, as an eigenvalue of  $A$ , is the dimension of the null space of  $I_N - A$ , the condition  $m_A > 1$  is equivalent to the condition  $\text{rank}(I_N - A) < N - 1$  in their proposition 5.

### 1.3.4 An Example: Interactions in Circulant Networks

This subsection considers a social interaction model in which social effects are transmitted through circulant networks. The model is more general than the circulant network models in the section 6.vii of Blume, Brock, Durlauf, and Ioannides (2011). The identification problem can be interpreted as an identification problem with partial knowledge of the network structure. The econometrician only know the form of the network (the form of the adjacency matrix) but do not know the strengths of connections (the values of matrix entries). In this example, even though the adjacency matrix for endogenous interactions does not have row sums equal to 1, differencing still works because we can find a differencing matrix that is simultaneously diagonalizable (diagonalizable by the same set of eigenvectors) with the adjacency matrix for endogenous interactions.

Individuals interacting in a circulant network means that individuals are located on a circles, and the social influences between them depend on the clockwise distance between them. The model is

$$\omega = \alpha 1_{N \times 1} + G\omega + Hx + \epsilon. \quad (1.34)$$

The endogenous interaction network  $G$  and the exogenous interaction network  $H$  are circulant:

$$G = \sum_{j=1}^N g_j W^{j-1} \text{ and } H = \sum_{j=1}^N h_j W^{j-1}, \quad (1.35)$$

where  $W$  is the  $N \times N$  circulant matrix

$$W = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (1.36)$$

The circulant matrix  $W$  has  $N$  distinct eigenvalues: 1,  $e^{\frac{2\pi i}{N}}$ ,  $e^{\frac{2\pi i}{N}2}$ , ..., and  $e^{\frac{2\pi i}{N}(N-1)}$ , where  $i^2 = -1$ . Both  $G$  and  $H$  have  $N$  distinct eigenvalues:

$$\xi_r = \sum_{j=1}^N g_j e^{\frac{2\pi i}{N}(j-1)r} \text{ and } \zeta_r = \sum_{j=1}^N h_j e^{\frac{2\pi i}{N}(j-1)r}, \quad (1.37)$$

where  $r = 0, \dots, N - 1$  (Biggs 1974, p.16). Assume  $|\xi_r| < 1$  for all  $r$  so that  $(I_N - G)$  is invertible (self-consistency assumption).

There is a group-level unobservable variable  $\alpha$  in the model (1.34). Since the entries of  $G$  are unknown, local differencing is not a choice. But we can choose any circulant matrix  $F$  with row sums equal to 1 as a differencing matrix:  $F = \sum_{j=1}^N f_j W^{j-1}$ , where  $f_1 + \dots + f_N = 1$ .

$F$ 's eigenvalues are of the same forms as the eigenvalues of  $G$  and  $H$ :  $\lambda_r = \sum_{j=1}^N f_j e^{\frac{2\pi i}{N}(j-1)r}$ , where  $r = 0, \dots, N - 1$ .  $F$  has  $N$  distinct eigenvalues, and the eigenvalue  $\lambda_0 = 1$  has multiplicity 1.  $F$ ,  $G$ , and  $H$  commute so they are simultaneously diagonalizable by a set of eigenvectors  $\{P(1), \dots, P(N)\}$ .<sup>36</sup> Note that the row sums of  $G$  are not equal to 1, but the differencing method still works because  $F$  and  $G$  commute.

Applying differencing to cancel out the group-level unobservable variable and substituting the spectral decompositions of  $F$ ,  $G$ , and  $H$  into the model give

$$(I_N - F)\omega = \sum_{r=1}^{N-1} (1 - \lambda_r) \Delta_r P(r) P^{-1}(r) x + \hat{\epsilon}, \quad (1.38)$$

where

$$\Delta_r \equiv \frac{\zeta_r}{1 - \xi_r}, \quad (1.39)$$

and

$$\hat{\epsilon} \equiv \sum_{r=1}^{N-1} (1 - \lambda_r) \frac{1}{1 - \xi_r} P(r) P^{-1}(r) \epsilon. \quad (1.40)$$

The summation in (1.38) start with 1, instead of 0, because  $\lambda_0 = 1$ . The eigenvalues of  $(I_N - G)^{-1}H$ ,  $\Delta_r$ , are identified. There are  $N - 1$  linearly independent equations of the form in (1.39) for identification. Substituting the explicit formulas of  $\xi_r$  and  $\zeta_r$  into (1.39) give

$$\sum_{j=1}^N g_j e^{\frac{2\pi i}{N}(j-1)r} + \Delta_r \sum_{j=1}^N h_j e^{\frac{2\pi i}{N}(j-1)r} = \Delta_r, \quad (1.41)$$

where  $r = 1, 2, \dots, N - 1$ .

If there are three groups of sizes  $N$ ,  $N + 1$ , and  $N + 2$ , then there are  $3N$  linearly independent equations. If  $N \geq 4$ , the set of  $2N + 4$  parameters:  $\{g_1, \dots, g_{N+2}, h_1, \dots, h_{N+2}\}$  is identified. The key is to have groups of different sizes so that there are enough equations for identification. For example, there are three groups of sizes  $N_1$ ,  $N_2$ , and  $N_3$  where  $N_1 < N_2 < N_3$ . If the number of equations  $(N_1 - 1) + (N_2 - 1) + (N_3 - 1)$  is larger than or equal to the number of parameters  $2N_3$ , then  $\{g_1, \dots, g_{N_3}, h_1, \dots, h_{N_3}\}$  is identified. The last remark is that it is reasonable to assume  $g_1 = 0$  since  $G$  represents the endogenous interactions.

---

<sup>36</sup>Let  $A$  and  $B$  be diagonalizable matrices.  $A$  and  $B$  commute if and only if  $A$  and  $B$  are simultaneously diagonalizable: there exists a basis of the vector space  $V$  such that the basis vectors are eigenvectors simultaneously for  $A$  and  $B$ . See the theorem D4 of Shaw (1982, p.81).

## 1.4 Disentangling the Effects of Correlated Unobservables and Endogenous Interactions

### 1.4.1 Covariance Matrix Decompositions

So far, we have considered models having observable exogenous variables. If there is no observable individual-level heterogeneity, then there is no instrument for the endogenous terms in the models. We have to rely on the information from higher-order moments of the observable dependent variables,  $\omega$ , to identify parameters. This subsection develops the method of identifying parameters by the spectral decompositions of covariance matrices. Graham's (2008) variance contrast method is a special case of the spectral decomposition of covariance matrix because the contrasted variances are scaled eigenvalues of  $Var(\omega)$ . The discussion on the connection to Graham's (2008) method will be in the next subsection.

The model is

$$\omega = k1_{N \times 1} + \sum_{d=1}^D J_d A^d \omega + \epsilon, \quad (1.42)$$

$$Var(\epsilon) = \sigma_0 I_N + \sum_{l=1}^L \sigma_l A^l. \quad (1.43)$$

$\sigma_l$  is the covariance of unobservable individual characteristics between agents at distance  $l$ . Assume the network is undirected so that  $A$  is symmetric. The covariance matrix of  $\epsilon$  in (1.43) is well-defined.<sup>37</sup>

Endogenous interactions are mixed with the effects of correlated unobservables in an infinite degree matrix polynomial:

$$Var(\omega) = \left( I_N - \sum_{d=1}^D J_d A^d \right)^{-1} Var(\epsilon) \left( I_N - \sum_{d=1}^D J_d A^d \right)^{-1}. \quad (1.44)$$

Since  $A$  is symmetric, it is (orthogonally) diagonalizable. After substituting a spectral decomposition of  $A$  into  $Var(\omega)$  and using the idempotence and pairwise orthogonality of  $P(i)P^{-1}(i)$ 's,  $Var(\omega)$  is a finite degree matrix polynomial:

$$Var(\omega) = \sum_{i=1}^N \Delta_i P(i)P^{-1}(i), \quad (1.45)$$

where

$$\Delta_i \equiv \frac{\sigma_0 + \sum_{l=1}^L \sigma_l \xi_i^l}{\left( 1 - \sum_{d=1}^D J_d \xi_i^d \right)^2}. \quad (1.46)$$

$Var(\omega)$  is decomposed into  $N$  pairwise orthogonal matrices. The coefficients of these components are eigenvalues of  $Var(\omega)$ . The only observable variables

<sup>37</sup>Also,  $\sigma_0, \sigma_1, \dots$ , and  $\sigma_L$  are of values so that  $Var(\epsilon)$  is positive definite.

are  $\omega_i$ 's.  $Var(\omega)$  is identified so its eigenvalues,  $\Delta_i$ . The number of distinct  $\xi_i$  determines the number of equations of the form in (1.46). That the terms having  $J_d$ 's in the equations are quadratic suggests that each  $J_d$  has two solutions. But the self-consistency assumption,  $\max_i |\sum_{d=1}^D J_d \xi_i^d| < 1$ , rules out one of them. The set of  $D + L + 1$  parameters  $\{J_1, \dots, J_D, \sigma_0, \dots, \sigma_L\}$  is identified if and only if  $A$  has at least  $D + L + 1$  distinct eigenvalues. This fact is stated below and proven in appendix 1.6.<sup>38</sup>

**Theorem 12** *Consider the model defined by (1.42) and (1.43). Suppose  $A$  is symmetric, and the self-consistency assumption holds.  $\{J_1, \dots, J_D, \sigma_0, \dots, \sigma_L\}$  is identified if and only if  $A$  has at least  $D + L + 1$  distinct eigenvalues, or equivalently,  $\{I_N, A, \dots, A^{D+L}\}$  is linearly independent.*

In appendix 1.7, the method is extended to models with group-level unobservables.

### 1.4.2 An Example: Graham's Linear-in-means Model

This subsection shows that the variance contrast method proposed by Graham (2008) can be treated as a special case of the spectral decomposition of covariance matrix because the contrasted variances are scaled eigenvalues of the covariance matrix of the dependent variables. Graham considered peer effects in classrooms. The variance contrast method was applied to the class size reduction experiment Project STAR. Students were randomly assigned to three types of classes: small (with 13 to 17 students), regular (with 22 to 25 students), and regular with a full time teacher's aide (with 22 to 25 students).

Suppose there are  $N(g, t)$  students in class (group)  $g$  of type  $t$ .  $\omega(g, t)$  is a  $N(g, t) \times 1$  column of academic achievements of students in class  $g$  of type  $t$ .  $\alpha(g, t)$  is a scalar of the teacher effectiveness of class  $g$  of type  $t$  (unobservable group characteristic).  $\epsilon(g, t)$  is a  $N(g, t) \times 1$  column of student productivities in class  $g$  of type  $t$  (unobservable individual characteristics). These variables are written as functions of  $(g, t)$  to allow them to vary with class and class type. Graham's model is the linear-in-means model

$$\omega(g, t) = J_0 S_0 \omega(g, t) + z(g, t), \quad (1.47)$$

where  $z$  is a column of "the total unobservables":

$$z(g, t) = \alpha(g, t) 1_{N(g, t) \times 1} + \epsilon(g, t), \quad (1.48)$$

and  $S_0$  is a matrix taking total class average:

$$S_0 = \frac{1}{N(g, t)} 1_{N(g, t) \times N(g, t)}. \quad (1.49)$$

The diagonal and the off-diagonal entries of  $Var(z(g, t))$ ,  $\gamma_1(g, t) \equiv Var(z_i(g, t))$  and  $\gamma_0(g, t) \equiv Cov(z_i(g, t), z_j(g, t))$ , are constant for all  $i$  and  $j$ :

$$Var(z(g, t)) = \gamma_0(g, t) B_0 + \gamma_1(g, t) B_1, \quad (1.50)$$

<sup>38</sup>The arguments in the proof work even if  $A$  has zero rows.

where

$$\gamma_0(g, t) = \sigma_\alpha^2(g, t) + 2\sigma_{\alpha\epsilon}(g, t) + \sigma_{\epsilon\epsilon}(g, t), \quad (1.51)$$

$$\gamma_1(g, t) = \sigma_\alpha^2(g, t) + 2\sigma_{\alpha\epsilon}(g, t) + \sigma_\epsilon^2(g, t), \quad (1.52)$$

$$B_0 = \mathbf{1}_{N(g,t) \times N(g,t)} - I_{N(g,t)}, \quad (1.53)$$

and

$$B_1 = I_{N(g,t)}. \quad (1.54)$$

In (1.51) and (1.52),  $\sigma_\alpha^2(g, t) \equiv \text{Var}(\alpha(g, t))$ ,  $\sigma_\epsilon^2(g, t) \equiv \text{Var}(\epsilon_i(g, t))$ ,  $\sigma_{\alpha\epsilon}(g, t) \equiv \text{Cov}(\alpha(g, t), \epsilon_i(g, t))$ , and  $\sigma_{\epsilon\epsilon}(g, t) \equiv \text{Cov}(\epsilon_i(g, t), \epsilon_j(g, t))$ .  $\mathbf{1}_{N(g,t) \times N(g,t)}$  is a  $N(g, t) \times N(g, t)$  matrix of 1's, and  $I_{N(g,t)}$  is the  $N(g, t) \times N(g, t)$  identity matrix.

The variance contrast method identifies the endogenous interaction parameter  $J_0$  under two assumptions. First, there is no within-class correlation between unobservable variables:  $\sigma_{\alpha\epsilon}(g, t) = \sigma_{\epsilon\epsilon}(g, t) = 0$ . Second, there is homoskedasticity across class types:  $\sigma_\alpha^2(g, t) = \sigma_\alpha^2$  does not vary with  $(g, t)$ . For the linear-in-means model defined by (1.47)-(1.54), the “within-class variance”,

$$\begin{aligned} & E(G_c^w | g, t) \\ & \equiv E \left[ \frac{1}{N(g, t)(N(g, t) - 1)} \sum_{i=1}^{N(g, t)} \left( \omega_i(g, t) - \frac{\sum_{j=1}^{N(g, t)} \omega_j(g, t)}{N(g, t)} \right)^2 \middle| g, t \right], \end{aligned} \quad (1.55)$$

and the “between-class variance”,

$$E(G_c^b | g, t) \equiv E \left[ \left( \frac{\sum_{i=1}^{N(g, t)} \omega_i(g, t)}{N(g, t)} - E(\omega_i(g, t) | t) \right)^2 \middle| g, t \right], \quad (1.56)$$

vary with class size (Graham 2008, p.649). Since the endogenous interaction parameter  $J_0$  is invariant in  $(g, t)$ ,  $J_0$  can be identified by contrasting (within-class and between-class) variances of classes of different sizes. While Graham shows identification by a particular estimator, Durlauf and Tanaka (2008) provide a general proof of identification for Graham's model.

This chapter's covariance matrix decomposition method allows within-class correlations and heteroskedasticity across class type. It does not only identify  $J_0$ , but also identifies  $\sigma_\alpha^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_\epsilon^2(g, t)$ , and  $\sigma_{\epsilon\epsilon}(g, t)$  for each class type. In other words,  $\sigma_\alpha^2(g, t)$ ,  $\sigma_\epsilon^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ , and  $\sigma_{\epsilon\epsilon}(g, t)$  are allowed to be different across class types; and these variances and covariances for each class type are separately identified. These parameters are of economic interest. Even if students and teachers are randomly assigned to classes as in the Project STAR, it is reasonable to expect that teaching effectiveness is correlated with students' characteristics and that the variances of and the covariances between teacher effectiveness and student productivity vary according to class type.

This subsection uses the version of spectral decomposition that groups terms with the same eigenvalue in one term.<sup>39</sup> The covariance matrix of  $\omega(g, t)$  is

$$\text{Var}(\omega(g, t)) = (I_{N(g,t)} - J_0 S_0)^{-1} \text{Var}(z(g, t)) (I_{N(g,t)} - J_0 S_0)^{-1}. \quad (1.57)$$

<sup>39</sup>See footnote 14 and Speed (1987).

The self-consistency assumption that  $|J_0| < 1$  is required for the invertibility of  $(I_{N(g,t)} - J_0 S_0)$ . A spectral decomposition of  $Var(z(g, t))$  is

$$Var(z(g, t)) = \xi_0(g, t)S_0 + \xi_1(g, t)S_1, \quad (1.58)$$

where

$$\xi_0(g, t) = (N(g, t) - 1)\gamma_0(g, t) + \gamma_1(g, t), \quad (1.59)$$

$$\xi_1(g, t) = -\gamma_0(g, t) + \gamma_1(g, t), \quad (1.60)$$

and

$$S_1 = I_{N(g,t)} - \frac{1}{N(g,t)} 1_{N(g,t) \times N(g,t)}. \quad (1.61)$$

$\xi_0$  and  $\xi_1$  are two distinct eigenvalues of  $Var(z(g, t))$ . Only  $\xi_0(g, t)$  depends on class size  $N(g, t)$ . Note that  $S_0$  in (1.58) is the  $S_0$  that calculates class average in the model (1.47).<sup>40</sup> Substituting (1.58) into (1.57) gives a spectral decomposition of  $Var(\omega(g, t))$ :

$$Var(\omega(g, t)) = \Delta_0(g, t)S_0 + \Delta_1(g, t)S_1 \quad (1.62)$$

where

$$\Delta_0(g, t) \equiv \frac{\xi_0(g, t)}{(1 - J_0)^2} \text{ and } \Delta_1(g, t) \equiv \xi_1(g, t). \quad (1.63)$$

$\Delta_0(g, t)$  and  $\Delta_1(g, t)$  are distinct eigenvalues of  $Var(\omega(g, t))$ .

Since  $Var(\omega(g, t))$  is identified,  $Var(\omega(g, t))$ 's eigenvalues are identified. For each  $(g, t)$ , there are two equations

$$\begin{aligned} N(g, t)\sigma_\alpha^2(g, t) + 2N(g, t)\sigma_{\alpha\epsilon}(g, t) + (N(g, t) - 1)\sigma_{\epsilon\epsilon}(g, t) + \sigma_\epsilon^2(g, t) \\ = \Delta_0(g, t)(1 - J_0)^2 \end{aligned} \quad (1.64)$$

and

$$-\sigma_{\epsilon\epsilon}(g, t) + \sigma_\epsilon^2(g, t) = \Delta_1(g, t). \quad (1.65)$$

But, for each  $(g, t)$ , there are five parameters:  $J_0$ ,  $\sigma_\alpha^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_\epsilon^2(g, t)$ , and  $\sigma_{\epsilon\epsilon}(g, t)$ . Suppose that  $\sigma_\alpha^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_\epsilon^2(g, t)$ , and  $\sigma_{\epsilon\epsilon}(g, t)$  vary with class type but are constant within each class type. Equation (1.64) depends on class size but equation (1.65) does not. If there are four classes of different sizes in each class type (the Project STAR satisfies this condition), then there are five equations to identify five parameters.

$g_1, g_2, g_3$ , and  $g_4$  are classes in a particular class type  $t'$ . These classes have different sizes:  $N_1, N_2, N_3$ , and  $N_4$ . Since these classes are in the same type, the unobservable variables' variances and covariances of these classes are constant:  $\sigma_\alpha^2(g_i, t') = \sigma_\alpha^2(t')$ ,  $\sigma_{\alpha\epsilon}(g_i, t') = \sigma_{\alpha\epsilon}(t')$ ,  $\sigma_\epsilon^2(g_i, t') = \sigma_\epsilon^2(t')$ , and  $\sigma_{\epsilon\epsilon}(g_i, t') = \sigma_{\epsilon\epsilon}(t')$ . There are five equations for identification:

$$N_i\sigma_\alpha^2(t') + 2N_i\sigma_{\alpha\epsilon}(t') + (N_i - 1)\sigma_{\epsilon\epsilon}(t') + \sigma_\epsilon^2(t') = \Delta_0(g_i, t')(1 - J_0)^2 \quad (1.66)$$

<sup>40</sup> $\{B_0, B_1\}$  in (1.53) and (1.54) and  $\{S_0, S_1\}$  in (1.49) and (1.61) are special cases of the set of commuting matrices  $\{A_r\}$  and the set of idempotent and pairwise orthogonal matrices  $\{S_r\}$  defined in the example in p.894 of Speed (1987). The sets of matrices in Speed's example can be used for more general form of linear-in-means models.

for  $i = 1, 2, 3, 4$ , and

$$-\sigma_{\epsilon\epsilon}(t') + \sigma_{\epsilon}^2(t') = \Delta_1(g_1, t') = \Delta_1(g_2, t') = \Delta_1(g_3, t') = \Delta_1(g_4, t'). \quad (1.67)$$

For this class type  $t'$ , the parameters,  $J_0$ ,  $\sigma_{\alpha}^2(t')$ ,  $\sigma_{\alpha\epsilon}(t')$ ,  $\sigma_{\epsilon}^2(t')$ , and  $\sigma_{\epsilon\epsilon}(t')$ , are identified. Hence, since  $J_0$  is identified already, for the remaining two class types, only three classes of different sizes are enough to identify  $\sigma_{\alpha}^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_{\epsilon}^2(g, t)$ , and  $\sigma_{\epsilon\epsilon}(g, t)$ .<sup>41</sup>

**Theorem 13** *Consider the model defined by (1.47)-(1.54). Suppose  $|J_0| < 1$  (self-consistency);  $\sigma_{\alpha}^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_{\epsilon}^2(g, t)$ , and  $\sigma_{\epsilon\epsilon}(g, t)$  vary with  $t$  (class type) but do not vary with  $g$  (class) for a given  $t$ . If*

- (i.) *there are four classes of different sizes for one class type, and*
- (ii.) *there are three classes of different sizes for the remaining class types,*

*then  $J_0$ ,  $\sigma_{\alpha}^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_{\epsilon}^2(g, t)$ , and  $\sigma_{\epsilon\epsilon}(g, t)$  for all class types are identified.*

The variance contrast method uses the within-class (1.55) and between-class (1.56) variances to identify the parameter  $J_0$ . We may treat the variance contrast method as a special case of covariance matrix decomposition because the scaled eigenvalues of  $Var(\omega(g, t))$  are the within-class and between-class variances:<sup>42</sup>

$$E(G_c^w|g, t) = \frac{\Delta_1(g, t)}{N(g, t)} \text{ and } E(G_c^b|g, t) = \frac{\Delta_0(g, t)}{N(g, t)}. \quad (1.68)$$

Indeed, Graham's method of variance contrasts can identify more parameters than the endogenous interaction parameter  $J_0$ . The diagonal elements,  $\gamma_0(g, t)$ , and the off-diagonal elements,  $\gamma_1(g, t)$ , of  $Var(z(g, t))$  in (1.50) can be identified. For a given class type  $t$ , if there are two classes  $g'$  and  $g''$  of sizes  $N'$  and  $N''$ , then Graham's Wald estimator (the equation 9 of Graham (2008)), that identifies  $J_0$ , can be written in terms of the eigenvalues of  $Var(\omega(g, t))$ :

$$\begin{aligned} & \frac{E(G_c^b|g', t) - E(G_c^b|g'', t)}{E(G_c^w|g', t) - E(G_c^w|g'', t)} \\ &= \frac{\frac{\Delta_0(g', t)}{N(g', t)} - \frac{\Delta_0(g'', t)}{N(g'', t)}}{\frac{\Delta_1(g', t)}{N(g', t)} - \frac{\Delta_1(g'', t)}{N(g'', t)}} = \frac{1}{(1 - J_0)^2} \frac{\gamma_0(g', t) - \gamma_0(g'', t)}{\frac{\xi_1(g', t)}{N(g', t)} - \frac{\xi_1(g'', t)}{N(g'', t)}} + \frac{1}{(1 - J_0)^2} \end{aligned} \quad (1.69)$$

Suppose there is no variation in  $\gamma_0(g, t)$  for a given class type,  $\gamma_0(g', t) = \gamma_0(g'', t)$ , the first term in (1.69) vanishes, then  $J_0$  is identified. We do not need  $\sigma_{\alpha\epsilon}(g, t) = \sigma_{\epsilon\epsilon}(g, t) = 0$ , which are implied by the "double randomization" assumption in Graham (2008, p.650). Then, with the identified  $J_0$ , we can calculate  $\xi_0(g', t)$ ,  $\xi_0(g'', t)$ ,  $\xi_1(g', t)$ , and  $\xi_1(g'', t)$  by (1.63), and hence  $\gamma_0(g', t) = \gamma_0(g'', t)$ ,  $\gamma_1(g', t)$ , and  $\gamma_1(g'', t)$  by (1.59) and (1.60).

<sup>41</sup>We may let  $J_0$  vary with class type. Then, for all class types, we need four classes of different sizes.

<sup>42</sup>Consider the spectral decomposition of  $Var(\omega)$  (1.62). The idempotence of  $S_0$  and  $S_1$  implies that  $Var(S_0\omega(g, t)) = \Delta_0(g, t)S_0$  and  $Var(S_1\omega(g, t)) = \Delta_1(g, t)S_1$ . (1.68) can be verified by equating the entries of  $Var(S_0\omega(g, t))$  and  $\Delta_0(g, t)S_0$  and by equating entries of  $Var(S_1\omega(g, t))$  and  $\Delta_1(g, t)S_1$ .

We can choose either  $\gamma_0(g, t)$  or  $\gamma_1(g, t)$  to be constant across  $g$ 's within a given class type. That is, we can choose to have  $\gamma_1(g', t) = \gamma_1(g'', t)$ . Then (1.59), (1.60), and (1.63) provide four equations to identify four parameters:  $J_0$ ,  $\gamma_0(g', t)$ ,  $\gamma_0(g'', t)$ , and  $\gamma_1(g', t) = \gamma_1(g'', t)$ . Hence, for the remaining class types, since  $J_0$  is known,  $\gamma_0(g, t)$  and  $\gamma_1(g, t)$  for any  $g$  are identified by (1.59), (1.60), and (1.63). The important point is that we only need either  $\gamma_0(g', t) = \gamma_0(g'', t)$  or  $\gamma_1(g', t) = \gamma_1(g'', t)$  for one class type. Within the remaining class types, parameters are allowed to vary with  $g$  and  $t$ .<sup>43</sup>

**Theorem 14** *Consider the model defined by (1.47)-(1.54). Suppose  $|J_0| < 1$  (self-consistency). There is a class type that has two classes,  $g'$  and  $g''$ , of different sizes. For this class type,*

*(i.) if  $\gamma_0(g', t) = \gamma_0(g'', t)$ , then  $J_0$ ,  $\gamma_0(g', t) = \gamma_0(g'', t)$ ,  $\gamma_1(g')$ , and  $\gamma_1(g'')$  are identified;*

*(ii.) if  $\gamma_1(g', t) = \gamma_1(g'', t)$ , then  $J_0$ ,  $\gamma_0(g', t)$ ,  $\gamma_0(g'', t)$ , and  $\gamma_1(g', t) = \gamma_1(g'', t)$  are identified.*

*Hence,  $\gamma_0(g, t)$  and  $\gamma_1(g, t)$  for all remaining classes within other class types are identified.*

## 1.5 Conclusions

This chapter does not only develop new methods for identifying social interactions, but also develops an elegant framework which unifies some important results in important works such as Bramouille, Djebbari, and Fortin (2009)<sup>44</sup> and Blume, Brock, Durlauf, and Ioannides (2011)<sup>45</sup>.

The key mathematical structure for the proposed methods is the simultaneous diagonalizability of the endogenous interaction matrix and the exogenous interaction matrix (such as  $\sum_{d=1}^D J_d A^d$  and  $c_0 I_N + \sum_{l=1}^L c_l A^l$  in section 1.2) and the simultaneous diagonalizability of the endogenous interaction matrix and the covariance matrix of unobservable variables (such as  $\sum_{d=1}^D J_d A^d$  and  $Var(\epsilon)$  in section 1.4.1, and  $J_0 S_0$  and  $Var(z)$  in section 1.4.2). That is, these matrices share the same set of eigenvectors. Due to the simultaneous diagonalizability, the numerators and the denominators of the  $\Lambda_i$ 's and  $\Delta_i$ 's in all sections are linear in the parameters.

<sup>43</sup>This chapter's  $1/(1 - J_0)$  is Graham's  $\gamma_0$ . And  $\alpha(g, t)/(1 - J_0)$  is Graham's  $\alpha_c$ . Perhaps, due to different specifications of the linear-in-means model, it is harder to recognize that  $\gamma_0(g, t)$  and  $\gamma_1(g, t)$  are identified in Graham (2008).

<sup>44</sup>Suppose the adjacency matrix is diagonalizable. Bramouille, Djebbari, and Fortin's proposition 1 (2009) are covered by theorems 3 and 4. Their proposition 2 is proven by the discussions concerning completely connected components in section 1.2.5. Their proposition 4 is covered by theorem 10. Their corollary 1 is covered by theorems 6 and 10. Their proposition 5 is a special case of theorem 11.

<sup>45</sup>Suppose the adjacency matrix is diagonalizable. Blume, Brock, Durlauf, and Ioannides's (2011) theorem 2 is covered by theorems 3 and 4. Their corollary 1 is proven by the discussions concerning completely connected components in section 1.2.5. The "if" parts of their theorem 3 are covered by theorem 9. Counterexamples to the statement ii of their theorem 3 are provided by theorem 9. Their theorem 4 is theorem 8.

The proposed methods rely on the diagonalizability of adjacency matrices. The methods are applicable to all undirected networks but only applicable to directed networks with diagonalizable adjacency matrices. The methods are extensible to models having non-diagonalizable adjacency matrices. Also, the generalization of the proposed methods to models having different networks for different types of interactions is left for future work.<sup>46</sup>

## 1.6 Appendix: Proofs of Theorems

**Proof of Theorem 3.** If  $A$  has no zero rows and  $A$ 's row sums are 1, then  $1_{N \times 1}$  is an eigenvector of  $A$  and the corresponding eigenvalue is 1. Assume without loss of generality that  $P(1) = 1_{N \times 1}$ . Then  $P^{-1}(i)1_{N \times 1} = 0$  for  $i = 2, \dots, N$  because  $P^{-1}P = I_N$ . The first term in (1.19),  $\Lambda_i P^{-1}(i)1_{N \times 1}$ , is zero for  $i = 2, \dots, N$ .

Assume the support of  $x$  is not in a proper subspace of  $\mathbf{R}^N$ . The eigenvalues  $\Lambda_1, \Delta_1, \dots$ , and  $\Delta_N$  in (1.19) are identified. We have a system of equations of the forms in (1.16) and (1.17). Write this system in a matrix form

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \Lambda_1 \xi_1 & \cdots & \Lambda_1 \xi_1^D \\ 0 & 1 & \xi_1 & \cdots & \xi_1^L & \Delta_1 \xi_1 & \cdots & \Delta_1 \xi_1^D \\ 0 & 1 & \xi_2 & \cdots & \xi_2^L & \Delta_2 \xi_2 & \cdots & \Delta_2 \xi_2^D \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \xi_N & \cdots & \xi_N^L & \Delta_N \xi_N & \cdots & \Delta_N \xi_N^D \end{pmatrix} \begin{pmatrix} k \\ c_0 \\ c_1 \\ \vdots \\ c_L \\ J_1 \\ \vdots \\ J_D \end{pmatrix} = \begin{pmatrix} \Lambda_1 \\ \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_N \end{pmatrix}. \quad (1.70)$$

Identification of  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$  requires that the  $(N+1) \times (D+L+2)$  matrix has rank  $D+L+2$ . Note that  $k$  appears only in the first equation of the system, and  $c_0, c_1, \dots$ , and  $c_L$  appears only in the last  $N$  equations. Thus the  $N \times (D+L+1)$  submatrix, obtained by cancelling the first row and the first column, is required to have rank  $D+L+1$ .

First, if  $A$  has less than  $D+L+1$  distinct eigenvalues, then the  $N \times (D+L+1)$  submatrix cannot have rank larger than or equal to  $D+L+1$ . The parameters are not identified. Second if  $A$  has at least  $D+L+1$  distinct eigenvalues, then we can show that the columns of the  $N \times (D+L+1)$  submatrix are linearly independent. This implies that  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified and then  $k$  is identified by the first equation.

Recall from (1.17) that  $\Delta_i \equiv (c_0 + \sum_{l=1}^L c_l \xi_i^l) / (1 - \sum_{d=1}^D J_d \xi_i^d)$ .  $\Delta_i$ 's are infinite degree polynomials in  $\xi_i$ . Define

$$\sum_{t=0}^{\infty} \beta_t \xi_i^t \equiv \frac{c_0 + \sum_{l=1}^L c_l \xi_i^l}{1 - \sum_{d=1}^D J_d \xi_i^d}. \quad (1.71)$$

<sup>46</sup>Blume, Brock, Durlauf, and Jayaraman (2011) consider the identification problem when endogenous and exogenous interactions are transmitted through different networks.

Suppose there are scalars  $\varphi_0, \dots, \varphi_L, \psi_1, \dots,$  and  $\psi_D$  so that

$$\begin{aligned} & \varphi_0 + \varphi_1 \xi_i + \dots + \varphi_L \xi_i^L \\ & + \psi_1 \xi_i \sum_{t=0}^{\infty} \beta_t \xi_i^t + \psi_2 \xi_i^2 \sum_{t=0}^{\infty} \beta_t \xi_i^t + \dots + \psi_D \xi_i^D \sum_{t=0}^{\infty} \beta_t \xi_i^t = 0, \end{aligned} \quad (1.72)$$

for all  $i$ . If  $\varphi_0 = \dots = \varphi_L = \psi_1 = \dots = \psi_D = 0$ , then the columns of the  $(D+L+1) \times N$  submatrix are linearly independent. Higher degree terms  $\xi_i^t$  where  $t > \max(L, D)$  appear only in the last  $D$  terms of (1.72). Since (1.72) holds for all  $\xi_i$ ,

$$\psi_1 \beta_{j-1} = \psi_2 \beta_{j-2} = \dots = \psi_D \beta_{j-D} = 0 \quad (1.73)$$

for  $j > \max(L, D)$ . Hence the first  $L+1$  terms of (1.72) force  $\varphi_i = 0$  for  $i = 0, 1, \dots, L$ .

**Proof of Theorem 4.** If  $A$  has zero rows and the row sums of the non-zero rows are 1,  $1_{N \times 1}$  cannot be an eigenvector corresponding to any eigenvalues of  $A$ .  $1_{N \times 1}$  must be in a space spanned by the eigenvectors corresponding to at least two distinct eigenvalues. Write  $1_{N \times 1}$  as a linear combination of the eigenvector of  $A$ :  $1_{N \times 1} = \sum_{i=1}^N \beta_i P(i)$ . There are  $s$  and  $t$  so that  $\beta_s \neq 0$ ,  $\beta_t \neq 0$ , and  $\xi_s \neq \xi_t$ . We have  $P^{-1}(s)1_{N \times 1} = \beta_s \neq 0$  and  $P^{-1}(t)1_{N \times 1} = \beta_t \neq 0$ . For these  $s$  and  $t$ ,  $\Lambda_s$  and  $\Lambda_t$  are identified.

Assume the support of  $x$  is not in a proper subspace of  $\mathbf{R}^N$ . We identify  $\Lambda_s$ ,  $\Lambda_t$ ,  $\Delta_1, \dots,$  and  $\Delta_N$  in (1.19). They give a system of equations

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 & \Lambda_s \xi_s & \dots & \Lambda_s \xi_s^D \\ 1 & 0 & 0 & \dots & 0 & \Lambda_t \xi_t & \dots & \Lambda_t \xi_t^D \\ 0 & 1 & \xi_1 & \dots & \xi_1^L & \Delta_1 \xi_1 & \dots & \Delta_1 \xi_1^D \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \xi_N & \dots & \xi_N^L & \Delta_N \xi_N & \dots & \Delta_N \xi_N^D \end{pmatrix} \begin{pmatrix} k \\ c_0 \\ c_1 \\ \vdots \\ c_L \\ J_1 \\ \vdots \\ J_D \end{pmatrix} = \begin{pmatrix} \Lambda_s \\ \Lambda_t \\ \Delta_1 \\ \vdots \\ \Delta_N \end{pmatrix}. \quad (1.74)$$

Note that the first two rows are linearly independent and cannot be in a space spanned by the last  $N$  rows. By a similar argument to the last paragraph of the proof of theorem 3, if  $A$  has  $D+L$  distinct eigenvalues, there are  $D+L+2$  linearly independent equations so that  $\{k, J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified. This proves the statement (i.).

The above arguments require  $k \neq 0$ . If  $k = 0$ ,  $\Lambda_i \equiv 0$  for all  $i$ ; the equations from  $\Lambda_s$  and  $\Lambda_t$  are not available. By the proof of theorem 3 without the equation from  $\Lambda_1$ ,  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has  $D+L+1$  distinct eigenvalues. This proves the statement (ii.).

**Proof of Theorem 5.** If  $A$ 's row sums are not normalized to be 1, we may have  $P^{-1}(i)1_{N \times 1} \neq 0$  for some  $i$ . Assume without loss of generality that  $P^{-1}(i)1_{N \times 1} \neq 0$  for  $i = 1, \dots, j$ .

Assume the support of  $x$  is not in a proper subspace of  $\mathbf{R}^N$ . The eigenvalues  $\Lambda_1, \dots, \Lambda_j, \Delta_1, \dots, \Delta_N$  in (1.19) are identified. We have a system of equations of the forms in (1.16) and (1.17). Write this system in a matrix form

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & \Lambda_1 \xi_1 & \cdots & \Lambda_1 \xi_1^D \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 & \Lambda_j \xi_j & \cdots & \Lambda_j \xi_j^D \\ 0 & 1 & \xi_1 & \cdots & \xi_1^L & \Delta_1 \xi_1 & \cdots & \Delta_1 \xi_1^D \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \xi_N & \cdots & \xi_N^L & \Delta_N \xi_N & \cdots & \Delta_N \xi_N^D \end{pmatrix} \begin{pmatrix} k \\ c_0 \\ c_1 \\ \vdots \\ c_L \\ J_1 \\ \vdots \\ J_D \end{pmatrix} = \begin{pmatrix} \Lambda_1 \\ \vdots \\ \Lambda_j \\ \Delta_1 \\ \vdots \\ \Delta_N \end{pmatrix}. \quad (1.75)$$

Suppose there are  $D + 1$  distinct  $\xi_i$  for  $i \in \{1, \dots, j\}$ . By an argument similar to the last paragraph in the proof of theorem 3, the submatrix consisting of the first  $j$  rows of the  $(j + N) \times (D + L + 2)$  matrix in (1.75) has rank  $D + 1$ . Therefore  $\{k, J_1, \dots, J_D\}$  is identified by the first  $j$  equations in (1.75). Suppose there are  $L + 1$  distinct  $\xi_i$  for  $i \in \{1, \dots, N\}$ . Then  $\{c_0, \dots, c_L\}$  is identified by the last  $N$  equations in (1.75). This proves the theorem.

**Proof of Theorem 10.** In (1.30), the first  $m_A$  coordinates are equal to zero because  $\xi_i = 1$  for  $i = 1, \dots, m_A$ . Assume the support of  $x$  is not in a proper subspace of  $\mathbf{R}^N$ .  $\Delta_i$  for  $i = m_A + 1, \dots, N$  are identified because  $\xi_i$ 's are known. We have a system of equations of the form in (1.28)

$$\begin{pmatrix} 1 & \xi_{m_A+1} & \cdots & \xi_{m_A+1}^L & \Delta_{m_A+1} \xi_{m_A+1} & \cdots & \Delta_{m_A+1} \xi_{m_A+1}^D \\ 1 & \xi_{m_A+2} & \cdots & \xi_{m_A+2}^L & \Delta_{m_A+2} \xi_{m_A+2} & \cdots & \Delta_{m_A+2} \xi_{m_A+2}^D \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_N & \cdots & \xi_N^L & \Delta_N \xi_N & \cdots & \Delta_N \xi_N^D \end{pmatrix} \times \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_L \\ J_1 \\ \vdots \\ J_D \end{pmatrix} = \begin{pmatrix} \Delta_{m_A+1} \\ \Delta_{m_A+2} \\ \vdots \\ \Delta_N \end{pmatrix}. \quad (1.76)$$

By an argument similar to that in the last paragraph of the proof of theorem 3, the  $(N - m_A) \times (D + L + 1)$  matrix in (1.76) has rank at least  $D + L + 1$  if and only if there are at least  $D + L + 1$  distinct  $\xi_i$  for  $i = m_A + 1, \dots, N$ . Thus

$\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has at least  $D + L + 2$  distinct eigenvalues.

**Proof of Theorem 11.** If we choose a difference matrix  $F$  with  $m_F = 1$ , then the reduced form of the model (1.22) after changing the basis is

$$\begin{aligned} & \begin{pmatrix} U^{-1}(2)(I_N - F)\omega \\ U^{-1}(3)(I_N - F)\omega \\ \vdots \\ U^{-1}(N)(I_N - F)\omega \end{pmatrix} \\ &= \begin{pmatrix} (1 - \lambda_2)\Delta_2 U^{-1}(2)P(2) & \cdots & (1 - \lambda_2)\Delta_N U^{-1}(2)P(N) \\ (1 - \lambda_3)\Delta_2 U^{-1}(3)P(2) & \cdots & (1 - \lambda_3)\Delta_N U^{-1}(3)P(N) \\ \vdots & \ddots & \vdots \\ (1 - \lambda_N)\Delta_2 U^{-1}(N)P(2) & \cdots & (1 - \lambda_N)\Delta_N U^{-1}(N)P(N) \end{pmatrix} \\ &\times \begin{pmatrix} P^{-1}(2)x \\ P^{-1}(3)x \\ \vdots \\ P^{-1}(N)x \end{pmatrix} + \begin{pmatrix} P^{-1}(2)\hat{\epsilon} \\ P^{-1}(3)\hat{\epsilon} \\ \vdots \\ P^{-1}(N)\hat{\epsilon} \end{pmatrix}. \end{aligned} \quad (1.77)$$

The first coordinate of  $(I_N - F)\omega$  with respect to  $U(1)$  is zero because the first eigenvalue of  $F$ ,  $\lambda_1$ , is 1.

Assume the support of  $x$  is not in a proper subspace of  $\mathbf{R}^N$ . The  $(N - 1) \times (N - 1)$  matrix in (1.77) is identified. For all  $j$  and  $i$ , the eigenvalue  $\lambda_j$  and the scalar  $U^{-1}(j)P(i)$  are known. So  $\Delta_i$  for  $i = 2, \dots, N$  are identified. We have equations of the form in (1.28) to identify  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$ . The number of equations that we have is the number of distinct  $\xi_i$  for  $i = 2, \dots, N$ . Recall that  $\xi_1 = \xi_2 = 1$ . So the number of equations that we have is the number of distinct eigenvalues of  $A$ . Note that the system of equations of the form in (1.28) is of a form similar to (1.76). A reasoning similar to that in the last paragraph of the proof of theorem 3 finishes the proof of theorem 11.

**Proof of Theorem 12.** The eigenvalues of  $Var(\omega)$  are identified. This give us a system of quadratic equations of the form in (1.46):

$$\begin{pmatrix} 1 & \xi_1 & \cdots & \xi_1^L \\ 1 & \xi_2 & \cdots & \xi_2^L \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_N & \cdots & \xi_N^L \end{pmatrix} \begin{pmatrix} \sigma_0 \\ \sigma_1 \\ \vdots \\ \sigma_L \end{pmatrix} = \begin{pmatrix} \Delta_1 \left(1 - \sum_{d=1}^D J_d \xi_1^d\right)^2 \\ \Delta_2 \left(1 - \sum_{d=1}^D J_d \xi_2^d\right)^2 \\ \vdots \\ \Delta_N \left(1 - \sum_{d=1}^D J_d \xi_N^d\right)^2 \end{pmatrix}. \quad (1.78)$$

If the number of distinct eigenvalues of  $A$  is smaller than the number of parameters  $D + L + 1$ , the number of equations in (1.78) is smaller than the number of parameters.  $\{J_1, \dots, J_D, \sigma_0, \dots, \sigma_L\}$  is not identified.

Suppose the number of distinct eigenvalues of  $A$  is larger than or equal to the number of parameters,  $D + L + 1$ . Define

$$\Xi_1 \equiv \begin{pmatrix} 1 & \xi_1 & \cdots & \xi_1^L \\ 1 & \xi_2 & \cdots & \xi_2^L \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_{L+1} & \cdots & \xi_{L+1}^L \end{pmatrix} \text{ and } \Xi_2 \equiv \begin{pmatrix} 1 & \xi_{L+2} & \cdots & \xi_{L+2}^L \\ 1 & \xi_{L+3} & \cdots & \xi_{L+3}^L \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_N & \cdots & \xi_N^L \end{pmatrix}. \quad (1.79)$$

The  $(L+1) \times (L+1)$  matrix,  $\Xi_1$ , in (1.79) is invertible since it is a Vandermonde matrix.<sup>47</sup> First, consider the first  $L+1$  equations in (1.78). Since  $\Xi_1$  is invertible, there is a unique solution of  $\{\sigma_0, \dots, \sigma_L\}$  in terms of  $\{J_1, \dots, J_D\}$ :

$$\begin{pmatrix} \sigma_0 \\ \sigma_1 \\ \vdots \\ \sigma_L \end{pmatrix} = \Xi_1^{-1} \begin{pmatrix} \Delta_1 \left(1 - \sum_{d=1}^D J_d \xi_1^d\right)^2 \\ \Delta_2 \left(1 - \sum_{d=1}^D J_d \xi_2^d\right)^2 \\ \vdots \\ \Delta_{L+1} \left(1 - \sum_{d=1}^D J_d \xi_{L+1}^d\right)^2 \end{pmatrix}. \quad (1.80)$$

Second, consider the last  $N - L - 1$  equations of (1.78):

$$\Xi_2 \Xi_1^{-1} \begin{pmatrix} \Delta_1 \left(1 - \sum_{d=1}^D J_d \xi_1^d\right)^2 \\ \Delta_2 \left(1 - \sum_{d=1}^D J_d \xi_2^d\right)^2 \\ \vdots \\ \Delta_{L+1} \left(1 - \sum_{d=1}^D J_d \xi_{L+1}^d\right)^2 \end{pmatrix} = \begin{pmatrix} \Delta_{L+2} \left(1 - \sum_{d=1}^D J_d \xi_{L+2}^d\right)^2 \\ \Delta_{L+3} \left(1 - \sum_{d=1}^D J_d \xi_{L+3}^d\right)^2 \\ \vdots \\ \Delta_N \left(1 - \sum_{d=1}^D J_d \xi_N^d\right)^2 \end{pmatrix}. \quad (1.81)$$

Denote the entries of the  $(N - L - 1) \times (L + 1)$  matrix  $\Xi_2 \Xi_1^{-1}$  as

$$\Xi_2 \Xi_1^{-1} \equiv \begin{pmatrix} \pi_{L+2,1} & \pi_{L+2,2} & \cdots & \pi_{L+2,L+1} \\ \pi_{L+3,1} & \pi_{L+3,2} & \cdots & \pi_{L+3,L+1} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{N,1} & \pi_{N,2} & \cdots & \pi_{N,L+1} \end{pmatrix}. \quad (1.82)$$

Suppose the size of the network is large enough so that  $N - L - 1 \geq D$ . There are at least  $D$  quadratic equations (1.81) to solve for  $\{J_1, \dots, J_D\}$ . Each  $J_d$  has two solutions that satisfy (1.81). But the self-consistency assumption,  $\max_i |\sum_{d=1}^D J_d \xi_i^d| < 1$ , rules out one of them. Taking square root on both sides of each equation in (1.81) gives

$$\begin{aligned} \sqrt{\pi_{j,1} \frac{\Delta_1}{\Delta_j} \left(1 - \sum_{d=1}^D J_d \xi_1^d\right)^2 + \cdots + \pi_{j,L+1} \frac{\Delta_{L+1}}{\Delta_j} \left(1 - \sum_{d=1}^D J_d \xi_{L+1}^d\right)^2} \\ = \pm \left(1 - \sum_{d=1}^D J_d \xi_j^d\right) \end{aligned} \quad (1.83)$$

<sup>47</sup>See the theorem 2 of Nicholson (2002, p.129).

for  $j = L + 2, \dots, N$ . Each solution of  $\{J_1, \dots, J_D\}$  is determined by a particular choice of “+” or “-” on the right side of (1.83) for each  $j$ . Thus we have  $2D$  solutions of  $\{J_1, \dots, J_D\}$ . Among these solutions,  $2D - 1$  of them are ruled out by the assumption that  $\max_i |\sum_{d=1}^D J_d \xi_i^d| < 1$ . If there is a  $j$  so that the right hand side of (1.83) has a minus sign, then

$$\begin{aligned} & \sum_{d=1}^D J_d \xi_j^d \\ &= 1 + \sqrt{\pi_{j,1} \frac{\Delta_1}{\Delta_j} \left(1 - \sum_{d=1}^D J_d \xi_1^d\right)^2 + \dots + \pi_{j,L+1} \frac{\Delta_{L+1}}{\Delta_j} \left(1 - \sum_{d=1}^D J_d \xi_{L+1}^d\right)^2} \\ &\geq 1. \end{aligned} \tag{1.84}$$

But (1.84) violates  $\max_i |\sum_{d=1}^D J_d \xi_i^d| < 1$ . The only solution remains is the one having “+” in the right hand side of (1.83) for  $j = L + 2, \dots, N$ . That is, the only solution is determined by

$$\begin{aligned} & \sqrt{\pi_{j,1} \frac{\Delta_1}{\Delta_j} \left(1 - \sum_{d=1}^D J_d \xi_1^d\right)^2 + \dots + \pi_{j,L+1} \frac{\Delta_{L+1}}{\Delta_j} \left(1 - \sum_{d=1}^D J_d \xi_{L+1}^d\right)^2} \\ &= 1 - \sum_{d=1}^D J_d \xi_j^d \end{aligned} \tag{1.85}$$

for  $j = L + 2, \dots, N$ . The set of parameters  $\{\sigma_0, \dots, \sigma_L, J_1, \dots, J_D\}$  is identified.

## 1.7 Appendix: Covariance Matrix Decompositions with Differencing

This appendix extends the covariance matrix decomposition method to models having group-level unobservable variables and discusses the differencing methods. For models without observable independent variables, the existence of group-level unobservable variables does not cause the endogeneity problem as in section 1.3. The differencing methods in this appendix suggest ways of estimation when the structure of the covariance matrix of the unobservable variables is not close to the structure of the endogenous interaction matrix. If the structure of the covariance matrix of the unobservable variables (including both the individual-level and group-level unobservable variables) is close enough to the structure of the endogenous interaction matrix, we do not need to use differencing to cancel out the group-level unobservable variables. The linear-in-means model in section 1.4.2 is an example that differencing is not necessary for identification of all parameters of interest. This is because the covariance matrix of the unobservable variables and the endogenous interaction matrix commute

(so that the covariance matrix of the unobservable variables and the endogenous interaction matrix are simultaneously diagonalizable by the same set of eigenvectors).

As in sections 1.3 and 1.4.1,  $A$  has no zero rows and the row sums of  $A$  are 1 so that differencing methods work;  $A$  is a symmetric matrix so that the covariance matrices are well-defined. The model is

$$\omega = \alpha 1_{N \times 1} + \sum_{d=1}^D J_d A^d \omega + \epsilon, \quad (1.86)$$

$$\text{Var}(\epsilon) = \sigma_0 I_N + \sum_{l=1}^L \sigma_l A^l, \quad (1.87)$$

$$\text{Var}(\alpha 1_{N \times 1} + \epsilon) = \sigma_\alpha^2 1_{N \times N} + \text{Var}(\epsilon). \quad (1.88)$$

$\sigma_\alpha^2 \equiv \text{Var}(\alpha^2)$ .  $\alpha$  and  $\epsilon_i$ 's are assumed to be uncorrelated.

First, we consider local differencing. Suppose the multiplicity of 1 is  $m_A$  and the first  $m_A$  eigenvalues are equal to 1, then the covariance matrix of the locally differenced  $\omega$  is

$$\text{Var}((I_N - A)\omega) = \sum_{i=m_A+1}^N (1 - \xi_i)^2 \Delta_i P(i) P^{-1}(i). \quad (1.89)$$

Recall from (1.46) that  $\Delta_i \equiv (\sigma_0 + \sum_{l=1}^L \sigma_l \xi_i^l) / (1 - \sum_{d=1}^D J_d \xi_i^d)^2$ .  $\Delta_i$ 's are eigenvalues of  $(I_N - \sum_{d=1}^D J_d A^d)^{-1} \text{Var}(\epsilon) (I_N - \sum_{d=1}^D J_d A^d)^{-1}$ . Since  $(1 - \xi_i)^2$ 's are known constants,  $\Delta_i$  for  $i = m_A + 1, \dots, N$  are identified.

We have a system of  $N - m_A$  equations, and this system of equations is exactly the system of equations in (1.78) without the first  $m_A$  equations. The number of distinct equations that we have is the number of distinct  $\xi_i$  for  $i = m_A + 1, \dots, N$ . The number of equations is the number of distinct eigenvalues of  $A$  minus one. By a proof similar to the proof of theorem 12,  $\{J_1, \dots, J_D, \sigma_0, \dots, \sigma_L\}$  is identified if and only if  $A$  has at least  $D + L + 2$  distinct eigenvalues. Once  $\{J_1, \dots, J_D, \sigma_0, \dots, \sigma_L\}$  is identified,  $\sigma_\alpha^2$  is identified by subtracting  $\text{Var}(\epsilon)$  from  $\text{Var}(\alpha 1_{N \times 1} + \epsilon)$ :

$$\begin{aligned} \sigma_\alpha^2 1_{N \times N} &= \text{Var}(\alpha 1_{N \times 1} + \epsilon) - \text{Var}(\epsilon) \\ &= \left( I_N - \sum_{d=1}^D J_d A^d \right) \text{Var}(\omega) \left( I_N - \sum_{d=1}^D J_d A^d \right) - \text{Var}(\epsilon) \end{aligned} \quad (1.90)$$

**Theorem 15** *Consider the model defined by (1.86), (1.87), and (1.88). Suppose  $A$  is symmetric,  $A$  has no zero rows,  $A$ 's row sums are 1, and the self-consistency assumption holds. With local differencing,  $\{J_1, \dots, J_D, \sigma_0, \dots, \sigma_L, \sigma_\alpha^2\}$  is identified if and only if  $A$  has at least  $D + L + 2$  distinct eigenvalues, or equivalently,  $\{I_N, A, \dots, A^{D+L+1}\}$  is linearly independent.*

Second, we choose a symmetric differencing matrix  $F$  so that  $F$  has no zero rows, the row sums of  $F$  are 1, and the eigenvalue 1 has multiplicity one:  $m_F = 1$ . Since  $F$  is symmetric,  $F$  is (orthogonally) diagonalizable. There exists an invertible (and orthogonal) matrix  $U$  so that  $F = U \text{diag}(\lambda_1, \dots, \lambda_N) U^{-1}$ , where  $\lambda_1, \lambda_2, \dots$ , and  $\lambda_N$  are eigenvalues of  $F$ . Denote the  $i^{\text{th}}$  column of  $U$  as  $U(i)$  and the  $i^{\text{th}}$  row of  $U^{-1}$  as  $U^{-1}(i)$ . Indeed  $U^{-1}(i) = U(i)'$  because  $U$  is orthogonal.<sup>48</sup> Assume without loss of generality that the first eigenvalue of  $F$  is 1:  $\lambda_1 = 1$ . The covariance matrix of the differenced  $\omega$  is

$$\begin{aligned} & \text{Var}((I_N - F)\omega) \\ &= \sum_{j=2}^N (1 - \lambda_j) U(j) U(j)' \sum_{i=2}^N \Delta_i P(i) P(i)' \sum_{h=2}^N (1 - \lambda_h) U(h) U(h)'. \end{aligned} \quad (1.91)$$

Since  $\lambda_1 = 1$ ,  $U(1) = 1_{N \times 1}$ .  $A$  has no zero rows and  $A$ 's row sums are 1, so  $1_{N \times 1}$  is also an eigenvector of  $A$  (and the corresponding eigenvalue is 1). Assume  $P(1) = 1_{N \times 1}$ . Then  $U(j)$  for  $j = 2, \dots, N$  are simultaneously orthogonal to  $P(1) = 1_{N \times 1}$  because  $UU' = I_N$ . This is the reason why the summation in middle of (1.91) start with  $i = 2$ , instead of 1.

Left multiplying and right multiplying  $\text{Var}((I_N - F)\omega)$  by  $U(j)U(j)'$  give

$$\begin{aligned} & U(j)U(j)' \text{Var}((I_N - F)\omega) U(j)U(j)' \\ &= (1 - \lambda_j) U(j)U(j)' \sum_{i=2}^N \Delta_i P(i) P(i)' (1 - \lambda_j) U(j)U(j)' \\ &= \left[ (1 - \lambda_j)^2 U(j)' \sum_{i=2}^N \Delta_i P(i) P(i)' U(j) \right] U(j)U(j)'. \end{aligned} \quad (1.92)$$

In the third line of (1.92), the term in the square bracket is a scalar, and this scalar is identified. We identified  $N - 1$  coefficients in the square bracket for  $j = 2, \dots, N$ . These identified coefficients give a system of  $N - 1$  equations

$$\begin{aligned} & \begin{pmatrix} \frac{U(2)' \text{Var}(I_N - F)\omega U(2)}{(1 - \lambda_2)^2} \\ \frac{U(3)' \text{Var}(I_N - F)\omega U(3)}{(1 - \lambda_3)^2} \\ \vdots \\ \frac{U(N)' \text{Var}(I_N - F)\omega U(N)}{(1 - \lambda_N)^2} \end{pmatrix} \\ &= \begin{pmatrix} U(2)' P(2) P(2)' U(2) & \cdots & U(2)' P(N) P(N)' U(2) \\ U(3)' P(2) P(2)' U(3) & \cdots & U(3)' P(N) P(N)' U(3) \\ \vdots & \ddots & \vdots \\ U(N)' P(2) P(2)' U(N) & \cdots & U(N)' P(N) P(N)' U(N) \end{pmatrix} \begin{pmatrix} \Delta_2 \\ \Delta_3 \\ \vdots \\ \Delta_N \end{pmatrix}. \end{aligned} \quad (1.93)$$

The eigenvalues  $\Delta_i$ 's for  $i = 2, \dots, N$  are identified if and only if the  $(N - 1) \times (N - 1)$  matrix in (1.93) is invertible, and it is indeed the case.

<sup>48</sup>Prime denotes matrix transpose.

First,  $\{P(1), \dots, P(N)\}$  and  $\{U(1), \dots, U(N)\}$  are bases of  $\mathbf{C}^N$ . Also,  $P(1) = U(1) = \mathbf{1}_{N \times 1}$ . For each  $j = 2, \dots, N$ , the vector  $U(j)$  cannot be simultaneously orthogonal to all  $P(i)$  for  $i = 1, \dots, N$ . There is no zero row in the  $(N-1) \times (N-1)$  matrix. Similarly, for each  $i = 2, \dots, N$ , the vector  $P(i)$  cannot be simultaneously orthogonal to all  $U(j)$  for  $j = 1, \dots, N$ . There is no zero column in the  $(N-1) \times (N-1)$  matrix. Second, using the fact that  $\{P(2), \dots, P(N)\}$  and  $\{U(2)', \dots, U(N)'\}$  are two sets of linearly independent vectors, we can prove that the  $(N-1) \times (N-1)$  matrix is invertible. So  $\Delta_i$  for  $i = 2, \dots, N$  are identified. We have equations of the form in (1.46),  $\Delta_i = (\sigma_0 + \sum_{l=1}^L \sigma_l \xi_i^l) / (1 - \sum_{d=1}^D J_d \xi_i^d)^2$ , to identify  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$ . The number of equations is the number of distinct  $\xi_i$  for  $i = 2, \dots, N$ .

Suppose  $m_A > 1$  and  $\xi_1 = \xi_2 = 1$ . The number of equations is the number of distinct eigenvalues of  $A$ . The system of equations is of the same form in (1.78) without the first equation. A reasoning similar to the proof of theorem 12 proves that  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified if and only if  $A$  has at least  $D + L + 1$  distinct eigenvalues. Once  $\{J_1, \dots, J_D, c_0, \dots, c_L\}$  is identified,  $\sigma_\alpha^2$  is identified by subtracting  $Var(\epsilon)$  from  $Var(\alpha \mathbf{1}_{N \times 1} + \epsilon)$ .

**Theorem 16** *Consider the model defined by (1.86), (1.87) and (1.88). Suppose  $A$  is symmetric,  $A$  has no zero rows,  $A$ 's row sums are 1,  $m_A > 1$ , and the self-consistency assumption holds. By choosing a symmetric differencing matrix  $F$  with  $m_F = 1$ ,  $\{J_1, \dots, J_D, \sigma_0, \dots, \sigma_L, \sigma_\alpha^2\}$  is identified if and only if  $A$  has at least  $D + L + 1$  distinct eigenvalues, or equivalently, the set of matrices  $\{I_N, A, \dots, A^{D+L}\}$  is linearly independent.*

## 1.8 Appendix: Linear-in-means models

This appendix extends the identification method in section 1.4.2 to more general forms of linear-in-means models, in which groups are divided into different levels of subgroups and individual outcomes depend on different subgroup level averages. A simple example will first illustrate the idea; then a more general model is in the subsequent discussions.

This example is a slight extension of the simple linear-in-means model in section 1.4.2. Suppose each group consists of two subgroups, and each subgroup consists of  $n(g, t)$  individuals. In group  $g$  of type  $t$ , the individual outcome of the individual  $j$  in subgroup  $i$  is denoted as  $\omega_{i,j}(g, t)$ , which depends on the group average and subgroup  $i$ 's average. The model is defined by

$$\omega(g, t) = J_0 S_0 \omega(g, t) + J_1 S_1 \omega(g, t) + z(g, t), \quad (1.94)$$

where  $z(g, t)$  is equal to the group unobservable  $\alpha(g, t)$  plus the individual unobservable  $\epsilon(g, t)$ :

$$z(g, t) = \alpha(g, t) \mathbf{1}_{N(g,t) \times 1} + \epsilon(g, t), \quad (1.95)$$

and the  $2n(g, t) \times 2n(g, t)$  matrices  $S_0$  and  $S_1$  are defined by

$$S_0 = \frac{1}{2n(g, t)} \begin{pmatrix} \mathbf{1}_{n(g,t) \times n(g,t)} & \mathbf{1}_{n(g,t) \times n(g,t)} \\ \mathbf{1}_{n(g,t) \times n(g,t)} & \mathbf{1}_{n(g,t) \times n(g,t)} \end{pmatrix}, \quad (1.96)$$

$$S_1 = \frac{1}{n(g,t)} \begin{pmatrix} \mathbf{1}_{n(g,t) \times n(g,t)} & \mathbf{0}_{n(g,t) \times n(g,t)} \\ \mathbf{0}_{n(g,t) \times n(g,t)} & \mathbf{1}_{n(g,t) \times n(g,t)} \end{pmatrix} - \frac{1}{2n(g,t)} \begin{pmatrix} \mathbf{1}_{n(g,t) \times n(g,t)} & \mathbf{1}_{n(g,t) \times n(g,t)} \\ \mathbf{1}_{n(g,t) \times n(g,t)} & \mathbf{1}_{n(g,t) \times n(g,t)} \end{pmatrix}. \quad (1.97)$$

$S_0\omega(g,t)$  is a column of group average and  $S_1\omega(g,t)$  is a column of subgroup average's deviations from group average.  $\mathbf{1}_{m \times n}$  and  $\mathbf{0}_{m \times n}$  are  $m \times n$  matrices of 1's and 0's.

The covariance matrix of the total unobservables  $z(g,t)$  consists of three terms: the inter-subgroup covariance  $\gamma_0(g,t)B_0$ , the intra-subgroup covariance  $\gamma_1(g,t)B_1$ , and the variance  $\gamma_2(g,t)B_2$ :

$$\text{Var}(z(g,t)) = \gamma_0(g,t)B_0 + \gamma_1(g,t)B_1 + \gamma_2(g,t)B_2, \quad (1.98)$$

$$B_0 = \begin{pmatrix} \mathbf{0}_{n(g,t) \times n(g,t)} & \mathbf{1}_{n(g,t) \times n(g,t)} \\ \mathbf{1}_{n(g,t) \times n(g,t)} & \mathbf{0}_{n(g,t) \times n(g,t)} \end{pmatrix}, \quad (1.99)$$

$$B_1 = \begin{pmatrix} \mathbf{1}_{n(g,t) \times n(g,t)} - I_{n(g,t)} & \mathbf{0}_{n(g,t) \times n(g,t)} \\ \mathbf{0}_{n(g,t) \times n(g,t)} & \mathbf{1}_{n(g,t) \times n(g,t)} - I_{n(g,t)} \end{pmatrix}, \quad (1.100)$$

$$B_2 = \begin{pmatrix} I_{n(g,t)} & \mathbf{0}_{n(g,t) \times n(g,t)} \\ \mathbf{0}_{n(g,t) \times n(g,t)} & I_{n(g,t)} \end{pmatrix}. \quad (1.101)$$

$I_n$  is the  $n \times n$  identity matrix.

(1.98) essentially assumes that the variances,  $\gamma_2(g,t) = \text{Var}(z_{i,j}(g,t))$ , the intra-subgroup covariances,  $\gamma_1(g,t) = \text{Cov}(z_{i,j}(g,t), z_{i,k}(g,t))$ , and the inter-subgroup covariances,  $\gamma_0(g,t) = \text{Cov}(z_{i,j}(g,t), z_{k,l}(g,t))$  where  $i \neq k$ , are constant for a given  $(g,t)$ . These variance and covariances are

$$\gamma_0(g,t) = \sigma_\alpha^2(g,t) + 2\sigma_{\alpha\epsilon}(g,t) + \sigma_{\epsilon\epsilon,0}(g,t), \quad (1.102)$$

$$\gamma_1(g,t) = \sigma_\alpha^2(g,t) + 2\sigma_{\alpha\epsilon}(g,t) + \sigma_{\epsilon\epsilon,1}(g,t), \quad (1.103)$$

and

$$\gamma_2(g,t) = \sigma_\alpha^2(g,t) + 2\sigma_{\alpha\epsilon}(g,t) + \sigma_\epsilon^2(g,t), \quad (1.104)$$

where  $\sigma_{\epsilon\epsilon,0}(g,t) = \text{Cov}(\epsilon_{i,j}(g,t), \epsilon_{k,l}(g,t))$  where  $i \neq k$  is the inter-subgroup covariance of  $\epsilon_{i,j}(g,t)$ ,  $\sigma_{\epsilon\epsilon,1}(g,t) = \text{Cov}(\epsilon_{i,j}(g,t), \epsilon_{i,k}(g,t))$  is the intra-subgroup covariance of  $\epsilon_{i,j}(g,t)$ , and  $\sigma_\epsilon^2(g,t) = \text{Var}(\epsilon_{i,j}(g,t))$  is the variance of  $\epsilon_{i,j}(g,t)$ .

A spectral decomposition of  $\text{Var}(z(g,t))$  is

$$\text{Var}(z(g,t)) = \xi_0(g,t)S_0 + \xi_1(g,t)S_1 + \xi_2(g,t)S_2 \quad (1.105)$$

where  $S_0$  and  $S_1$  are defined in (1.96) and (1.97),

$$S_2 = \begin{pmatrix} I_{n(g,t)} & \mathbf{0}_{n(g,t) \times n(g,t)} \\ \mathbf{0}_{n(g,t) \times n(g,t)} & I_{n(g,t)} \end{pmatrix} - \frac{1}{n(g,t)} \begin{pmatrix} \mathbf{1}_{n(g,t) \times n(g,t)} & \mathbf{0}_{n(g,t) \times n(g,t)} \\ \mathbf{0}_{n(g,t) \times n(g,t)} & \mathbf{1}_{n(g,t) \times n(g,t)} \end{pmatrix}, \quad (1.106)$$

and

$$\begin{pmatrix} \xi_0(g, t) \\ \xi_1(g, t) \\ \xi_2(g, t) \end{pmatrix} = \begin{pmatrix} 1 & n(g, t) - 1 & n(g, t) \\ 1 & n(g, t) - 1 & -n(g, t) \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \gamma_2(g, t) \\ \gamma_1(g, t) \\ \gamma_0(g, t) \end{pmatrix}. \quad (1.107)$$

$\{B_0, B_1, B_2\}$  is a set of commuting matrices, and  $\{S_0, S_1, S_2\}$  is a set of idempotent and pairwise orthogonal matrices.<sup>49</sup> (1.105) implies that  $Var(S_r z(g, t)) = S_r Var(z(g, t)) S_r = \xi_r S_r$ . That is,  $\xi_r(g, t) S_r$  is the covariance matrix of  $S_r z(g, t)$ .

Assume the self-consistency assumption that  $|J_0| < 1$  and  $|J_1| < 1$ . A spectral decomposition of  $Var(\omega(g, t))$  is

$$Var(\omega(g, t)) = \Delta_0(g, t) S_0 + \Delta_1(g, t) S_1 + \Delta_2(g, t) S_2, \quad (1.108)$$

where  $\Delta_r$ 's are the distinct eigenvalues of  $Var(\omega)$ :

$$\Delta_0(g, t) \equiv \frac{1}{(1 - J_0)^2} \xi_0(g, t), \quad (1.109)$$

$$\Delta_1(g, t) \equiv \frac{1}{(1 - J_1)^2} \xi_1(g, t), \quad (1.110)$$

$$\Delta_2(g, t) \equiv \xi_2(g, t). \quad (1.111)$$

For each class  $(g, t)$ , we have three equations for each group to identify parameters:

$$\Delta_0(g, t) = \frac{1}{(1 - J_0)^2} [2n(g, t) \sigma_\alpha^2(g, t) + 4n(g, t) \sigma_{\alpha\epsilon}(g, t) + \sigma_\epsilon^2(g, t) + (n(g, t) - 1) \sigma_{\epsilon\epsilon,1}(g, t) + n(g, t) \sigma_{\epsilon\epsilon,0}(g, t)], \quad (1.112)$$

$$\Delta_1(g, t) = \frac{1}{(1 - J_1)^2} [\sigma_\epsilon^2(g, t) + (n(g, t) - 1) \sigma_{\epsilon\epsilon,1}(g, t) - n(g, t) \sigma_{\epsilon\epsilon,0}(g, t)], \quad (1.113)$$

$$\Delta_2(g, t) = \sigma_\epsilon^2(g, t) - \sigma_{\epsilon\epsilon,1}(g, t). \quad (1.114)$$

But, for each  $(g, t)$ , we have seven parameters to be identified:  $J_0, J_1, \sigma_\alpha^2(g, t), \sigma_{\alpha\epsilon}(g, t), \sigma_{\epsilon\epsilon,0}(g, t), \sigma_{\epsilon\epsilon,1}(g, t)$ , and  $\sigma_\epsilon^2(g, t)$ . Note that only  $\Delta_0$  and  $\Delta_1$  vary with subgroup size  $n(g, t)$ .

Recall that there are three class types in the Project STAR: small (with 13 to 17 students), regular (with 22 to 25 students), and regular with a full time teacher's aide (with 22 to 25 students). Suppose  $\sigma_\alpha^2(g, t), \sigma_{\alpha\epsilon}(g, t), \sigma_\epsilon^2(g, t), \sigma_{\epsilon\epsilon,1}(g, t)$ , and  $\sigma_{\epsilon\epsilon,0}(g, t)$  are constant for all classes  $g$  within a given class type, but  $\sigma$ 's can vary with class type. Then there are seven parameters to be identified for each class type. For each  $g$ , there are three equations but only two of

<sup>49</sup>The sets of matrices,  $\{B_0, B_1, B_2\}$  and  $\{S_0, S_1, S_2\}$ , are the sets of matrices in p.888 of Speed (1987).  $B_0, B_1$ , and  $B_2$  are commuting, symmetric, and summing to the  $2n \times 2n$  matrix of 1's ( $B_0 + B_1 + B_2 = I_{2n \times 2n}$ );  $S_0, S_1$ , and  $S_2$  are idempotent ( $S_r S_r = S_r$  for all  $r$ ), pairwise orthogonal ( $S_i S_j = 0_{2n \times 2n}$  for  $i \neq j$ ), symmetric, and summing to the  $2n \times 2n$  identity matrix ( $S_0 + S_1 + S_2 = I_{2n}$ ).

them vary with class size. Thus, for a given class type, we need three classes of different sizes so that we have seven equations.

Suppose  $g_1$ ,  $g_2$ , and  $g_3$  are classes in a given class type, and these classes have different subgroup sizes:  $n_1$ ,  $n_2$ , and  $n_3$ . Since these classes are in the same type, the unobservable variables' variances and covariances of these classes are constant:  $\sigma_\alpha^2(g_i, t) = \sigma_\alpha^2(t)$ ,  $\sigma_{\alpha\epsilon}(g_i, t) = \sigma_{\alpha\epsilon}(t)$ ,  $\sigma_\epsilon^2(g_i, t) = \sigma_\epsilon^2(t)$ ,  $\sigma_{\epsilon\epsilon,1}(g_i, t) = \sigma_{\epsilon\epsilon,1}(t)$ , and  $\sigma_{\epsilon\epsilon,0}(g_i, t) = \sigma_{\epsilon\epsilon,0}(t)$ . We have the following seven equations for identification:

$$\Delta_0(g_i, t) = \frac{1}{(1 - J_0)^2} [2n_i(t)\sigma_\alpha^2(t) + 4n_i(t)\sigma_{\alpha\epsilon}(t) + \sigma_\epsilon^2(t) + (n_i(t) - 1)\sigma_{\epsilon\epsilon,1}(t) + n_i(t)\sigma_{\epsilon\epsilon,0}(t)], \quad (1.115)$$

$$\Delta_1(g_i, t) = \frac{1}{(1 - J_1)^2} [\sigma_\epsilon^2(t) + (n_i(t) - 1)\sigma_{\epsilon\epsilon,1}(t) - n_i(t)\sigma_{\epsilon\epsilon,0}(t)], \quad (1.116)$$

for  $i = 1, 2, 3$ , and

$$\Delta_2(g_1, t) = \Delta_2(g_2, t) = \Delta_2(g_3, t) = \sigma_\epsilon^2(t) - \sigma_{\epsilon\epsilon,1}(t). \quad (1.117)$$

The parameters:  $J_0$ ,  $J_1$ ,  $\sigma_\alpha^2(t)$ ,  $\sigma_{\alpha\epsilon}(t)$ ,  $\sigma_{\epsilon\epsilon,0}(t)$ ,  $\sigma_{\epsilon\epsilon,1}(t)$ , and  $\sigma_\epsilon^2(t)$  for this class type  $t$  are identified by these equations. For the remaining class types, since  $J_0$  and  $J_1$  are identified already, we just need two classes of different sizes to identify  $\sigma_\alpha^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_\epsilon^2(g, t)$ ,  $\sigma_{\epsilon\epsilon,1}(g, t)$ , and  $\sigma_{\epsilon\epsilon,0}(g, t)$  for  $g$ 's within the remaining class type.

**Theorem 17** *Consider the model defined by (1.94)-(1.101). Suppose  $|J_0| < 1$  and  $|J_1| < 1$  (self-consistency);  $\sigma_\alpha^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_\epsilon^2(g, t)$ ,  $\sigma_{\epsilon\epsilon,1}(g, t)$ , and  $\sigma_{\epsilon\epsilon,0}(g, t)$  vary with  $t$  (class type) but do not vary with  $g$  (class) for a given  $t$ . If (i.) there are three classes of different sizes for one class type, and (ii.) there are two classes of different sizes for the remaining class types, then  $J_0$ ,  $J_1$ ,  $\sigma_\alpha^2(g, t)$ ,  $\sigma_{\alpha\epsilon}(g, t)$ ,  $\sigma_\epsilon^2(g, t)$ ,  $\sigma_{\epsilon\epsilon,1}(g, t)$ , and  $\sigma_{\epsilon\epsilon,0}(g, t)$  for all class types are identified.*

In general, groups can be divided  $R$  times. The subscript  $i$  of individual outcome  $\omega_i(g, t)$  is defined as a nested index  $i = i_1, i_2, \dots, i_R$  where  $1 \leq i_k \leq n_k(g, t)$ . And  $n_k(g, t)$  is the number of  $k$ -level subgroups in a  $(k - 1)$ -level subgroup. Denote the group size as  $N(g, t) \equiv n_1(g, t)n_2(g, t) \cdots n_R(g, t)$ .

Define two sets of matrices before stating the generalized linear-in-means model. Use  $\otimes$  to denote the Kronecker product operator. First,  $\{B_0, B_1, \dots, B_R\}$  is a set of symmetric and commuting matrices which are defined by

$$B_0 = (1_{n_1 \times n_1} - I_{n_1}) \otimes 1_{n_2 \times n_2} \otimes \cdots \otimes 1_{n_R \times n_R}, \quad (1.118)$$

$$B_1 = I_{n_1 \times n_1} \otimes (1_{n_2 \times n_2} - I_{n_2}) \otimes 1_{n_3 \times n_3} \otimes \cdots \otimes 1_{n_R \times n_R}, \quad (1.119)$$

$$B_r = I_{n_1 \times n_1} \otimes \cdots \otimes I_{n_r \times n_r} \otimes (1_{n_{r+1} \times n_{r+1}} - I_{n_{r+1}}) \otimes 1_{n_{r+1} \times n_{r+1}} \otimes \cdots \otimes 1_{n_R \times n_R} \quad (1.120)$$

and

$$B_R = I_{n_1} \otimes I_{n_2} \otimes \cdots \otimes I_{n_R}. \quad (1.121)$$

The  $(i, j)$ -entry of  $B_r$  is 1 if  $i_h = j_h$  for  $h = 1, 2, \dots, r$  and  $i_{r+1} \neq j_{r+1}$ ; the  $(i, j)$ -entry of  $B_r$  is 0 otherwise. Second,  $\{S_1, S_2, \dots, S_R\}$  is a set of idempotent, pairwise orthogonal, and symmetric matrices which are defined by

$$S_0 = \frac{1}{n_1 \cdots n_R} R_0, \quad (1.122)$$

$$S_r = \frac{1}{n_{r+1} \cdots n_R} R_r - \frac{1}{n_r \cdots n_R} R_{r-1}, \quad (1.123)$$

$$S_R = I_N - \frac{1}{n_R} R_{R-1}, \quad (1.124)$$

where

$$R_r = B_r + B_{r+1} + \cdots + B_R. \quad (1.125)$$

Note that  $\sum_{r=0}^R B_r = 1_{N \times N}$  and  $\sum_{r=0}^R S_r = I_N$ . These are the sets of matrices defined in the example in p.894 of Speed (1987).

The generalized linear-in-means model is

$$\omega(g, t) = \sum_{r=0}^{R-1} J_r S_r \omega(g, t) + z(g, t), \quad (1.126)$$

$$z(g, t) = \alpha(g, t) 1_{N(g,t) \times 1} + \epsilon(g, t), \quad (1.127)$$

$$\text{Var}(z(g, t)) = \sum_{r=0}^R \gamma_r(g, t) B_r. \quad (1.128)$$

$\gamma_0(g, t)$  is the covariance between  $z_i(g, t)$  and  $z_j(g, t)$  where  $i$  and  $j$  have different first index.  $\gamma_r(g, t)$  is the covariance between  $z_i(g, t)$  and  $z_j(g, t)$  where  $i_h = j_h$  for  $h = 1, 2, \dots, r$  and  $i_{r+1} \neq j_{r+1}$ .  $\gamma_R(g, t)$  is the variance of  $z_i(g, t)$ . And these  $\gamma_r(g, t)$ 's are equal to

$$\gamma_r(g, t) = \sigma_\alpha^2(g, t) + 2\sigma_{\alpha\epsilon}(g, t) + \sigma_{\epsilon\epsilon, r}(g, t), \quad (1.129)$$

$$\gamma_R(g, t) = \sigma_\alpha^2(g, t) + 2\sigma_{\alpha\epsilon}(g, t) + \sigma_\epsilon^2(g, t), \quad (1.130)$$

where  $\sigma_{\epsilon\epsilon, r}$  is the covariance between  $\epsilon_i(g, t)$  and  $\epsilon_j(g, t)$  where  $i_h = j_h$  for  $h = 1, 2, \dots, r$  and  $i_{r+1} \neq j_{r+1}$ , and  $\sigma_\epsilon^2(g, t)$  is the variance of  $z_i(g, t)$ .

A spectral decomposition of  $\text{Var}(z(g, t))$  is

$$\text{Var}(z(g, t)) = \sum_{r=0}^R \xi_r(g, t) S_r, \quad (1.131)$$

where

$$\xi_R(g, t) = \gamma_R(g, t) - \gamma_{R-1}(g, t), \quad (1.132)$$

$$\xi_{R-1}(g, t) = \gamma_R(g, t) + (n_R(g, t) - 1)\gamma_{R-1}(g, t) - n_R(g, t)\gamma_{R-2}(g, t), \quad (1.133)$$

$$\begin{aligned} \xi_{i+1}(g, t) &= \gamma_R(g, t) + (n_R(g, t) - 1)\gamma_{R-1}(g, t) \\ &\quad + (n_{R-1}(g, t) - 1)n_R(g, t)\gamma_{R-2}(g, t) + \cdots \\ &\quad + [(n_{i+2}(g, t) - 1)n_{i+3}(g, t) \cdots n_R(g, t)]\gamma_{i+1}(g, t) \\ &\quad - [n_{i+2}(g, t)n_{i+3}(g, t) \cdots n_R(g, t)]\gamma_i(g, t), \end{aligned} \quad (1.134)$$

$$\begin{aligned} \xi_i(g, t) &= \gamma_R(g, t) + (n_R(g, t) - 1)\gamma_{R-1}(g, t) \\ &\quad + (n_{R-1}(g, t) - 1)n_R(g, t)\gamma_{R-2}(g, t) + \cdots \\ &\quad + [(n_{i+1}(g, t) - 1)n_{i+2}(g, t) \cdots n_R(g, t)]\gamma_i(g, t) \\ &\quad - [n_{i+1}(g, t)n_{i+2}(g, t) \cdots n_R(g, t)]\gamma_{i-1}(g, t), \end{aligned} \quad (1.135)$$

$$\begin{aligned} \xi_{i-1}(g, t) &= \gamma_R(g, t) + (n_R(g, t) - 1)\gamma_{R-1}(g, t) \\ &\quad + (n_{R-1}(g, t) - 1)n_R(g, t)\gamma_{R-2}(g, t) + \cdots \\ &\quad + [(n_i(g, t) - 1)n_{i+1}(g, t) \cdots n_R(g, t)]\gamma_{i-1}(g, t) \\ &\quad - [n_i(g, t)n_{i+1}(g, t) \cdots n_R(g, t)]\gamma_{i-2}(g, t), \end{aligned} \quad (1.136)$$

and

$$\begin{aligned} \xi_0(g, t) &= \gamma_R(g, t) + (n_R(g, t) - 1)\gamma_{R-1}(g, t) \\ &\quad + (n_{R-1}(g, t) - 1)n_R(g, t)\gamma_{R-2}(g, t) + \cdots \\ &\quad + [(n_2(g, t) - 1)n_3(g, t) \cdots n_R(g, t)]\gamma_1(g, t) \\ &\quad + [(n_1(g, t) - 1)n_2(g, t) \cdots n_R(g, t)]\gamma_0(g, t), \end{aligned} \quad (1.137)$$

Note that the eigenvalues are of the forms:

$$\xi_{R-1}(g, t) = \xi_R(g, t) + n_R(g, t)[\gamma_{R-1}(g, t) - \gamma_{R-2}(g, t)], \quad (1.138)$$

$$\xi_{i+1}(g, t) = \xi_{i+2}(g, t) + [n_{i+2}(g, t)n_{i+3}(g, t) \cdots n_R(g, t)][\gamma_{i+1}(g, t) - \gamma_i(g, t)], \quad (1.139)$$

$$\xi_i(g, t) = \xi_{i+1}(g, t) + [n_{i+1}(g, t)n_{i+2}(g, t) \cdots n_R(g, t)][\gamma_i(g, t) - \gamma_{i-1}(g, t)], \quad (1.140)$$

$$\xi_{i-1}(g, t) = \xi_i(g, t) + [n_i(g, t)n_{i+1}(g, t) \cdots n_R(g, t)][\gamma_{i-1}(g, t) - \gamma_{i-2}(g, t)], \quad (1.141)$$

and

$$\xi_0(g, t) = \xi_1(g, t) + [n_1(g, t)n_2(g, t) \cdots n_R(g, t)]\gamma_0(g, t). \quad (1.142)$$

Substitute (1.129) and (1.130) into (1.132) and (1.138)-(1.142).

$$\xi_R(g, t) = \sigma_\epsilon^2(g, t) - \sigma_{\epsilon\epsilon, R-1}(g, t), \quad (1.143)$$

$$\xi_{R-1}(g, t) = \sigma_\epsilon^2(g, t) - \sigma_{\epsilon\epsilon, R-1}(g, t) + n_R(g, t)[\sigma_{\epsilon\epsilon, R-1}(g, t) - \sigma_{\epsilon\epsilon, R-2}(g, t)], \quad (1.144)$$

$$\begin{aligned}\xi_r(g, t) &= \sigma_\epsilon^2(g, t) - \sigma_{\epsilon\epsilon, R-1}(g, t) \\ &\quad + \sum_{i=r}^{R-1} [n_{i+1}(g, t) \cdots n_R(g, t)] [\sigma_{\epsilon\epsilon, i}(g, t) - \sigma_{\epsilon\epsilon, i-1}(g, t)],\end{aligned}\quad (1.145)$$

$$\begin{aligned}\xi_0(g, t) &= [\sigma_\epsilon^2(g, t) - \sigma_{\epsilon\epsilon, R-1}(g, t)] \\ &\quad + \sum_{i=1}^{R-1} [n_{i+1}(g, t) \cdots n_R(g, t)] [\sigma_{\epsilon\epsilon, i}(g, t) - \sigma_{\epsilon\epsilon, i-1}(g, t)] \\ &\quad + [n_1(g, t) \cdots n_R(g, t)] [\sigma_\alpha^2(g, t) + 2\sigma_{\alpha\epsilon}(g, t) + \sigma_{\epsilon\epsilon, 0}(g, t)]\end{aligned}\quad (1.146)$$

Assume  $|J_r| < 1$  for all  $r$ . A spectral decomposition of  $Var(\omega(g, t))$  is

$$Var(\omega(g, t)) = \sum_{r=0}^{R-1} \Delta_r S_r + \Delta_R S_R \quad (1.147)$$

where

$$\Delta_r(g, t) \equiv \frac{\xi_r}{(1 - J_r)^2} \text{ and } \Delta_R(g, t) \equiv \xi_R(g, t). \quad (1.148)$$

These eigenvalues of  $Var(\omega(g, t))$  are identified. We have  $R + 1$  equations for each  $(g, t)$ . But we have  $R + (R + 1) + 2$  parameters for each  $(g, t)$ :  $J_0, \dots, J_{R-1}, \sigma_{\epsilon\epsilon, 0}(g, t), \dots, \sigma_{\epsilon\epsilon, R-1}(g, t), \sigma_\epsilon^2(g, t), \sigma_\alpha^2(g, t)$ , and  $\sigma_{\alpha\epsilon}(g, t)$ .

Suppose all  $\sigma(g, t)$ 's vary with class type but does not vary with  $g$  for  $g$ 's within a class type. Then, for a given class type, we have  $R + (R + 1) + 2$  parameters to identify. Note that, from (1.143)-(1.146),  $\Delta_r(g, t)$ 's depend on subgroup sizes but  $\Delta_R(g, t)$  does not depend on subgroup size. Thus, for a given class type, if there are three classes with different class sizes, then we have  $3(R + 1) - 2$  equations for this class type to identify all parameters for this class type. Hence, since  $J_0, \dots, J_{R-1}$  are identified already, for the remaining class types, we need only two classes of different sizes.

**Theorem 18** *Consider the model defined by (1.126)-(1.128). Suppose  $|J_r| < 1$  for all  $r$  (self-consistency);  $\sigma_{\epsilon\epsilon, 0}(g, t), \dots, \sigma_{\epsilon\epsilon, R-1}(g, t), \sigma_\epsilon^2(g, t), \sigma_\alpha^2(g, t)$ , and  $\sigma_{\alpha\epsilon}(g, t)$  vary with  $t$  (class type) but do not vary with  $g$  (class) for a given  $t$ . If*

- (i.) *there are three classes of different sizes for one class type, and*
- (ii.) *there are two classes of different sizes for the remaining class types,*

*then  $J_0, \dots, J_{R-1}, \sigma_{\epsilon\epsilon, 0}(g, t), \dots, \sigma_{\epsilon\epsilon, R-1}(g, t), \sigma_\epsilon^2(g, t), \sigma_\alpha^2(g, t)$ , and  $\sigma_{\alpha\epsilon}(g, t)$  for all class types are identified.*

## Chapter 2

# Estimating Covariance Matrices with Correlated Disturbances

**JEL classification:** C13

**Keywords:** covariance matrix, network

### 2.1 Introduction

This chapter derives a sufficient condition so that the covariance matrices can be consistently estimated. Consider a linear regression model where individuals interact in a social network so that the regression disturbances are correlated. The disturbances of two observations are correlated if the two individuals are connected in the network. The model is

$$y_t = X_t \beta_0 + \epsilon_t \quad (2.1)$$

for  $t = 1, 2, \dots, n$ .  $n$  is the number of observations.  $y_t$  is the dependent variable.  $X_t = (X_{t1}, X_{t2}, \dots, X_{tk})$  is a  $1 \times k$  row of independent variables.  $\beta_0$  is a  $k \times 1$  column of parameters to be estimated.  $\epsilon_t$  is a scale of disturbance. The subscript  $t$  denote the  $t^{\text{th}}$  observation.  $\epsilon_t$  and  $\epsilon_{t'}$  are correlated if individuals  $t$  and  $t'$  are connected in the network. The model can be written in a matrix form

$$y = X \beta_0 + \epsilon. \quad (2.2)$$

$y = (y_1, y_2, \dots, y_n)'$  is a  $n \times 1$  column.  $X$  is a  $n \times k$  matrix where the  $t^{\text{th}}$  row is  $X_t$ . The  $n \times 1$  column  $\epsilon$  is defined similarly.

The independent variables and the disturbance are potentially correlated,  $E(X_t' \epsilon_t) \neq 0$ . Suppose  $Z_t = (Z_{t1}, Z_{t2}, \dots, Z_{tl})$  is a  $1 \times l$  row of instruments so

that  $E(Z_t'\epsilon_t) = 0$ .  $Z$  is a  $n \times l$  matrix where the  $t^{\text{th}}$  row is  $Z_t$ . Estimate  $\beta_0$  by choosing a  $k \times 1$  column  $\beta$  that minimizes

$$d_n(\beta) \equiv (y - X\beta)'Z\hat{P}_nZ'(y - X\beta), \quad (2.3)$$

where  $\hat{P}_n$  is a symmetric  $l \times l$  positive definite norming matrix. The minimizing  $\beta$  is

$$\tilde{\beta}_n = (X'Z\hat{P}_nZ'X)^{-1}X'Z\hat{P}_nZ'y. \quad (2.4)$$

Define  $V_n \equiv \text{var}(n^{-1/2}Z'\epsilon)$ . The chapter IV of White (1984) shows that the optimal choice of  $P_n$  is  $V_n^{-1}$ . The main purpose of this chapter is to derive a sufficient condition under which  $V_n$  can be consistently estimated.

## 2.2 Estimating Covariance Matrices

Suppose the  $n$  observations are arranged so that the last  $n - N(n)$  disturbances are not correlated. That is, only the first  $N(n)$  disturbances are potentially correlated.  $N(n)$  is written explicitly as a function of  $n$  because  $N$  is growing with  $n$ . We will see that  $V_n$  can be consistently estimated if  $N$  grows slow enough.

Since the instruments are not correlated with the disturbances,  $E(Z_t'\epsilon_t) = 0$ ,

$$\begin{aligned} V_n &= \frac{1}{n}E(Z'\epsilon\epsilon'Z) \\ &= \frac{1}{n}\sum_{t=1}^n E(Z_t'\epsilon_t\epsilon_tZ_t) + \frac{1}{n}\sum_{i=1}^{N(n)-1}\sum_{t=i+1}^{N(n)} E(Z_i'\epsilon_i\epsilon_tZ_t + Z_t'\epsilon_t\epsilon_iZ_i). \end{aligned} \quad (2.5)$$

Consider an estimator of  $V_n$ :

$$\hat{V}_n = \frac{1}{n}\sum_{t=1}^n Z_t'\tilde{\epsilon}_t\tilde{\epsilon}_tZ_t + \frac{1}{n}\sum_{i=1}^{N(n)-1}\sum_{t=i+1}^{N(n)} (Z_i'\tilde{\epsilon}_i\tilde{\epsilon}_tZ_t + Z_t'\tilde{\epsilon}_t\tilde{\epsilon}_iZ_i), \quad (2.6)$$

where

$$\tilde{\epsilon}_t = y_t - X_t\tilde{\beta}_n = \epsilon_t - X_t(\tilde{\beta}_n - \beta_0). \quad (2.7)$$

$\tilde{\beta}_n$  is a consistent estimator of  $\beta_0$ . That is,  $\tilde{\beta}_n \rightarrow \beta_0$  in probability as  $n \rightarrow \infty$ .

This section derives the order of  $N(n)$  so that  $\hat{V}_n \rightarrow V_n$  in probability as  $n \rightarrow \infty$ . Split the difference between  $V_n$  and  $\hat{V}_n$  into three terms:

$$V_n - \hat{V}_n = v_{1n} + v_{2n} + v_{3n}, \quad (2.8)$$

where

$$\begin{aligned} v_{1n} &\equiv \frac{1}{n}\sum_{t=1}^n [Z_t'\epsilon_t\epsilon_tZ_t - E(Z_t'\epsilon_t\epsilon_tZ_t)] \\ &\quad + \frac{1}{n}\sum_{t=1}^n [-Z_t'X_t(\tilde{\beta}_n - \beta_0)\epsilon_tZ_t - Z_t'\epsilon_t(\tilde{\beta}_n - \beta_0)'X_t'Z_t] \\ &\quad + \frac{1}{n}\sum_{t=1}^n [Z_t'X_t(\tilde{\beta}_n - \beta_0)(\tilde{\beta}_n - \beta_0)'X_t'Z_t], \end{aligned} \quad (2.9)$$

$$\begin{aligned}
v_{2n} &\equiv \frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z'_i \epsilon_i \epsilon_t Z_t - E(Z'_i \epsilon_i \epsilon_t Z_t)] \\
&+ \frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [-Z'_i X_i (\tilde{\beta}_n - \beta_0) \epsilon_t Z_t - Z'_i \epsilon_i (\tilde{\beta}_n - \beta_0)' X'_t Z_t] \\
&+ \frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z'_i X_i (\tilde{\beta}_n - \beta_0) (\tilde{\beta}_n - \beta_0)' X'_t Z_t], \tag{2.10}
\end{aligned}$$

and

$$\begin{aligned}
v_{3n} &\equiv \frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z'_t \epsilon_t \epsilon_i Z_i - E(Z'_t \epsilon_t \epsilon_i Z_i)] \\
&+ \frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [-Z'_t X_t (\tilde{\beta}_n - \beta_0) \epsilon_i Z_i - Z'_t \epsilon_t (\tilde{\beta}_n - \beta_0)' X'_i Z_i] \\
&+ \frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z'_t X_t (\tilde{\beta}_n - \beta_0) (\tilde{\beta}_n - \beta_0)' X'_i Z_i]. \tag{2.11}
\end{aligned}$$

The sufficient conditions for each of these three terms converging to zero in probability will be derived in the subsequent discussions.

**Variance Terms.**  $v_{1n}$  consists of the variance terms, the diagonal entries of  $V_n - \hat{V}_n$ . Consider the first term of  $v_{1n}$ :

$$\frac{1}{n} \sum_{t=1}^n [Z'_t \epsilon_t \epsilon_t Z_t - E(Z'_t \epsilon_t \epsilon_t Z_t)]. \tag{2.12}$$

Suppose  $\mathcal{G}$  and  $\mathcal{H}$  are  $\sigma$ -fields. Define

$$\phi(\mathcal{G}, \mathcal{H}) \equiv \sup_{G \in \mathcal{G}, H \in \mathcal{H}: P(G) > 0} |P(H|G) - P(H)|, \tag{2.13}$$

and

$$\alpha(\mathcal{G}, \mathcal{H}) \equiv \sup_{G \in \mathcal{G}, H \in \mathcal{H}} |P(G \cap H) - P(G)P(H)|. \tag{2.14}$$

Denote the smallest  $\sigma$ -field generated by the random variables from  $a^{th}$  to  $b^{th}$  observations as  $\mathcal{B}_a^b$ . Define

$$\phi(m) \equiv \sup_n \phi(\mathcal{B}_{-\infty}^n, \mathcal{B}_{n+m}^\infty) \tag{2.15}$$

and

$$\alpha(m) \equiv \sup_n \alpha(\mathcal{B}_{-\infty}^n, \mathcal{B}_{n+m}^\infty). \tag{2.16}$$

Let  $1 \leq r \leq \infty$ . If  $\phi(m) = O(m^{-\lambda})$  for  $\lambda > r/(2r-1)$ , then  $\phi(m)$  is of size  $r/(2r-1)$ . Let  $r > 1$ . If  $\alpha(m) = O(m^{-\lambda})$  for  $\lambda > r/(r-1)$ , then  $\alpha(m)$  is of size  $r/(r-1)$ .

Since only the first  $N(n)$  observations are potentially correlated,  $\phi(m) > 0$  and  $\alpha(m) > 0$  for  $m < N(n)$ , and  $\phi(m) = 0$  and  $\alpha(m) = 0$  for  $m \geq N(n)$ . Therefore,  $\phi(m)$  is of size  $r/(2r-1)$  for all any  $r \geq 1$  and  $\alpha(m)$  is of size  $r/(r-1)$  for all any  $r > 1$ . By the corollary 3.48 of White (1984),  $E|Z'_t \epsilon_t \epsilon_t Z_t|^{r+\delta} < \Delta < \infty$  for some  $\delta > 0$  and for all  $t$ ; then (2.12) converges to 0 almost surely.

Consider the second term of  $v_{1n}$ :

$$\frac{1}{n} \sum_{t=1}^n [-Z'_t X_t (\tilde{\beta}_n - \beta_0) \epsilon_t Z_t - Z'_t \epsilon_t (\tilde{\beta}_n - \beta_0)' X'_t Z_t] \quad (2.17)$$

In (2.17), the  $(i, j)$ -entry of the  $\lambda^{th}$  term is

$$\frac{1}{n} \sum_{t=1}^n [-Z_{ti} X_{\lambda t} (\tilde{\beta}_{n\lambda} - \beta_{0\lambda}) \epsilon_t Z_{tj}] \quad (2.18)$$

By an argument similar to that in previous paragraph, the corollary 3.48 of White (1984) implies that  $\frac{1}{n} \sum_{t=1}^n [-Z'_{ti} X_{\lambda t} \epsilon_t Z_{tj}] \rightarrow E(-Z'_{ti} X_{\lambda t} \epsilon_t Z_{tj})$  almost surely. Since  $\tilde{\beta}_n$  is a consistent estimator of  $\beta_0$ ,  $\tilde{\beta}_{n\lambda} \rightarrow \beta_{0\lambda}$  in probability. Therefore, (2.18) converges to 0 in probability and hence (2.17) converges to 0 in probability.

The third term of  $v_{1n}$  is

$$\frac{1}{n} \sum_{t=1}^n [Z'_t X_t (\tilde{\beta}_n - \beta_0) (\tilde{\beta}_n - \beta_0)' X'_t Z_t]. \quad (2.19)$$

This term converge to 0 in probability by an argument similar to that for the second term of (2.17).

**Covariance Terms**  $v_{2n}$  consists of the covariance terms, the off-diagonal entries of  $V_n - \hat{V}_n$ . Consider the first term of  $v_{2n}$ :

$$\frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z'_i \epsilon_i \epsilon_t Z_t - E(Z'_i \epsilon_i \epsilon_t Z_t)]. \quad (2.20)$$

Define  $\mathcal{Z}_{t,i} \equiv Z'_i \epsilon_i \epsilon_t Z_t - E(Z'_i \epsilon_i \epsilon_t Z_t)$ . By the correct version of the lemma 6.19 of White (1984), which is proved in the appendix,

$$E \left( \sum_{t=i+1}^{N(n)} \mathcal{Z}_{t,i} \right)^2 \leq (N(n) - i)(i + 2)\Delta^*, \quad (2.21)$$

where  $\Delta^* < \infty$ . The probability of the absolute value of (2.20) is large than a small  $\epsilon$  is

$$\begin{aligned}
P \left[ \left| \frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} \mathcal{Z}_{t,i} \right| \geq \epsilon \right] &\leq P \left[ \sum_{i=1}^{N(n)-1} \left| \sum_{t=i+1}^{N(n)} \mathcal{Z}_{t,i} \right| \geq n\epsilon \right] \\
&\leq \sum_{i=1}^{N(n)-1} P \left[ \left| \sum_{t=i+1}^{N(n)} \mathcal{Z}_{t,i} \right| \geq \frac{n}{N(n)-1} \epsilon \right] \\
&\leq \sum_{i=1}^{N(n)-1} \frac{E \left( \sum_{t=i+1}^{N(n)} \mathcal{Z}_{t,i} \right)^2}{\frac{n^2}{(N(n)-1)^2} \epsilon^2} \\
&\leq \sum_{i=1}^{N(n)-1} \frac{(N(n)-i)(i+2)\Delta^*}{\frac{n^2}{(N(n)-1)^2} \epsilon^2} \\
&< \sum_{i=1}^{N(n)-1} \frac{N(n)(N(n)-1)^2 \Delta^*}{n^2 \epsilon^2} (i+2) \\
&= \frac{N(n)(N(n)-1)^3 (N(n)+4) \Delta^*}{2n^2 \epsilon^2}. \quad (2.22)
\end{aligned}$$

Therefore, if  $N(n) = o(n^{2/5})$  the probability of in the left hand side of (2.22) converges to 0.

Consider the second term of  $v_{2n}$ :

$$\frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [-Z'_i X_i (\tilde{\beta}_n - \beta_0) \epsilon_t Z_t - Z'_i \epsilon_i (\tilde{\beta}_n - \beta_0)' X'_t Z_t] \quad (2.23)$$

The  $(a, b)$ -entry of the  $\lambda^{th}$  term is

$$\frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [-Z_{ia} X_{i\lambda} (\tilde{\beta}_{n\lambda} - \beta_{0\lambda}) \epsilon_t Z_{tb}] \quad (2.24)$$

We need to derive a sufficient condition for (2.24) to converge to 0 in probability.

First, an argument similar to that for the first term of  $v_{2n}$  shows that shows that

$$\frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z_{ia} X_{i\lambda} \epsilon_t Z_{tb} - E(Z_{ia} X_{i\lambda} \epsilon_t Z_{tb})] \rightarrow 0 \quad (2.25)$$

in probability if  $N(n) = o(n^{2/5})$ . Second, we know that  $n^{1/2}(\tilde{\beta}_{n\lambda} - \beta_{0\lambda}) = O_p(1)$ .

Third, we have

$$\left| \frac{1}{n^{3/2}} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} E(Z_{ia} X_{i\lambda} \epsilon_t Z_{tb}) \right| \leq \frac{1}{n^{3/2}} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} |E(Z_{ia} X_{i\lambda} \epsilon_t Z_{tb})|$$

$$\begin{aligned}
&\leq \frac{1}{n^{3/2}} \sum_{i=1}^{N(n)-1} (N(n) - i) \Delta \\
&= \frac{1}{n^{3/2}} \frac{N(n)(N(n) - 1)}{2} \Delta, \tag{2.26}
\end{aligned}$$

where  $|E(Z_{ia}X_{i\lambda}\epsilon_t Z_{tb})| \leq \Delta < \infty$ . If  $N(n) = o(n^{3/4})$ , then the left hand side of (2.26) converge to 0.

Consider the third term of  $v_{2n}$ :

$$\frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z'_i X_i (\tilde{\beta}_n - \beta_0) (\tilde{\beta}_n - \beta_0)' X'_t Z_t]. \tag{2.27}$$

The  $(a, b)$ -entry of the  $(\lambda, \kappa)^{th}$  term is

$$\frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z_{ia} X_{i\lambda} (\tilde{\beta}_{n\lambda} - \beta_{0\lambda}) (\tilde{\beta}_{n\kappa} - \beta_{0\kappa}) X_{t\kappa} Z_{tb}]. \tag{2.28}$$

We need to derive a sufficient condition for (2.28) to converge to 0 in probability.

First, an argument similar to that for the first term of  $v_{2n}$  shows that

$$\frac{1}{n} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} [Z_{ia} X_{i\lambda} X_{t\kappa} Z_{tb} - E(Z_{ia} X_{i\lambda} X_{t\kappa} Z_{tb})] \rightarrow 0 \tag{2.29}$$

in probability if  $N(n) = o(n^{2/5})$ . Second, we know that  $n^{1/2}(\tilde{\beta}_{n\lambda} - \beta_{0\lambda}) = O_p(1)$  and  $n^{1/2}(\tilde{\beta}_{n\kappa} - \beta_{0\kappa}) = O_p(1)$ . Third,

$$\begin{aligned}
\left| \frac{1}{n^2} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} E(Z_{ia} X_{i\lambda} X_{t\kappa} Z_{tb}) \right| &\leq \frac{1}{n^2} \sum_{i=1}^{N(n)-1} \sum_{t=i+1}^{N(n)} E|Z_{ia} X_{i\lambda} X_{t\kappa} Z_{tb}| \\
&\leq \frac{1}{n^2} \frac{N(n)(N(n) - 1)}{2} \Delta, \tag{2.30}
\end{aligned}$$

where  $E|Z_{ia} X_{i\lambda} X_{t\kappa} Z_{tb}| \leq \Delta < \infty$ . Therefore, if  $N(n) = o(n)$ , the left hand side (2.30) converge to 0.

The same argument for the second term of  $v_{2n}$  shows that  $v_{3n} \rightarrow 0$  in probability if  $N(n) = o(n)$ . As a result, the difference between  $V_n$  and  $\hat{V}_n$  converge to zero in probability if  $N(n) = o(n)$ .

**Theorem 19** *If  $N(n) = o(n)$ , then  $\hat{V}_n$  is a consistent estimator of  $V_n$ .*

## 2.3 Appendix: Corrected Lemma 6.19 of White (1984)

The corrected statement of the lemma 6.19 of White (1984) is

**Theorem 20** *There exists a constant  $\Delta^* < \infty$  such that*

$$E \left( \sum_{t=\tau+1}^n \mathcal{Z}_{t,\tau} \right)^2 \leq (n - \tau)(\tau + 2)\Delta^*. \quad (2.31)$$

**Proof.** Correct the summation in the proof the lemma 6.19 (White 1984, p.153):

$$\begin{aligned} E \left( \sum_{t=\tau+1}^n \mathcal{Z}_{t,\tau} \right)^2 &= \sum_{t=\tau+1}^n E(\mathcal{Z}_{t,\tau}^2) \\ &\quad + \sum_{m=1}^{n-\tau-1} \sum_{t=\tau+1+m}^n [E(\mathcal{Z}_{t,\tau}\mathcal{Z}_{t-m,\tau}) + E(\mathcal{Z}_{t-m,\tau}\mathcal{Z}_{t,\tau})] \\ &\leq \sum_{t=\tau+1}^n E|\mathcal{Z}_{t,\tau}^2| \\ &\quad + \sum_{m=1}^{n-\tau-1} \sum_{t=\tau+1+m}^n [E|\mathcal{Z}_{t,\tau}\mathcal{Z}_{t-m,\tau}| + E|\mathcal{Z}_{t-m,\tau}\mathcal{Z}_{t,\tau}|] \end{aligned} \quad (2.32)$$

The corollary 6.17 of the revised edition of White (2001) (or the corollary 6.16 of White (1984)) implies that

$$E|\mathcal{Z}_{t-m,\tau}\mathcal{Z}_{t,\tau}| \leq 2\phi_\tau(m)^{1-1/q} \text{var}(\mathcal{Z}_{t-m,\tau})^{1/2} \|\mathcal{Z}_{t,\tau}\|_q \quad (2.33)$$

or

$$E|\mathcal{Z}_{t-m,\tau}\mathcal{Z}_{t,\tau}| \leq 2(2^{1/2} + 1)\alpha_\tau(m)^{1/2-1/q} \text{var}(\mathcal{Z}_{t-m,\tau})^{1/2} \|\mathcal{Z}_{t,\tau}\|_q. \quad (2.34)$$

Then substituting (2.33) into (2.32) yields

$$\begin{aligned} &E \left( \sum_{t=\tau+1}^n \mathcal{Z}_{t,\tau} \right)^2 \\ &\leq \sum_{t=\tau+1}^n E|\mathcal{Z}_{t,\tau}^2| + 2 \sum_{m=1}^{n-\tau-1} \sum_{t=\tau+1+m}^n 2\phi_\tau(m)^{1-1/q} \text{var}(\mathcal{Z}_{t-m,\tau})^{1/2} \|\mathcal{Z}_{t,\tau}\|_q \\ &\leq (n - \tau)\Delta_2^2 + 4 \sum_{m=1}^{n-\tau-1} \sum_{t=\tau+1+m}^n \phi_\tau(m)^{1-1/q} \Delta_2 \Delta_q \\ &= (n - \tau)\Delta_2^2 + 4\Delta_2 \Delta_q \sum_{m=1}^{n-\tau-1} (n - \tau - m)\phi_\tau(m)^{1-1/q}, \end{aligned} \quad (2.35)$$

where  $\Delta_2^2 \geq E|\mathcal{Z}_{t,\tau}^2|$ ,  $\Delta_2 \geq \text{var}(\mathcal{Z}_{t-m,\tau})^{1/2}$  and  $\Delta_q \geq \|\mathcal{Z}_{t,\tau}\|_q$ . The lemma 6.18 of White (1984) states that  $\phi_\tau(m) \leq 1$  for  $m \leq \tau$  and  $\phi_\tau(m) \leq \phi(m - \tau)$  for

$m > \tau$ . Hence,

$$\begin{aligned}
E \left( \sum_{t=\tau+1}^n \mathcal{Z}_{t,\tau} \right)^2 &\leq (n-\tau)\Delta_2^2 + 4\Delta_2\Delta_q \sum_{m=1}^{\tau} (n-\tau-m) \\
&\quad + 4\Delta_2\Delta_q \sum_{m=\tau+1}^{n-\tau-1} (n-\tau-m)\phi(m-\tau)^{1-1/q} \\
&< (n-\tau)\Delta_2^2 + 4\Delta_2\Delta_q(n-\tau)\tau \\
&\quad + 4\Delta_2\Delta_q(n-\tau) \sum_{m=\tau+1}^{n-\tau-1} \phi(m-\tau)^{1-1/q} \\
&< (n-\tau)(\tau+2)\Delta^*, \tag{2.36}
\end{aligned}$$

where  $\Delta^* \geq \Delta_2^2$ ,  $\Delta^* \geq 4\Delta_2\Delta_q$ , and  $\Delta^* \geq 4\Delta_2\Delta_q \sum_{m=\tau+1}^{n-\tau-1} \phi(m-\tau)^{1-1/q}$ . This proves theorem 20. Note that the inequality (2.36) can be derived by using (2.34) also.

## Chapter 3

# Robust Ramsey Taxation under Implementation Lag Uncertainty

**JEL classification:** C61, E62, H21

**Keywords:** Fourier transform, implementation lag, Ramsey taxation, robust control

### 3.1 Introduction

This chapter develops a robust control theory under implementation lag uncertainty. The following simple control problem illustrates the idea. A decision maker maximizes a function of a vector of endogenous state variables  $x_t$  and a vector of control variables  $u_t$ . The vector  $x_t$  evolves according to the model

$$x_{t+1} = A(L)x_t + B(L)u_t + \epsilon_{t+1}, \quad (3.1)$$

where  $\epsilon_t$  is a vector of exogenous state variables,  $L$  is the lag operator,  $A(L)$  and  $B(L)$  are matrices of lag polynomials. We may think of consumption and labor supply as the endogenous state variables, tax rate as the control variable, technological level and government expenditure as the exogenous state variables. But the model (3.1) may be misspecified. There may be misspecification of the distribution of  $\{\epsilon_t\}$ , or the lag polynomials in  $A(L)$  and  $B(L)$ . Hansen and Sargent (2008) developed a robust control theory for the misspecification of the distribution of  $\{\epsilon_t\}$ . This chapter develops a robust control theory for the misspecification of  $B(L)$ . This kind of model uncertainty was proposed in the section 3.iii of Brock and Durlauf (2004).

In Hansen and Sargent's robust control theory, there is a set of distributions of  $\{\epsilon_t\}$ . A robust decision rule refers to the decision rule under the worst situa-

tion in this set of distributions. The robust decision maker assumes that there is an adversarial agent who chooses a distribution, within the given set of distributions, to minimize the robust decision maker's objective function. Precisely, for any decision rule chosen by the robust decision maker, the adversarial agent chooses a distribution to minimize the robust decision maker's objective function; the robust decision maker chooses a decision rule subject to the adversarial agent's malevolent decision rule.

This chapter develops a robust control theory under implementation lag uncertainty, which refers to the model uncertainty with respect to the lag polynomial matrix  $B(L)$  in the model (3.1). For any decision rule chosen by the robust decision maker, the adversarial agent chooses a lag polynomial, within a given set of lag polynomials, to minimize the robust decision maker's objective function.

The robust control theory under implementation lag uncertainty is developed through a Ramsey taxation problem without capital. The Ramsey planner chooses a sequence of labor income tax rates  $\{\tau_t\}$  to maximize welfare, but the effectively implemented sequence of tax rates is  $\{D(L)\tau_t\}$ , where  $D(L)$  is an estimate of the implementation lag polynomial. The estimate may not be the true implementation lag. The Ramsey planner seeks for a robust decision rule over a set of implementation lag polynomials that are close to  $D(L)$ . The closeness is in the  $l^2$  sense. The adversarial agent chooses an implementation lag  $\hat{D}(L)$  to minimize welfare subject to

$$\sum_{j=0}^{\infty} \beta^j (\hat{D}_j - D_j)^2 \leq \eta. \quad (3.2)$$

A discount factor  $\beta$  appears in (3.2) for an analytical reason, which will become clear in the subsequent discussions.

Analytically tractability of the robust taxation problem requires Fourier transform. Due to the existence of the implementation lag, the first order conditions (FOCs) of the Ramsey planner and the adversarial agent contain convolutions of the forms

$$\sum_{k=0}^{\infty} \varphi_{t+k} \hat{D}_k \quad \text{and} \quad \sum_{t=0}^{\infty} \varphi_t \tau_{t-j} \quad (3.3)$$

(where  $\varphi_t$  is a Lagrangian multiplier of the Ramsey problem). The FOCs with these convolutions are hard to manipulate. Fourier transforms turn convolutions into multiplications. Fourier transforms make the problem analytically tractable. The Fourier transforms of the convolutions in (3.3) are

$$\varphi(\omega) \hat{D}(-\omega) \quad \text{and} \quad \varphi(\omega) \tau(-\omega). \quad (3.4)$$

The Fourier transform of a sequence  $\{\beta^{t/2} x_t\}$  is  $x(\omega) \equiv \sum_{t=-\infty}^{\infty} \beta^{t/2} x_t e^{-i\omega t}$ , where  $i^2 = -1$  and  $\omega \in [-\pi, \pi]$ .

This chapter focuses on linear-quadratic (LQ) control problems.<sup>1</sup> The adversarial agent's distortion  $\hat{D}(L)$  depends on the second moments of the control variable, which is the tax rate in the Ramsey problem. The optimal  $\hat{D}(L)$  depends on the variance of  $\{\tau_t\}$  and the covariances between  $\{\tau_t\}$  and the exogenous state variables. This result is intuitive because distorting the lag structure of the stochastic process  $\{D(L)\tau_t\}$  is equivalent to distorting the spectral density of  $\{D(L)\tau_t\}$ , which contains the information of the second moments. In contrast to Hansen and Sargent's robust control theory, for LQ control problems, first moment distortions are sufficient to derive robust decision rules.

## 3.2 Ramsey Taxation

This section formulates a Ramsey taxation problem without capital. The formulation of the problem follows Chari and Kehoe (1999). A LQ approximation of the Ramsey problem will be derived in section 3.4 using the approaches in Benigno and Woodford (2006, 2008). It is well-known that Ramsey problems are not time-consistent in general. Deriving a LQ approximation of the Ramsey problem requires a time-consistent version of the Ramsey problem. The source of time-inconsistency is the forward-looking behaviors of the households. This section formulates a conventional version of the Ramsey problem, which is not time-consistent. Section 3.3 formulates a time-consistent version of the Ramsey problem.

The Ramsey planner maximizes welfare subject to the decision rules in the private sectors. That is, the Ramsey planner's maximization is subject to the household's and the firm's FOCs. The economy is populated by large numbers of homogeneous households and homogeneous firms. Time is discrete. The state of the economy in period  $t$  is  $s_t$ . The history of states up to period  $t$  is  $s^t = (\dots, s_{-1}, s_0, s_1, \dots, s_t)$ . In period 0, the probability of  $s^t$  is  $Pr(s^t)$ . The initial state in period 0,  $s_0$ , is given.

Denote consumption as  $c_t$ , labor as  $l_t$ , and government bond as  $b_t$ .<sup>2</sup> The household's maximization problem is

$$\max_{c_t, l_t, b_t} \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) u(c_t, l_t), \quad (3.5)$$

where the maximization is subject to the household budgets

$$(1 - D(L)\tau_t)w_t l_t + R_t b_{t-1} - c_t - b_t \geq 0, \quad (3.6)$$

for  $t \geq 0$ . The utility function  $u$  is increasing and concave in  $(c_t, l_t)$ . In the household's problem, the initial holding of government bond  $b_{-1}$ , the gross

<sup>1</sup>Linear-quadratic control problems are control problems with quadratic objective functions and linear constraints.

<sup>2</sup>Throughout this chapter, a random variable in period  $t$ , such as  $c_t$ , is a function of the states up to period  $t$ ,  $s^t$ . I do not explicitly write  $c(s^t)$  for notational simplicity. That is, for any random variable  $x_t$  in this chapter,  $x_t \equiv x(s^t)$ .

returns of bonds  $\{R_t\}$ , the wage  $\{w_t\}$ , the labor income tax rates  $\{\tau_t\}$ , and the implementation lag  $D(L)$  are given. The household's FOCs of the maximization problem are

$$u_{l,t} + u_{c,t}(1 - D(L)\tau_t)w_t = 0 \quad (3.7)$$

and the Euler equations

$$-Pr(s^t)u_{c,t} + \sum_{s^{t+1}|s^t} \beta Pr(s^{t+1})u_{c,t+1}R_{t+1} = 0, \quad (3.8)$$

for  $t \geq 0$ .<sup>3</sup>

Firms in the economy have a linear production function  $z_t l_t$  where  $z_t$  is the technological level, an exogenous state variable. The zero profit condition implies that

$$w_t = z_t. \quad (3.9)$$

The household budgets (3.6) and the FOCs (3.7), (3.8), and (3.9) are in the set of constraints for the Ramsey planner's welfare maximization problem. We can get rid of  $b_t$  and  $R_t$  from the Ramsey planner's problem by substituting the Euler equations (3.8) into the household budgets (3.6). If we multiply the household's budgets (3.6) by  $\beta^t Pr(s^t)u_{c,t}$ , sum the budgets over  $s^t$  and  $t$ , and use the Euler equations (3.8), then we have the implementability constraint

$$\sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t)(u_{c,t}c_t + u_{l,t}l_t) = u_{c,0}R_0b_{-1}. \quad (3.10)$$

After solving a Ramsey planner's problem without  $b_t$  and  $R_t$ , we can use the household budgets (3.6) and the Euler equations (3.8) to derive the optimal  $b_t$  and  $R_t$ .

Precisely, the Ramsey planner's problem is to maximize welfare subject to the FOCs (3.7) and (3.9), the implementability constraint (3.10), and the feasibility constraints

$$z_t l_t - c_t - g_t \geq 0, \quad (3.11)$$

for  $t \geq 0$ . In (3.11), the government expenditure  $g_t$  is an exogenous state variable.

### 3.3 Time-consistency

Benigno and Woodford's (2006, 2008) LQ approximation method requires a time-consistent version of the Ramsey problem. This section formulates a time-consistent version of the Ramsey problem.

Ramsey problems are not time-consistent. The households' forward-looking behaviors are the source of time-inconsistency. The Euler equations (3.8), which represent the households' forward-looking behaviors, are used to derive the implementability constraint (3.10). Because of the implementability constraint,

<sup>3</sup>Subscripts denote partial derivatives. For example,  $u_{c,t} \equiv \partial u(c_t, l_t) / \partial c_t$ .

the Ramsey planner's decision rule (FOC) in period 0 is different from that in period  $t \geq 1$ . Thus the optimal policy follows time-invariant rules only for  $t \geq 1$ . If we pay attention to the right hand side of the implementability constraint (3.10), we can see that the Ramsey planner's FOCs with respect to  $c_0$  and  $l_0$  have additional terms from differentiating  $u_{c,0}$ . These additional terms will not appear in the FOCs with respect to  $c_t$  and  $l_t$  for  $t \geq 1$ . Thus the optimal policy rules for  $t = 0$  are different from that for  $t \geq 1$ . This suggests that we may modify the implementability constraint to make the optimal policy follows time-invariant rules for all  $t$ . The modification of the implementability constraint (3.10) follows the method in Benigno and Woodford (2006).

Denote the Ramsey planner's committed asset value to the households in period  $t$  as  $W_t$ :

$$W_t \equiv u_{c,t}R_t b_{t-1} = \sum_{T=t}^{\infty} \sum_{s^T} \beta^{T-t} Pr(s^T | s^t) (u_{c,T} c_T + u_{l,T} l_T). \quad (3.12)$$

The equality in (3.12) follows from the household budgets (3.6), and FOCs (3.7) and (3.8); the equality is derived by a similar method for deriving the implementability constraint. The commitment  $W_t$  can be written recursively as

$$W_t = (u_{c,t} c_t + u_{l,t} l_t) + \sum_{s^{t+1}} \beta Pr(s^{t+1} | s^t) W_{t+1}. \quad (3.13)$$

The implementability constraint (3.10) is rewritten as

$$\sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) (u_{c,t} c_t + u_{l,t} l_t) = W_0, \quad (3.14)$$

where  $W_0$  is the state-contingent committed value of the initial asset

$$W_0 \equiv u_{c,0} R_0 b_{-1}. \quad (3.15)$$

With the modified implementability constraint (3.14), the Ramsey planner's FOCs with respect to  $c_0$  and  $l_0$  do not have the additional terms from the right hand side of the original implementability constraint (3.10).

The choice of initial commitment  $W_0$  is not arbitrary. Since we want to formulate a time-consistent version of the Ramsey taxation problem,  $W_0$  should be self-consistent in the sense that the Ramsey planner will choose the same state-contingent commitment  $W_t$  for all  $t \geq 1$ . In other words,  $W_0$  as a function of state is equal to the function  $W_t$  determined by

$$W_t = \sum_{T=t}^{\infty} \sum_{s^T} \beta^{T-t} Pr(s^T | s^t) (u_{c,T} c_T + u_{l,T} l_T). \quad (3.16)$$

### 3.4 Linear-quadratic Approximation

This section derives a LQ approximation of the time-consistent version of the Ramsey problem by Benigno and Woodford's (2006, 2008) method. The derived

objective function in the LQ approximate problem will correctly rank policies up to second order. Essentially, the LQ approximate problem is maximizing the quadratic terms of the Lagrangian of the original problem subject to the linearized constraints of the original problem. The FOCs of the LQ approximate problem are exactly the linearized FOCs of the original problem. These are nicely explained in the section 1 of Benigno and Woodford (2008). The approximation is around a steady state. The steady states of the Ramsey problem are calculated in appendices 3.11 and 3.12.

The deviation of a variable  $x_t$  from its steady state value  $x$  is denoted as<sup>4</sup>

$$\tilde{x}_t \equiv x_t - x. \quad (3.17)$$

Recall that the time-consistent version of the Ramsey planner's problem is subject to three sets of constraints: the household's FOCs (3.7), the modified implementability constraint (3.14), and the feasibility constraints (3.11). To simplify notations, define

$$H(c_t, l_t, D(L)\tau_t, z_t) \equiv u_{l,t} + u_{c,t}(1 - D(L)\tau_t)z_t \quad (3.18)$$

and

$$\Omega(c_t, l_t) \equiv u_{c,t}c_t + u_{l,t}l_t. \quad (3.19)$$

The Ramsey planner's constraint: the household's FOCs (3.7) and the modified implementability constraint (3.14) can be written as

$$H_t = 0 \quad (3.20)$$

and

$$\sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \Omega_t = W_0. \quad (3.21)$$

The Lagrangian of the Ramsey planner's problem is

$$\begin{aligned} L = & \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) u(c_t, l_t) \\ & + \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \varphi_t H_t \\ & + \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \xi_t (z_t l_t - c_t - g_t) \\ & + \lambda \left( \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \Omega_t - W_0 \right), \end{aligned} \quad (3.22)$$

where  $\varphi_t$ ,  $\xi_t$ , and  $\lambda$  are Lagrangian multipliers.

<sup>4</sup>Throughout this chapter, a random variable without subscript  $t$  denotes a steady state value of this random variable.

The objective function of the LQ approximate Ramsey problem consists of the quadratic terms of the Lagrangian of the original problem (3.22):

$$\max_{c_t, l_t, \tau_t} \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \left( \frac{1}{2} Q_{cc} \tilde{c}_t^2 + \frac{1}{2} Q_{ll} \tilde{l}_t^2 + Q_{cl} \tilde{c}_t \tilde{l}_t + \xi \tilde{l}_t \tilde{z}_t \right) \quad (3.23)$$

where

$$Q_{cc} \equiv u_{cc} + \lambda \Omega_{cc}, \quad (3.24)$$

$$Q_{ll} \equiv u_{ll} + \lambda \Omega_{ll}, \quad (3.25)$$

and

$$Q_{cl} \equiv u_{cl} + \lambda \Omega_{cl}. \quad (3.26)$$

Terms in the household's FOCs (3.20) do not appear in the objective function because  $\varphi = 0$  (which is shown in appendix 3.11, (3.120)). The maximization in (3.23) is subject to the linearized constraints of the original problem: the linearized household's FOCs

$$H_c \tilde{c}_t + H_l \tilde{l}_t + H_\tau D(L) \tilde{\tau}_t + H_z \tilde{z}_t = 0 \quad (3.27)$$

for  $t \geq 0$ , the linearized feasibility constraints

$$z \tilde{l}_t + l \tilde{z}_t - \tilde{c}_t - \tilde{g}_t = 0 \quad (3.28)$$

for  $t \geq 0$ , and the linearized implementability constraint

$$\sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) (\Omega_c \tilde{c}_t + \Omega_l \tilde{l}_t) = \tilde{W}_0. \quad (3.29)$$

We want the initial commitment  $\tilde{W}_0$  to be self-consistent in the sense that  $\tilde{W}_0$  as a function of state is equal to the function

$$\tilde{W}_t = \sum_{T=t}^{\infty} \sum_{s^T} \beta^{T-t} Pr(s^T | s^t) (\Omega_c \tilde{c}_t + \Omega_l \tilde{l}_t). \quad (3.30)$$

A self-consistent initial commitment will make the linearized implementability constraint (3.29) not binding. If we solve the Ramsey problem without the linearized implementability constraint (3.29), then the derived commitment  $\tilde{W}_t$  by substituting the optimal  $\tilde{c}_t$  and  $\tilde{l}_t$  into (3.30) will be self-consistent. Therefore, we can ignore the linearized implementability constraint (3.29).

### 3.5 Robust Ramsey Taxation

This section formulates the robust Ramsey taxation problem under implementation lag uncertainty. The Ramsey planner fears the misspecification of implementation lag polynomial. Although he has an estimate of the implementation lag polynomial,  $D(L)$ , he thinks that any lag polynomial close to the estimate

could be the true implementation lag. He seeks for a robust taxation rule over a set of alternative implementation lag polynomials close to an estimate  $D(L)$ . The closeness is in the  $l^2$  sense. He assumes that there is a hypothetical adversarial agent chooses an implementation lag polynomial  $\hat{D}(L)$  to minimize welfare subject to

$$\sum_{j=0}^{\infty} \beta^j (\hat{D}_j - D_j)^2 \leq \eta. \quad (3.31)$$

A larger  $\eta$  represents a greater fear of misspecification.

The robust Ramsey taxation problem is

$$\begin{aligned} \max_{\tilde{c}_t, \tilde{l}_t, \tilde{\tau}_t} \min_{\hat{D}_j} & \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \left( \frac{1}{2} Q_{cc} \tilde{c}_t^2 + \frac{1}{2} Q_{ll} \tilde{l}_t^2 + Q_{cl} \tilde{c}_t \tilde{l}_t + \xi \tilde{l}_t \tilde{z}_t \right) \\ & + \Theta \sum_{j=0}^{\infty} \beta^j (\hat{D}_j - D_j)^2 \end{aligned} \quad (3.32)$$

where the optimizations in (3.32) are subject to the linearized household's FOCs (3.27), the linearized feasibility constraints (3.28), and the linearized implementability constraint (3.29). The parameter  $\Theta$  in (3.32) measures the fear of model misspecification. A smaller  $\Theta$  penalizes the adversarial agent less for choosing a  $\hat{D}(L)$  that is far away from  $D(L)$ . A smaller  $\Theta$  represents a greater fear of misspecification. So  $\Theta$  is inversely related to  $\eta$ .

Recall that the linearized implementability constraint (3.29) is not binding if the initial commitment  $\tilde{W}_0$  is self-consistent. We can solve the problem without the linearized implementability constraint (3.29). The Lagrangian of the problem (3.32) is

$$\begin{aligned} L_{LQ} &= \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \left( \frac{1}{2} Q_{cc} \tilde{c}_t^2 + \frac{1}{2} Q_{ll} \tilde{l}_t^2 + Q_{cl} \tilde{c}_t \tilde{l}_t + \xi \tilde{l}_t \tilde{z}_t \right) \\ &+ \Theta \sum_{j=0}^{\infty} \beta^j (\hat{D}_j - D_j)^2 \\ &+ \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \tilde{\varphi}_t (H_c \tilde{c}_t + H_l \tilde{l}_t + H_\tau \hat{D}(L) \tilde{\tau}_t + H_z \tilde{z}_t) \\ &+ \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \tilde{\xi}_t (z \tilde{l}_t + l \tilde{z}_t - \tilde{c}_t - \tilde{g}_t). \end{aligned} \quad (3.33)$$

The Ramsey planner's FOCs are

$$Q_{cc} \tilde{c}_t + Q_{cl} \tilde{l}_t + \tilde{\varphi}_t H_c - \tilde{\xi}_t = 0, \quad (3.34)$$

$$Q_{ll} \tilde{l}_t + Q_{cl} \tilde{c}_t + \xi \tilde{z}_t + \tilde{\varphi}_t H_l + \tilde{\xi}_t z = 0, \quad (3.35)$$

and

$$\sum_{k=0}^{\infty} \sum_{s^{t+k}|s^t} \beta^{t+k} Pr(s^{t+k}) \tilde{\varphi}_{t+k} H_\tau \hat{D}_k = 0, \quad (3.36)$$

for all  $t \geq 0$ . The adversarial agent's FOCs are

$$2\Theta\beta^j(\hat{D}_j - D_j) + \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \tilde{\varphi}_t H_{\tau} \tilde{\tau}_{t-j} = 0 \quad (3.37)$$

for all  $j \geq 0$ .

(3.36) and (3.37) are the key equations for solving the robust taxation and distorted implementation lag. The multiplier  $\tilde{\varphi}_t$  is an important variable. Notice the convolutions in (3.36) and (3.37). These convolutions make the equations hard to manipulate. Fourier transform turns these convolutions into multiplications. The problem is tractable with the Fourier transforms of the FOCs.

Solve for the allocations and multipliers. First, the constraints (3.27) and (3.28) implies that

$$\tilde{c}_t = \frac{(-H_z z + H_l l) \tilde{z}_t - H_l \tilde{g}_t - (H_{\tau} z) \hat{D}(L) \tilde{\tau}_t}{H_c z + H_l}, \quad (3.38)$$

$$\tilde{l}_t = \frac{(-H_c l - H_z) \tilde{z}_t + H_c \tilde{g}_t - H_{\tau} \hat{D}(L) \tilde{\tau}_t}{H_c z + H_l}. \quad (3.39)$$

Second, Ramsey Planner's FOCs (3.34) and (3.35) implies that

$$\tilde{\varphi}_t = \frac{(-Q_{cc} z - Q_{cl}) \tilde{c}_t + (-Q_{cl} z - Q_{ll}) \tilde{l}_t - \xi \tilde{z}_t}{H_c z + H_l}, \quad (3.40)$$

$$\tilde{\xi}_t = \frac{(-Q_{cl} H_c + Q_{cc} H_l) \tilde{c}_t + (-Q_{ll} H_c + Q_{cl} H_l) \tilde{l}_t - H_c \xi \tilde{z}_t}{H_c z + H_l}. \quad (3.41)$$

Third, substituting (3.38) and (3.39) into (3.40) and (3.41) yields

$$\tilde{\varphi}_t = \frac{K_z^{\varphi} \tilde{z}_t + K_g^{\varphi} \tilde{g}_t + K_{\tau}^{\varphi} \hat{D}(L) \tilde{\tau}_t}{(H_c z + H_l)^2} \quad (3.42)$$

and

$$\tilde{\xi}_t = \frac{K_z^{\xi} \tilde{z}_t + K_g^{\xi} \tilde{g}_t + K_{\tau}^{\xi} \hat{D}(L) \tilde{\tau}_t}{(H_c z + H_l)^2}, \quad (3.43)$$

where  $K_a^b$  denotes  $b$ 's coefficient of  $a$ :

$$K_z^{\varphi} \equiv (-Q_{cc} z - Q_{cl})(-H_z z + H_l l) + (-Q_{cl} z - Q_{ll})(-H_c l - H_z) - \xi(H_c z + H_l), \quad (3.44)$$

$$K_g^{\varphi} \equiv (-Q_{cc} z - Q_{cl})(-H_l) + (-Q_{cl} z - Q_{ll}) H_c, \quad (3.45)$$

$$K_{\tau}^{\varphi} \equiv (-Q_{cc} z - Q_{cl})(-H_{\tau} z) + (-Q_{cl} z - Q_{ll})(-H_{\tau}), \quad (3.46)$$

$$K_z^{\xi} \equiv (-Q_{cl} H_c + Q_{cc} H_l)(-H_z z + H_l l) + (-Q_{ll} H_c + Q_{cl} H_l)(-H_c l - H_z) + (-H_c \xi)(H_c z + H_l), \quad (3.47)$$

$$K_g^{\xi} \equiv (-Q_{cl} H_c + Q_{cc} H_l)(-H_l) + (-Q_{ll} H_c + Q_{cl} H_l) H_c, \quad (3.48)$$

and

$$K_{\tau}^{\xi} \equiv (-Q_{cl} H_c + Q_{cc} H_l)(-H_{\tau} z) + (-Q_{ll} H_c + Q_{cl} H_l)(-H_{\tau}). \quad (3.49)$$

### 3.6 Equilibrium

For a given state  $s = (\dots, s_{-1}, s_0, s_1, \dots)$ , the Fourier transform of the process  $\{\beta^{t/2}x_t\}$  is denoted as

$$x(\omega, s) \equiv \sum_{t=-\infty}^{\infty} \beta^{\frac{t}{2}} x_t e^{-i\omega t}, \quad (3.50)$$

where  $\omega \in [-\pi, \pi]$ .<sup>5</sup> <sup>6</sup> The Fourier transform of (3.42) is

$$\varphi(\omega, s) = \frac{K_z^\varphi z(\omega, s) + K_g^\varphi g(\omega, s) + K_\tau^\varphi \hat{D}(\omega) \tau(\omega, s)}{(H_c z + H_l)^2}. \quad (3.51)$$

The Fourier transforms of the Ramsey planner's FOC with respect to the tax rate (3.36) and the adversarial agent's FOC (3.37) are

$$\sum_s Pr(s) H_\tau \varphi(\omega, s) \hat{D}(-\omega) = 0 \quad (3.52)$$

and

$$2\Theta(\hat{D}(\omega) - D(\omega)) + \left[ \sum_s Pr(s) H_\tau \varphi(\omega, s) \tau(-\omega, s) \right]_+ = 0. \quad (3.53)$$

$[ \ ]_+$  is the annihilation operator which annihilates the negative degree terms of a polynomial.<sup>7</sup> The negative degree terms are dropped because  $\hat{D}(\omega)$  and  $D(\omega)$  are one-sided polynomials with non-negative degree terms only. Denote expectation as  $E \equiv \sum_s Pr(s)$ . Assume  $H_\tau \neq 0$  and  $(H_c z + H_l)^2 \neq 0$ . Substituting  $\varphi(\omega, s)$  into (3.52) and (3.53) yields

$$[K_z^\varphi E z(\omega, s) + K_g^\varphi E g(\omega, s) + K_\tau^\varphi \hat{D}(\omega) E \tau(\omega, s)] \hat{D}(-\omega) = 0 \quad (3.54)$$

and

$$2\Theta(\hat{D}(\omega) - D(\omega)) = \left[ \begin{aligned} & -\frac{H_\tau K_z^\varphi}{(H_c z + H_l)^2} E z(\omega, s) \tau(-\omega, s) \\ & -\frac{H_\tau K_g^\varphi}{(H_c z + H_l)^2} E g(\omega, s) \tau(-\omega, s) \\ & -\frac{H_\tau K_\tau^\varphi}{(H_c z + H_l)^2} E \tau(\omega, s) \tau(-\omega, s) \hat{D}(\omega) \end{aligned} \right]_+. \quad (3.55)$$

<sup>5</sup>Note  $x(L)$  and  $x(\omega)$  are different polynomials in this chapter. The discounting factor  $\beta$  is in  $x(\omega)$ .  $x(L) = \sum_{j=0}^{\infty} x_j L^j$  but  $x(\omega) = \sum_{j=0}^{\infty} \beta^{j/2} x_j e^{-i\omega j}$ .

<sup>6</sup>The state  $s = (\dots, s_{-1}, s_0, s_1, \dots)$  is given in the summation in (3.50). In (3.50), the summation is for the sequence  $\{\dots, x_{-1}, x_0, x_1, \dots\}$  of a particular sample path. Thus  $x(\omega, s)$  is written explicitly as a function of  $s$ .

<sup>7</sup>For example,  $[x_{-2}(e^{-i\omega})^{-2} + x_{-1}(e^{-i\omega})^{-1} + x_0 + x_1 e^{-i\omega} + x_2(e^{-i\omega})^2]_+ = x_0 + x_1 e^{-i\omega} + x_2(e^{-i\omega})^2$ .

The adversarial agent's lag distortion depends on the second moments of the tax rate, the government expenditure, and the technological level. This is intuitive because distorting the lag structure of  $\{D(L)\tilde{\tau}_t\}$  is equivalent to distorting the spectrum density of  $\{D(L)\tilde{\tau}_t\}$ , which contains the information of second moments. Suppose  $\{x_t\}$  and  $\{y_t\}$  are two second-order stationary processes. Define  $\sigma_{xy}(t-k) \equiv E x_t y_k$ .  $E x(\omega, s) y(-\omega, s)$  is the “discounted cross spectrum” of  $\{x_t\}$  and  $\{y_t\}$  because

$$\begin{aligned} E x(\omega, s) y(-\omega, s) &= E \sum_{t=0}^{\infty} \beta^{\frac{t}{2}} x_t e^{-i\omega t} \sum_{k=0}^{\infty} \beta^{\frac{k}{2}} y_k e^{i\omega k} \\ &= \sum_{t=0}^{\infty} \sum_{k=0}^{\infty} \beta^{\frac{t+k}{2}} \sigma_{xy}(t-k) e^{-i\omega(t-k)} \\ &= \frac{1}{1-\beta} \left( \sum_{t=0}^{\infty} \beta^{\frac{t}{2}} \sigma_{xy}(t) e^{-i\omega t} + \sum_{k=1}^{\infty} \beta^{\frac{k}{2}} \sigma_{xy}(-k) e^{i\omega k} \right) \end{aligned} \quad (3.56)$$

Thus,  $E\tau(\omega, s)\tau(-\omega, s)$  is the “discounted spectrum” of  $\{\tilde{\tau}_t\}$ .  $Ez(\omega, s)\tau(-\omega, s)$  is the “discounted cross spectrum” of  $\{\tilde{z}_t\}$  and  $\{\tilde{\tau}_t\}$ .  $Eg(\omega, s)\tau(-\omega, s)$  is the “discounted cross spectrum” of  $\{\tilde{g}_t\}$  and  $\{\tilde{\tau}_t\}$ .

The adversarial agent may shut down taxation by choosing  $\hat{D}(\omega) = 0$ . This possibility can be ruled out by setting small enough  $\eta$  or large enough  $\Theta$ . If  $\hat{D}(-\omega) \neq 0$ , then (3.54) becomes

$$\hat{D}(\omega) E\tau(\omega, s) = \frac{-K_z^\varphi}{K_\tau^\varphi} E z(\omega, s) + \frac{-K_g^\varphi}{K_\tau^\varphi} E g(\omega, s) \quad (3.57)$$

First, consider the adversarial agent's FOC (3.55) and the “discounted cross spectrum” equation (3.56). The adversarial agent puts a weight on  $\hat{D}_j$  whenever an exogenous state, such as  $\tilde{z}_t$ , is correlated with  $\tilde{\tau}_{t-j}$ . Second, consider the terms relating to  $\tilde{z}_t$  in the adversarial agent's FOC (3.55) and the Ramsey planner's FOC (3.57); the coefficient in the adversarial agent's FOC  $-H_\tau K_z^\varphi / (H_c z + H_l)^2$  has an opposite sign to the coefficient  $-K_z^\varphi / K_\tau^\varphi$  in the Ramsey planner's FOC because  $H_\tau < 0$  and  $K_\tau^\varphi > 0$  (see appendices 3.12). The adversarial agent puts a negative weight whenever the Ramsey planner wants the tax rate to be an increasing function of  $\tilde{z}_t$ , and vice versa. That is, whenever  $\tilde{z}_t$  is correlated with  $\tilde{\tau}_{t-j}$ , the adversarial agent can put a weight, in a direction undesirable to the Ramsey planner, on  $\hat{D}_j$ , so that the effective tax rate  $\hat{D}(L)\tilde{\tau}_t$  is correlated to  $\tilde{\tau}_{t-j}$ , to reduce welfare. A similar comment holds for  $\tilde{g}_t$ .

### 3.7 Purely Stochastic Exogenous Variables

This section derives the robust policy and the optimal policy when the technological level and the government expenditure are purely stochastic. Optimal policies refer to the welfare-maximizing policies when the Ramsey planner totally trust the estimated implementation lag  $D(L)$ . Suppose the technological

level  $\{\tilde{z}_t\}$  and the government expenditure  $\{\tilde{g}_t\}$  are moving average processes:

$$\tilde{z}_t = \sum_{j=0}^{\infty} \alpha_{z,j} \epsilon_{z,t-j} \quad \text{and} \quad \tilde{g}_t = \sum_{j=0}^{\infty} \alpha_{g,j} \epsilon_{g,t-j}, \quad (3.58)$$

where  $\{\epsilon_{z,t}\}$  and  $\{\epsilon_{g,t}\}$  are two uncorrelated white noise processes. The Fourier transforms of  $\{\beta^{t/2} \tilde{z}_t\}$  and  $\{\beta^{t/2} \tilde{g}_t\}$  are

$$z(\omega, s) = \alpha_z(\omega) \epsilon_z(\omega, s) \quad \text{and} \quad g(\omega, s) = \alpha_g(\omega) \epsilon_g(\omega, s). \quad (3.59)$$

Since we are considering a LQ control problem, the policy rule is linear:

$$\tilde{\tau}_t = \sum_{j=0}^{\infty} F_j \epsilon_{z,t-j} + \sum_{j=0}^{\infty} G_j \epsilon_{g,t-j}. \quad (3.60)$$

The Fourier transform of  $\{\beta^{t/2} \tilde{\tau}_t\}$  is

$$\tilde{\tau}(\omega, s) = F(\omega) \epsilon_z(\omega, s) + G(\omega) \epsilon_g(\omega, s). \quad (3.61)$$

The FOC with respect to the tax rate (3.57) implies that the robust policy is determined by

$$F(\omega) = \frac{-K_z^\varphi}{K_\tau^\varphi} \left[ \frac{\alpha_z(\omega)}{\hat{D}(\omega)} \right]_+ \quad \text{and} \quad G(\omega) = \frac{-K_g^\varphi}{K_\tau^\varphi} \left[ \frac{\alpha_g(\omega)}{\hat{D}(\omega)} \right]_+. \quad (3.62)$$

There are annihilation operators because  $\tilde{\tau}_t$  cannot depend on the white noises in the future.  $F(\omega)$  and  $G(\omega)$  are one-sided polynomials with non-negative degree terms only. The optimal policy is given by replacing  $\hat{D}(\omega)$  with  $D(\omega)$ :

$$F(\omega) = \frac{-K_z^\varphi}{K_\tau^\varphi} \left[ \frac{\alpha_z(\omega)}{D(\omega)} \right]_+ \quad \text{and} \quad G(\omega) = \frac{-K_g^\varphi}{K_\tau^\varphi} \left[ \frac{\alpha_g(\omega)}{D(\omega)} \right]_+. \quad (3.63)$$

Given that the exogenous states are moving average and the policy rule is linear, the FOC of the adversarial agent is

$$2\Theta(\hat{D}(\omega) - D(\omega)) = [V_z \alpha_z(\omega) F(-\omega) \sigma_z + V_g \alpha_g(\omega) G(-\omega) \sigma_g + V_\tau [F(\omega) F(-\omega) \sigma_z + G(\omega) G(-\omega) \sigma_g] \hat{D}(\omega)]_+, \quad (3.64)$$

where

$$V_z \equiv -\frac{H_\tau K_z^\varphi}{(H_c z + H_l)^2}, \quad (3.65)$$

$$V_g \equiv -\frac{H_\tau K_g^\varphi}{(H_c z + H_l)^2}, \quad (3.66)$$

$$V_\tau \equiv -\frac{H_\tau K_\tau^\varphi}{(H_c z + H_l)^2}, \quad (3.67)$$

$$\sigma_z \equiv \frac{E\epsilon_{z,t}^2}{1-\beta}, \quad \text{and} \quad \sigma_g \equiv \frac{E\epsilon_{g,t}^2}{1-\beta}. \quad (3.68)$$

(3.56) is used to calculate  $\sigma_z$  and  $\sigma_g$ .

Now suppose that the technological level and the government expenditure are MA(1) processes:  $\alpha_z(\omega) = 1 + \beta^{1/2}\alpha_{z,1}e^{-i\omega}$  and  $\alpha_g(\omega) = 1 + \beta^{1/2}\alpha_{g,1}e^{-i\omega}$ , and that the estimated implementation lag polynomial is  $D(\omega) = D_0 + \beta^{1/2}D_1e^{-i\omega}$  with  $|D_0/D_1| < 1$ .<sup>8</sup> Guess that  $F(\omega) = F_0$  and  $G(\omega) = G_0$ . Then the adversarial agent's FOC implies that

$$\hat{D}(\omega) = \hat{D}_0 + \beta^{\frac{1}{2}}\hat{D}_1e^{-i\omega}, \quad (3.71)$$

where

$$\hat{D}_0 = \frac{2\Theta D_0 + V_z F_0 \sigma_z + V_g G_0 \sigma_g}{2\Theta - V_\tau (F_0^2 \sigma_z + G_0^2 \sigma_g)}, \quad (3.72)$$

and

$$\hat{D}_1 = \frac{2\Theta D_1 + V_z \alpha_{z,1} F_0 \sigma_z + V_g \alpha_{g,1} G_0 \sigma_g}{2\Theta - V_\tau (F_0^2 \sigma_z + G_0^2 \sigma_g)}. \quad (3.73)$$

Suppose  $\Theta$  is large enough (the choice set of the adversarial agent is small enough) so that  $|\hat{D}_0/\hat{D}_1| < 1$ . Then (3.62) implies that the robust policy is determined by

$$F(\omega) = \frac{-K_z^\varphi \alpha_{z,1}}{K_\tau^\varphi \hat{D}_1} \quad \text{and} \quad G(\omega) = \frac{-K_g^\varphi \alpha_{g,1}}{K_\tau^\varphi \hat{D}_1}. \quad (3.74)$$

The guess that  $F(\omega) = F_0$  and  $G(\omega) = G_0$  are confirmed. Use (3.72), (3.73), and (3.74) to solve for  $F_0$  and  $G_0$ . The robust policy is given by

$$F(\omega) = \frac{-K_z^\varphi \alpha_{z,1}}{K_\tau^\varphi D_1} \quad \text{and} \quad G(\omega) = \frac{-K_g^\varphi \alpha_{g,1}}{K_\tau^\varphi D_1}. \quad (3.75)$$

Hence, the distorted implementation lag polynomial is determined by

$$\hat{D}_0 = \frac{2\Theta D_0 + V_z \left( \frac{-K_z^\varphi \alpha_{z,1}}{K_\tau^\varphi D_1} \right) \sigma_z + V_g \left( \frac{-K_g^\varphi \alpha_{g,1}}{K_\tau^\varphi D_1} \right) \sigma_g}{2\Theta - V_z \frac{-K_z^\varphi}{K_\tau^\varphi} \left( \frac{\alpha_{z,1}}{D_1} \right)^2 \sigma_z - V_g \frac{-K_g^\varphi}{K_\tau^\varphi} \left( \frac{\alpha_{g,1}}{D_1} \right)^2 \sigma_g}, \quad (3.76)$$

<sup>8</sup>For a given polynomial  $A_0 + A_1L$ , if  $|A_0/A_1| > 1$  then

$$(A_0 + A_1L)^{-1} = \left[ A_0 \left( 1 + \frac{A_1}{A_0}L \right) \right]^{-1} = \frac{1}{A_0} \sum_{j=0}^{\infty} \left( \frac{A_1}{A_0} \right)^j L^j. \quad (3.69)$$

If  $|A_0/A_1| < 1$  then

$$(A_0 + A_1L)^{-1} = \left[ A_1L \left( \frac{A_0}{A_1L} + 1 \right) \right]^{-1} = \frac{1}{A_1L} \sum_{j=0}^{\infty} \left( \frac{-A_0}{A_1} \right)^j L^{-j}. \quad (3.70)$$

See Sargent (1987, p.178-179) for more discussions on lag polynomials.

and

$$\hat{D}_1 = \frac{2\Theta D_1 + V_z \alpha_z \left( \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{D_1} \right) \sigma_z + V_g \alpha_g \left( \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{D_1} \right) \sigma_g}{2\Theta - V_z \frac{-K_z^\varphi}{K_\tau^\varphi} \left( \frac{\alpha_{z,1}}{D_1} \right)^2 \sigma_z - V_g \frac{-K_g^\varphi}{K_\tau^\varphi} \left( \frac{\alpha_{g,1}}{D_1} \right)^2 \sigma_g}, \quad (3.77)$$

$V_z$  has opposite sign to  $F_0$  and  $V_g$  has opposite sign to  $G_0$  because  $H_\tau < 0$  and  $K_\tau^\varphi > 0$  (see appendix 3.12). That is, the adversarial agent puts negative weights when the Ramsey planner wants the tax to be increasing in the exogenous state variables, and vice versa. The magnitudes of distortions depend on the fluctuations of the exogenous state variables,  $\sigma_z$  and  $\sigma_g$ . Since the exogenous state variables are MA(1) processes ( $\alpha_{z,1} \neq 0$  and  $\alpha_{g,1} \neq 0$ ), the  $\tilde{\tau}_t$  is correlated with  $\tilde{\tau}_{t-1}$ . Thus, the adversarial agent puts a weight on the lagged tax rate.

We can show that the robust policy is exactly equal to the optimal policy by using (3.63). The adversarial agent will choose to distort the implementation lag but the robust policy is equal to the optimal policy because the robust policy cannot depend on the white noises in the future.

This section assumes  $|D_0/D_1| < 1$  so that  $|\hat{D}_0/\hat{D}_1| < 1$  because the problem will be uninteresting otherwise. If  $|\hat{D}_0/\hat{D}_1| > 1$ , then the inverse of  $\hat{D}(\omega)$  is an one-sided polynomial with non-negative degree only; the annihilation operator is not useful in (3.62). This means that the Ramsey planner can always choose a policy rule to totally invert the adversarial agent's distortion so that the Ramsey planner can achieve the same welfare regardless of the choice of the adversarial agent. It is easy to show that, if  $|\hat{D}_0/\hat{D}_1| > 1$ , the adversarial agent will not choose to distort:  $\hat{D}(\omega) = D(\omega)$ , and the Ramsey plan will choose the optimal rule.

### 3.8 General Exogenous Variables

This section derives the robust policy and the optimal policy when there are deterministic components in the exogenous state variables. Suppose the technological level and the government expenditure are

$$\tilde{z}_t = z_t^d + \sum_{j=0}^{\infty} \alpha_{z,j} \epsilon_{z,t-j} \quad \text{and} \quad \tilde{g}_t = g_t^d + \sum_{j=0}^{\infty} \alpha_{g,j} \epsilon_{g,t-j}, \quad (3.78)$$

where  $\{z_t^d\}$  and  $\{g_t^d\}$  are deterministic, and  $\{\epsilon_{z,t-j}\}$  and  $\{\epsilon_{g,t-j}\}$  are uncorrelated white noises. The Fourier transforms of  $\{\beta^{t/2} \tilde{z}_t\}$  and  $\{\beta^{t/2} \tilde{g}_t\}$  are

$$z(\omega, s) = z^d(\omega) + \alpha_z(\omega) \epsilon_z(\omega, s) \quad \text{and} \quad g(\omega, s) = g^d(\omega) + \alpha_g(\omega) \epsilon_g(\omega, s). \quad (3.79)$$

The tax rate can depend on the future values of the deterministic components  $z_t^d$  and  $g_t^d$  but cannot for the future values of the white noises. Write the policy rule as

$$\tilde{\tau}_t = \sum_{j=-\infty}^{\infty} F_j^d z_{t-j}^d + \sum_{j=0}^{\infty} F_j^s \epsilon_{z,t-j} + \sum_{j=-\infty}^{\infty} G_j^d g_{t-j}^d + \sum_{j=0}^{\infty} G_j^s \epsilon_{g,t-j}. \quad (3.80)$$

Coefficients with superscripts  $d$  denote the coefficients of the deterministic components, and superscripts  $s$  denote the stochastic components. The Fourier transform of  $\beta^{t/2}\tilde{\tau}_t$  is

$$\tau(\omega, s) = F^d(\omega)z^d(\omega) + F^s(\omega)\epsilon_z(\omega, s) + G^d(\omega)g^d(\omega) + G^s(\omega)\epsilon_g(\omega, s). \quad (3.81)$$

The deterministic parts of the policy  $F^d(\omega)$  and  $G^d(\omega)$  are two-sided polynomials, while the stochastic parts of the policy  $F^s(\omega)$  and  $G^s(\omega)$  are one-sided polynomials with non-negative degree terms only. The FOC with respect to the tax rate (3.57) implies that the robust policy is determined by

$$F^d(\omega) = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{1}{\hat{D}(\omega)}, \quad F^s(\omega) = \frac{-K_z^\varphi}{K_\tau^\varphi} \left[ \frac{\alpha_z(\omega)}{\hat{D}(\omega)} \right]_+, \quad (3.82)$$

$$G^d(\omega) = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{1}{\hat{D}(\omega)} \quad \text{and} \quad G^s(\omega) = \frac{-K_g^\varphi}{K_\tau^\varphi} \left[ \frac{\alpha_z(\omega)}{\hat{D}(\omega)} \right]_+. \quad (3.83)$$

The optimal policy rule is given by replacing  $\hat{D}(\omega)$  with  $D(\omega)$ :

$$F^d(\omega) = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{1}{D(\omega)}, \quad F^s(\omega) = \frac{-K_z^\varphi}{K_\tau^\varphi} \left[ \frac{\alpha_z(\omega)}{D(\omega)} \right]_+, \quad (3.84)$$

$$G^d(\omega) = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{1}{D(\omega)} \quad \text{and} \quad G^s(\omega) = \frac{-K_g^\varphi}{K_\tau^\varphi} \left[ \frac{\alpha_z(\omega)}{D(\omega)} \right]_+. \quad (3.85)$$

Simple algebra shows that the adversarial agent's FOC is the same as in the case where the technological level and government expenditure are purely stochastic:

$$2\Theta(\hat{D}(\omega) - D(\omega)) = [V_z\alpha_z(\omega)F^s(-\omega)\sigma_z + V_g\alpha_g(\omega)G^s(-\omega)\sigma_g + V_\tau[F^s(\omega)F^s(-\omega)\sigma_z + G^s(\omega)G^s(-\omega)\sigma_g]\hat{D}(\omega)] \quad (3.86)$$

where  $V_z$ ,  $V_g$ ,  $V_\tau$ ,  $\sigma_z$ , and  $\sigma_g$  are defined in (3.65), (3.66), (3.67), and (3.68).

Suppose the stochastic parts of the technological level and the government expenditure are MA(1) processes:  $\alpha_z(\omega) = 1 + \beta^{1/2}\alpha_{z,1}e^{-i\omega}$  and  $\alpha_g(\omega) = 1 + \beta^{1/2}\alpha_{g,1}e^{-i\omega}$ , and the estimated implementation lag polynomial is  $D(\omega) = D_0 + \beta^{1/2}D_1e^{-i\omega}$ . Guess stochastic parts of robust policy rule are  $F^s(\omega) = F_0^s$  and  $G^s(\omega) = G_0^s$ . Then the adversarial agent's FOC (3.86) implies that

$$\hat{D}(\omega) = \hat{D}_0 + \beta^{1/2}\hat{D}_1e^{-i\omega}, \quad (3.87)$$

where

$$\hat{D}_0 = \frac{2\Theta D_0 + V_z F_0^s \sigma_z + V_g G_0^s \sigma_g}{2\Theta - V_\tau ((F_0^s)^2 \sigma_z + (G_0^s)^2 \sigma_g)}, \quad (3.88)$$

and

$$\hat{D}_1 = \frac{2\Theta D_1 + V_z \alpha_{z,1} F_0^s \sigma_z + V_g \alpha_{g,1} G_0^s \sigma_g}{2\Theta - V_\tau ((F_0^s)^2 \sigma_z + (G_0^s)^2 \sigma_g)}. \quad (3.89)$$

Suppose  $\Theta$  is large enough (the choice set of the adversarial agent is small enough) so that  $|\hat{D}_0/\hat{D}_1| < 1$ . (3.82) and (3.83) imply that the robust policy is determined by

$$F^d(L) = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{1}{\hat{D}(\omega)} = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{1}{\hat{D}_1} \sum_{j=0}^{\infty} \left( \frac{-\hat{D}_0}{\hat{D}_1} \right)^j L^{-j-1}, \quad (3.90)$$

$$F^s(L) = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{\hat{D}_1}, \quad (3.91)$$

$$G^d(L) = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{1}{\hat{D}(\omega)} = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{1}{\hat{D}_1} \sum_{j=0}^{\infty} \left( \frac{-\hat{D}_0}{\hat{D}_1} \right)^j L^{-j-1}, \quad (3.92)$$

and

$$G^s(L) = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{\hat{D}_1}. \quad (3.93)$$

The guess that  $F^s(\omega) = F_0^s$  and  $G^s(\omega) = G_0^s$  are confirmed. Use (3.88), (3.89), (3.91), and (3.93) to solve for  $F_0^s$  and  $G_0^s$ . We have

$$F_0^s = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{D_1} \quad \text{and} \quad G_0^s = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{D_1}. \quad (3.94)$$

Consider the optimal policy (3.84) and (3.85) imply that the optimal policy is determined by

$$F^d(L) = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{1}{D(\omega)} = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{1}{D_1} \sum_{j=0}^{\infty} \left( \frac{-D_0}{D_1} \right)^j L^{-j-1}, \quad (3.95)$$

$$F^s(L) = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{D_1}, \quad (3.96)$$

$$G^d(L) = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{1}{D(\omega)} = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{1}{D_1} \sum_{j=0}^{\infty} \left( \frac{-D_0}{D_1} \right)^j L^{-j-1}, \quad (3.97)$$

and

$$G^s(L) = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{D_1}. \quad (3.98)$$

The robust policy and the optimal policy have the same stochastic parts, while the deterministic parts of them are different.

### 3.9 Dynamics of Allocations

This section considers the dynamics of consumption and labor supply under the optimal and robust policies when the estimated lag  $D(L)$  or the distorted lag

$\hat{D}(L)$  is the true lag. To simplify notations, rewrite the equilibrium consumption (3.38) and the equilibrium labor supply (3.39) as

$$\tilde{c}_t = K_z^c \tilde{z}_t + K_g^c \tilde{z}_t + K_\tau^c \tilde{\tau}_t \quad (3.99)$$

and

$$\tilde{l}_t = K_z^l \tilde{z}_t + K_g^l \tilde{z}_t + K_\tau^l \tilde{\tau}_t, \quad (3.100)$$

where

$$K_z^c \equiv \frac{-H_z z + H_l l}{H_c z + H_l}, \quad K_g^c \equiv \frac{-H_l}{H_c z + H_l}, \quad K_\tau^c \equiv \frac{-H_\tau z}{H_c z + H_l}, \quad (3.101)$$

$$K_z^l \equiv \frac{-H_c l - H_z}{H_c z + H_l}, \quad K_g^l \equiv \frac{H_c}{H_c z + H_l}, \quad \text{and} \quad K_\tau^l \equiv \frac{-H_\tau}{H_c z + H_l}. \quad (3.102)$$

Assume the exogenous state variables have both deterministic and stochastic parts, and the stochastic parts are MA(1) processes as in section 3.8. Thus, we consider the optimal and robust policies computed in section 3.8. First, we consider the dynamics under the optimal policy. The dynamic of consumption under the optimal policy with the estimated lag  $D(L)$  is

$$\begin{aligned} \tilde{c}_t &= \left( K_z^c + K_\tau^c \frac{-K_z^\varphi}{K_\tau^\varphi} \right) z_t^d \\ &+ \left[ K_z^c (1 + \alpha_{z,1} L) + K_\tau^c \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{D_1} D(L) \right] \epsilon_{z,t} \\ &+ \left( K_g^c + K_\tau^c \frac{-K_g^\varphi}{K_\tau^\varphi} \right) g_t^d \\ &+ \left[ K_g^c (1 + \alpha_{g,1} L) + K_\tau^c \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{D_1} D(L) \right] \epsilon_{g,t}. \end{aligned} \quad (3.103)$$

The dynamic of labor supply under the optimal policy with the estimated lag  $D(L)$  is

$$\begin{aligned} \tilde{l}_t &= \left( K_z^l + K_\tau^l \frac{-K_z^\varphi}{K_\tau^\varphi} \right) z_t^d \\ &+ \left[ K_z^l (1 + \alpha_{z,1} L) + K_\tau^l \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{D_1} D(L) \right] \epsilon_{z,t} \\ &+ \left( K_g^l + K_\tau^l \frac{-K_g^\varphi}{K_\tau^\varphi} \right) g_t^d \\ &+ \left[ K_g^l (1 + \alpha_{g,1} L) + K_\tau^l \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{D_1} D(L) \right] \epsilon_{g,t}. \end{aligned} \quad (3.104)$$

The dynamic of consumption under the optimal policy with the distorted lag  $\hat{D}(L)$  is

$$\tilde{c}_t = \left( K_z^c + K_\tau^c \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\hat{D}(L)}{D(L)} \right) z_t^d$$

$$\begin{aligned}
& + \left[ K_z^c(1 + \alpha_{z,1}L) + K_\tau^c \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{D_1} \hat{D}(L) \right] \epsilon_{z,t} \\
& + \left( K_g^c + K_\tau^c \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\hat{D}(L)}{D(L)} \right) g_t^d \\
& + \left[ K_g^c(1 + \alpha_{g,1}L) + K_\tau^c \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{D_1} \hat{D}(L) \right] \epsilon_{g,t}. \quad (3.105)
\end{aligned}$$

The dynamic of labor supply under the optimal policy with the distorted lag  $\hat{D}(L)$  is

$$\begin{aligned}
\tilde{l}_t & = \left( K_z^l + K_\tau^l \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\hat{D}(L)}{D(L)} \right) z_t^d \\
& + \left[ K_z^l(1 + \alpha_{z,1}L) + K_\tau^l \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{D_1} \hat{D}(L) \right] \epsilon_{z,t} \\
& + \left( K_g^l + K_\tau^l \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\hat{D}(L)}{D(L)} \right) g_t^d \\
& + \left[ K_g^l(1 + \alpha_{g,1}L) + K_\tau^l \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{D_1} \hat{D}(L) \right] \epsilon_{g,t}. \quad (3.106)
\end{aligned}$$

Second, we consider the dynamics under robust policy. The dynamic of consumption under the robust policy with the estimated lag  $D(L)$  is

$$\begin{aligned}
\tilde{c}_t & = \left( K_z^c + K_\tau^c \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{D(L)}{\hat{D}(L)} \right) z_t^d \\
& + \left[ K_z^c(1 + \alpha_{z,1}L) + K_\tau^c \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{\hat{D}_1} D(L) \right] \epsilon_{z,t} \\
& + \left( K_g^c + K_\tau^c \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{D(L)}{\hat{D}(L)} \right) g_t^d \\
& + \left[ K_g^c(1 + \alpha_{g,1}L) + K_\tau^c \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{\hat{D}_1} D(L) \right] \epsilon_{g,t}. \quad (3.107)
\end{aligned}$$

The dynamic of labor supply under the robust policy with the estimated lag  $D(L)$  is

$$\begin{aligned}
\tilde{l}_t & = \left( K_z^l + K_\tau^l \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{D(L)}{\hat{D}(L)} \right) z_t^d \\
& + \left[ K_z^l(1 + \alpha_{z,1}L) + K_\tau^l \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{\hat{D}_1} D(L) \right] \epsilon_{z,t} \\
& + \left( K_g^l + K_\tau^l \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{D(L)}{\hat{D}(L)} \right) g_t^d \\
& + \left[ K_g^l(1 + \alpha_{g,1}L) + K_\tau^l \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{\hat{D}_1} D(L) \right] \epsilon_{g,t}. \quad (3.108)
\end{aligned}$$

The dynamic of consumption under the robust policy with the distorted lag  $\hat{D}(L)$  is

$$\begin{aligned}
\tilde{c}_t &= \left( K_z^c + K_\tau^c \frac{-K_z^\varphi}{K_\tau^\varphi} \right) z_t^d \\
&+ \left[ K_z^c (1 + \alpha_{z,1} L) + K_\tau^c \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{\hat{D}_1} \hat{D}(L) \right] \epsilon_{z,t} \\
&+ \left( K_g^c + K_\tau^c \frac{-K_g^\varphi}{K_\tau^\varphi} \right) g_t^d \\
&+ \left[ K_g^c (1 + \alpha_{g,1} L) + K_\tau^c \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{\hat{D}_1} \hat{D}(L) \right] \epsilon_{g,t}. \tag{3.109}
\end{aligned}$$

The dynamic of labor supply under the robust policy with the distorted lag  $\hat{D}(L)$  is

$$\begin{aligned}
\tilde{l}_t &= \left( K_z^l + K_\tau^l \frac{-K_z^\varphi}{K_\tau^\varphi} \right) z_t^d \\
&+ \left[ K_z^l (1 + \alpha_{z,1} L) + K_\tau^l \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,1}}{\hat{D}_1} \hat{D}(L) \right] \epsilon_{z,t} \\
&+ \left( K_g^l + K_\tau^l \frac{-K_g^\varphi}{K_\tau^\varphi} \right) g_t^d \\
&+ \left[ K_g^l (1 + \alpha_{g,1} L) + K_\tau^l \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,1}}{\hat{D}_1} \hat{D}(L) \right] \epsilon_{g,t}. \tag{3.110}
\end{aligned}$$

We can observe that the dynamics of consumption and labor supply do not depend on the future values of the deterministic parts of the exogenous state variables when the true lag coincide with the anticipated lag, for example, the dynamics under robust policy do not depend on the future values of the deterministic parts when the true lag is  $\hat{D}(L)$ .

### 3.10 Remarks

The complete set of equilibria for all situations is hard to find. Sections 3.7, 3.8, and 3.9 just show a particular set of conditions so that an equilibrium for the zero sum game between the Ramsey planner and the adversarial agent exists. Consider the case where the exogenous state variables have deterministic and stochastic parts as in section 3.8. Assume that the stochastic parts of the technological level and the government expenditure are MA( $h$ ) processes:  $\alpha_z(L) = \sum_{j=0}^h \alpha_{z,j} L^j$  and  $\alpha_g(L) = \sum_{j=0}^h \alpha_{g,j} L^j$ , and that the estimated lag polynomial has highest degree  $k$ :  $D(L) = \sum_{j=0}^k D_j L^j = \prod_{j=0}^k (1 - d_j L)$  where  $|d_j| > 1$  for all  $j$  (that is, the inverse of  $D(L)$  has no positive degree term). As shown in section 3.7 and 3.8, the adversarial agent's FOC for case where the exogenous states have deterministic parts (3.86) is the same as the FOC for the case where the exogenous have no deterministic part (3.64). That is,

existence of the deterministic parts does not affect the distortion chosen by the adversarial agent. Thus we focus on the stochastic parts of the robust policy, and the deterministic parts can be calculated easily by (3.82) and (3.83).

First, suppose  $h < k$ . Then the stochastic parts of the robust policy are  $F_s(L) = 0$  and  $G_s(L) = 0$  and  $\hat{D}(L) = D(L)$ . That is, the robust taxation is zero (precisely, the tax rate is constant at a steady state value) and the adversarial agent does not distort the implementation lag.

Second, suppose  $h = k$ . Then  $\hat{D}(L) = \sum_{j=0}^k \hat{D}_j L^j = \prod_{j=0}^k (1 - \hat{d}_j L)$  where

$$\hat{D}_j = \frac{2\Theta D_j + V_z \alpha_{z,j} F_0^s \sigma_z + V_g \alpha_{g,j} G_0^s \sigma_g}{2\Theta - V_\tau ((F_0^s)^2 \sigma_z + (G_0^s)^2 \sigma_g)}, \quad (3.111)$$

for  $j = 0, 1, \dots, h$ . Assume  $\Theta$  is large enough so that  $|\hat{d}_j| > 1$  for all  $j$ ; the inverse of  $\hat{D}(L)$  has no positive degree term. The stochastic parts of the robust policy are  $F_s(L) = F_0^s$  and  $G_s(L) = G_0^s$  where

$$F_0^s = \frac{-K_z^\varphi}{K_\tau^\varphi} \frac{\alpha_{z,h}}{D_h} \quad \text{and} \quad G_0^s = \frac{-K_g^\varphi}{K_\tau^\varphi} \frac{\alpha_{g,h}}{D_h}. \quad (3.112)$$

Note the stochastic section 3.8 is a special where  $h = 1$ . As in section 3.8, the stochastic parts of the robust policy are the same as the stochastic parts of the optimal policy.

Third, suppose  $h > k$ . Then an equilibrium does not exist.

### 3.11 Appendix: Steady States

This appendix derives the equations that determine the set of steady states. We will see that there is one degree of freedom to choose a steady state labor income tax rate or government debt. The steady state values of allocations and multipliers are calculated in the appendix 3.12. In the time-consistent version of the Ramsey problem derived in section 3.3, the Ramsey planner maximizes welfare subject to the FOCs (3.7) and (3.9), the modified implementability constraint (3.14), and the feasibility constraints (3.11). The Lagrangian of the Ramsey planner's problem is

$$\begin{aligned} L = & \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) u(c_t, l_t) \\ & + \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \varphi_t [u_{l,t} + u_{c,t} (1 - D(L) \tau_t) z_t] \\ & + \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) \xi_t [z_t l_t - c_t - g_t] \\ & + \lambda \left[ \sum_{t=0}^{\infty} \sum_{s^t} \beta^t Pr(s^t) (u_{c,t} c_t + u_{l,t} l_t) - W_0 \right]. \end{aligned} \quad (3.113)$$

The FOCs of the Ramsey planner's problem are

$$u_{c,t} + \varphi_t[u_{lc,t} + u_{cc,t}(1 - D(L)\tau_t)z_t] - \xi_t + \lambda[u_{c,t} + u_{cc,t}c_t + u_{lc,t}l_t] = 0, \quad (3.114)$$

$$u_{l,t} + \varphi_t[u_{ll,t} + u_{cl,t}(1 - D(L)\tau_t)z_t] + \xi_t z_t + \lambda[u_{cl,t}c_t + u_{l,t} + u_{ll,t}l_t] = 0, \quad (3.115)$$

and

$$\sum_{k=0}^{\infty} \sum_{s^{t+k}|s^t} \beta^{t+k} P_T(s^{t+k}) \varphi_{t+k} u_{c,t+k} (-D_k) z_{t+k} = 0. \quad (3.116)$$

Evaluating the FOCs (3.114), (3.115), and (3.116) at a steady state yields

$$u_c + \varphi[u_{lc} + u_{cc}(1 - D(L)\tau)z] - \xi + \lambda[u_c + u_{cc}c + u_{lc}l] = 0, \quad (3.117)$$

$$u_l + \varphi[u_{ll} + u_{cl}(1 - D(L)\tau)z] + \xi z + \lambda[u_{cl}c + u_l + u_{ll}l] = 0, \quad (3.118)$$

and

$$\sum_{k=0}^{\infty} \beta^{t+k} \varphi u_c (-D_k) z = 0. \quad (3.119)$$

In the above equations, a variable without subscript  $t$  refers to a steady state value of that variable. If  $D_k \neq 0$  for some  $k$ , (3.119) implies that

$$\varphi = 0. \quad (3.120)$$

Substituting (3.120) into (3.117) and (3.118) gives

$$u_c - \xi + \lambda[u_c + u_{cc}c + u_{lc}l] = 0, \quad (3.121)$$

and

$$u_l + \xi z + \lambda[u_{cl}c + u_l + u_{ll}l] = 0. \quad (3.122)$$

It remains to evaluate the constraints at a steady state. The FOC of household with respect to labor (3.7), the feasibility constraint (3.11), and the implementability constraint (3.14) at a steady state are

$$u_l + u_c(1 - D(L)\tau)z = 0, \quad (3.123)$$

$$zl - c - g = 0, \quad (3.124)$$

and

$$\frac{u_c c + u_l l}{1 - \beta} = W. \quad (3.125)$$

We have five equations: (3.121), (3.122), (3.123), (3.124), and (3.125) to determine six variables:  $c$ ,  $l$ ,  $W$ ,  $\tau$ ,  $\xi$ , and  $\lambda$ . We have one degree of freedom to choose the steady state tax rate  $\tau$  or the steady state government debt  $b$ .

We can use the steady state commitment  $W$  to solve for the steady state value of government debt  $b$ . By the definition of commitment (3.12), the steady state value of government bond  $b$  is determined by

$$W = u_c R b. \quad (3.126)$$

In (3.126), the steady state gross return of government  $R$  is determined by the Euler equation (3.8):

$$R = \beta^{-1}. \quad (3.127)$$

### 3.12 Appendix: Steady States with Log Utility

This appendix calculates the steady state allocations, the steady state multipliers, and the important constants and coefficients for a given utility function. The utility function is

$$u(c, l) = \log c + a \log(1 - l). \quad (3.128)$$

The derivatives of the utility function are

$$u_c(c, l) = \frac{1}{c}, \quad (3.129)$$

$$u_{cc}(c, l) = \frac{-1}{c^2}, \quad (3.130)$$

$$u_{ccc}(c, l) = \frac{2}{c^3}, \quad (3.131)$$

$$u_l(c, l) = \frac{-a}{1-l}, \quad (3.132)$$

$$u_{ll}(c, l) = \frac{-a}{(1-l)^2}, \quad (3.133)$$

$$u_{lll}(c, l) = \frac{-2a}{(1-l)^3}. \quad (3.134)$$

$\Omega$  is defined in (3.19) in section 3.4. At a steady state,  $\Omega$  and its derivatives are

$$\Omega = u_c c + u_l l = 1 - \frac{al}{1-l}, \quad (3.135)$$

$$\Omega_c = u_c + u_{cc}c + u_{lc}l = 0, \quad (3.136)$$

$$\Omega_l = u_{cl} + u_l + u_{ll}l = \frac{-a}{(1-l)^2}, \quad (3.137)$$

$$\Omega_{cc} = 2u_{cc} + u_{ccc}c + u_{lcc}l = 0, \quad (3.138)$$

$$\Omega_{ll} = u_{cll}c + 2u_{ll} + u_{lll}l = \frac{-2a}{(1-l)^3}, \quad (3.139)$$

$$\Omega_{cl} = u_{cl} + u_{ccl}c + u_{lc} + u_{lcl}l = 0. \quad (3.140)$$

In appendix 3.11, (3.121) and (3.122) determine the steady state values of  $\lambda$  and  $\xi$ :

$$\lambda = \frac{u_l + u_c z}{-\Omega_c z - \Omega_l} \quad (3.141)$$

and

$$\xi = \frac{-\Omega_l u_c + u_l \Omega_c}{-\Omega_c z - \Omega_l}. \quad (3.142)$$

The denominator of them is

$$-\Omega_c z - \Omega_l = \frac{a}{(1-l)^2} > 0. \quad (3.143)$$

The numerator of  $\lambda$  is

$$u_l + u_c z = \frac{-a}{1-l} + \frac{z}{c}. \quad (3.144)$$

The numerator of  $\xi$  is

$$-\Omega_l u_c + u_l \Omega_c = \frac{a}{c(1-l)^2} > 0. \quad (3.145)$$

Recall that we have one degree of freedom to choose the steady state tax rate  $\tau$  or the steady state commitment  $W$ . We may express the steady state allocations in terms of  $\tau$  or  $W$ . The Ramsey planner's has three constraints. They are the household's FOC

$$u_l + u_c(1 - D(L)\tau)z = \frac{-a}{1-l} + \frac{1}{c}(1 - D(L)\tau)z = 0, \quad (3.146)$$

the feasibility constraint

$$zl - c - g = 0, \quad (3.147)$$

and the implementability constraint

$$\frac{u_c c + u_l l}{1 - \beta} = \frac{1 - \frac{al}{1-l}}{1 - \beta} = W. \quad (3.148)$$

Use the household FOC (3.146) and the feasibility constraint (3.147) to solve for the steady state consumption and labor in term of the steady state tax rate  $\tau$ . The steady state consumption and labor are

$$c = \frac{(z - g)(1 - D(L)\tau)}{a + (1 - D(L)\tau)} \quad (3.149)$$

and

$$l = \frac{ag + (1 - D(L)\tau)z}{az + (1 - D(L)\tau)z}. \quad (3.150)$$

Use the feasibility constraint (3.147) and the modified implementability (3.148) to solve for the steady state consumption and labor in terms of steady state commitment  $W$ . The steady state consumption and labor are

$$c = z \frac{1 - W(1 - \beta)}{1 + a - W(1 - \beta)} - g \quad (3.151)$$

and

$$l = \frac{1 - W(1 - \beta)}{1 + a - W(1 - \beta)}. \quad (3.152)$$

The steady state tax rate  $\tau$  and the steady state commitment  $W$  are related by

$$\frac{az - ag - a[ag + (1 - D(L)\tau)z]}{az - ag} = W(1 - \beta). \quad (3.153)$$

Now, we can express the steady state values of  $\lambda$  and  $\xi$  in terms of exogenous variables. The steady state  $\lambda$  is

$$\lambda = -(1-l) + \frac{z(1-l)^2}{ac} = \frac{a(z-g)D(L)\tau}{[a+(1-D(L)\tau)]z(1-D(L)\tau)} > 0. \quad (3.154)$$

The steady state  $\xi$  is

$$\xi = \frac{1}{c} = \frac{a+(1-D(L)\tau)}{(z-g)(1-D(L)\tau)} > 0. \quad (3.155)$$

Note that  $\xi$  is equal to the marginal value of consumption  $u_c = 1/c$ .

In section 3.4,  $H$  is defined in (3.18) to represent the household FOC with respect to labor. The coefficients in the linearized household FOC are

$$H_c = u_{lc} + u_{cc}(1-D(L)\tau)z = \frac{-[a+(1-D(L)\tau)]^2 z}{(z-g)^2(1-D(L)\tau)} < 0, \quad (3.156)$$

$$H_l = u_{ll} + u_{cl}(1-D(L)\tau)z = \frac{-[a+(1-D(L)\tau)]^2 z^2}{a(z-g)^2} < 0, \quad (3.157)$$

$$H_z = u_c(1-D(L)\tau) = \frac{a+(1-D(L)\tau)}{(z-g)} > 0, \quad (3.158)$$

$$H_\tau = u_c(-1)z = \frac{-[a+(1-D(L)\tau)]z}{(z-g)(1-D(L)\tau)} < 0. \quad (3.159)$$

In the LQ approximate problem (3.32), the coefficients in the quadratic objective are

$$Q_{cc} = \frac{-[a+(1-D(L)\tau)]^2}{[(z-g)(1-D(L)\tau)]^2} < 0, \quad (3.160)$$

$$Q_{ll} = \frac{-[a+(1-D(L)\tau)]^2 z^2(1+D(L)\tau)}{a(z-g)^2(1-D(L)\tau)} < 0, \quad (3.161)$$

$$Q_{cl} = 0. \quad (3.162)$$

Calculate the constants appear in  $\tilde{c}_t$ ,  $\tilde{l}_t$ ,  $\tilde{\varphi}_t$  and  $\tilde{\xi}_t$ . In (3.38) and (3.39), we have

$$-H_z z + H_l l = \frac{-[a+(1-D(L)\tau)]^2 z^2}{a(z-g)^2} < 0, \quad (3.163)$$

$$-H_{cl} - H_z = \frac{[a+(1-D(L)\tau)]^2 g}{(z-g)^2(1-D(L)\tau)} > 0, \quad (3.164)$$

$$H_c z + H_l = \frac{-[a+(1-D(L)\tau)]^3 z^2}{(z-g)^2(1-D(L)\tau)} < 0. \quad (3.165)$$

In (3.40) and (3.41), we have

$$-Q_{cc} z - Q_{cl} = \frac{[a+(1-D(L)\tau)]^2 z}{(z-g)^2(1-D(L)\tau)^2} > 0, \quad (3.166)$$

$$-Q_{cl}z - Q_{ll} = \frac{[a + (1 - D(L)\tau)]^2 z^2 (1 + D(L)\tau)}{a(z - g)^2 (1 - D(L)\tau)} > 0, \quad (3.167)$$

$$-Q_{cl}H_c + Q_{cc}H_l = \frac{[a + (1 - D(L)\tau)]^4 z^2}{a(z - g)^4 (1 - D(L)\tau)^2} > 0, \quad (3.168)$$

$$-Q_{ll}H_c + Q_{cl}H_l = \frac{-[a + (1 - D(L)\tau)]^4 z^3 (1 + D(L)\tau)}{a(z - g)^4 (1 - D(L)\tau)^2} < 0, \quad (3.169)$$

$$-H_c \xi = \frac{[a + (1 - D(L)\tau)]^3 z}{(z - g)^3 (1 - D(L)\tau)^2} > 0. \quad (3.170)$$

The coefficients of  $\tilde{\varphi}_t$  are

$$K_z^\varphi = \frac{[a + (1 - D(L)\tau)]^4 z^2 ((a - 1)(z - g) + D(L)\tau g)}{a(z - g)^4 (1 - D(L)\tau)^2}, \quad (3.171)$$

$$K_g^\varphi = \frac{-[a + (1 - D(L)\tau)]^4 z^3 D(L)\tau}{a(z - g)^4 (1 - D(L)\tau)^2} < 0, \quad (3.172)$$

$$K_\tau^\varphi = \frac{[a + (1 - D(L)\tau)]^3 z^3 [a + (1 + D(L)\tau)(1 - D(L)\tau)]}{a(z - g)^3 (1 - D(L)\tau)^3} > 0. \quad (3.173)$$

Note that  $K_z^\varphi > 0$  if  $(a - 1)(z - g) + D(L)\tau g > 0$ . For  $K_z^\varphi > 0$ , we need  $a > 1$ , small  $z - g$ , or large  $D(L)\tau g$ . The coefficients of  $\tilde{\xi}_t$  are

$$K_z^\xi = \frac{-[a + (1 - D(L)\tau)]^6 z^3 [z(1 - D(L)\tau) + ag(1 + D(L)\tau) + a^2(z - g)]}{a^2(z - g)^6 (1 - D(L)\tau)^3} < 0, \quad (3.174)$$

$$K_g^\xi = \frac{[a + (1 - D(L)\tau)]^6 z^4 [(1 - D(L)\tau) + a(1 + D(L)\tau)]}{a^2(z - g)^6 (1 - D(L)\tau)^3} > 0, \quad (3.175)$$

$$K_\tau^\xi = \frac{-[a + (1 - D(L)\tau)]^5 z^4 D(L)\tau}{a(z - g)^5 (1 - D(L)\tau)^3} < 0. \quad (3.176)$$

In order to calculate the coefficients in the robust or optimal taxation, we need

$$\frac{K_z^\varphi}{K_\tau^\varphi} = \frac{[a + (1 - D(L)\tau)][(a - 1)(z - g) + D(L)\tau g](1 - D(L)\tau)}{(z - g)z[a + (1 + D(L)\tau)(1 - D(L)\tau)]} \quad (3.177)$$

$$\frac{K_g^\varphi}{K_\tau^\varphi} = \frac{[a + (1 - D(L)\tau)](-D(L)\tau)(1 - D(L)\tau)}{(z - g)[a + (1 + D(L)\tau)(1 - D(L)\tau)]} < 0. \quad (3.178)$$

Then the coefficients in the dynamics of  $\tilde{c}_t$  and  $\tilde{l}_t$  are

$$\begin{aligned} & (-H_z z + H_l l) + H_\tau z \frac{K_z^\varphi}{K_\tau^\varphi} \\ &= \frac{-[a + (1 - D(L)\tau)]^2 z}{a(z - g)^2 [a + (1 + D(L)\tau)(1 - D(L)\tau)]}, \\ & \times \{z[a + (1 + D(L)\tau)(1 - D(L)\tau)] + a[(a - 1)(z - g) + D(L)\tau g]\} \end{aligned}$$

$$< 0, \quad (3.179)$$

$$\begin{aligned} & -H_l + H_\tau z \frac{K_g^\varphi}{K_\tau^\varphi} \\ &= \frac{z^2[a + (1 - D(L)\tau)]^2 \{ [a + (1 + D(L)\tau)(1 - D(L)\tau)] + aD(L)\tau \}}{a(z - g)^2 [a + (1 + D(L)\tau)(1 - D(L)\tau)]} \\ & > 0, \end{aligned} \quad (3.180)$$

$$\begin{aligned} & (-H_c l - H_z) + H_\tau \frac{K_z^\varphi}{K_\tau^\varphi} \\ &= \frac{[a + (1 - D(L)\tau)]^2 [(2ag - az + z) + (az - ag - z)D(L)\tau]}{(z - g)^2 (1 - D(L)\tau) [a + (1 + D(L)\tau)(1 - D(L)\tau)]}, \end{aligned} \quad (3.181)$$

$$\begin{aligned} & H_c + H_\tau \frac{K_g^\varphi}{K_\tau^\varphi} \\ &= \frac{[a + (1 - D(L)\tau)]^2 z [-a - 1 - D(L)\tau]}{(z - g)^2 (1 - D(L)\tau) [a + (1 + D(L)\tau)(1 - D(L)\tau)]} < 0. \end{aligned} \quad (3.182)$$

# Bibliography

- [1] Pierpaolo Benigno and Michael Woodford. Optimal taxation in an RBC model: A linear-quadratic approach. *Journal of Economic Dynamics and Control*, 30(9-10):1445–89, September-October 2006.
- [2] Pierpaolo Benigno and Michael Woodford. Linear-quadratic approximation of optimal policy problems. August 2008.
- [3] Norman Biggs. *Algebraic Graph Theory*. Cambridge University Press, Cambridge, Great Britain, 1974.
- [4] Lawrence E. Blume, William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides. Identification of social interactions. In Jess Benhabib, Alberto Bisin, and Matthew Jackson, editors, *Handbook of Social Economics*, volume 1B, chapter 18, pages 853–964. North-Holland, Amsterdam, 2011.
- [5] Lawrence E. Blume, William A. Brock, Steven N. Durlauf, and Rajshri Jayaraman. Linear social network models, May 2011.
- [6] Yann Bramoulle, Habiba Djebbari, and Bernard Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55, May 2009.
- [7] William A. Brock and Steven N. Durlauf. Interaction-based models. In J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics*, volume 5, chapter 54, pages 3297–380. North-Holland, Amsterdam, 2001.
- [8] William A. Brock and Steven N. Durlauf. Elements of a theory of design limits to optimal policy. *NBER Working Paper Series*, Working Paper 10495, May 2004.
- [9] V. V. Chari and Patrick J. Kehoe. Optimal fiscal and monetary policy. In John B. Taylor and Michael Woodford, editors, *Handbook of Macroeconomics*, volume 1C, chapter 26, pages 1671–745. North-Holland, Amsterdam, 1999.
- [10] Jane Cooley. Desegregation and the achievement gap: Do diverse peers help?, March 2007.

- [11] Charles W. Curtis. *Linear Algebra: An Introductory Approach*. Springer-Verlag, New York, 1984.
- [12] Linda Datcher. Effects of community and family background on achievement. *The Review of Economics and Statistics*, 64(1):32–41, February 1982.
- [13] Laurent Davezies, Xavier D’Haultfoeuille, and Denis Fougere. Identification of peer effects using group size variation. *Econometrics Journal*, 12(3):397–413, November 2009.
- [14] Steven N. Durlauf. Poverty traps and Appalachia, August 2010.
- [15] Steven N. Durlauf and Hisatoshi Tanaka. Understanding regression versus variance tests for social interactions. *Economic Inquiry*, 46(1):25–28, January 2008.
- [16] Masahisa Fujita and Hideaki Ogawa. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics*, 12(2):161–96, May 1982.
- [17] Bryan S. Graham. Identifying social interactions through conditional variance restrictions. *Econometrica*, 76(3):643–60, May 2008.
- [18] Paul R. Halmos. *Finite-dimensional Vector Spaces*. D. Van Nostrand Company, Inc., Princeton, New Jersey, second edition, 1958.
- [19] Lars Peter Hansen and Thomas J. Sargent. *Robustness*. Princeton University Press, Princeton, New Jersey, 2008.
- [20] Cheng Hsiao. Identification. In Zvi Griliches and Michael D. Intriligator, editors, *Handbook of Econometrics*, volume 1, chapter 4, pages 223–83. North-Holland, Amsterdam, 1983.
- [21] Matthew O. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton, New Jersey, 2008.
- [22] Lung-fei Lee. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2):333–74, October 2007.
- [23] R. E. Lucas, Jr. Externalities and cities. *Review of Economic Dynamics*, 4(2):245–74, April 2001.
- [24] Robert E. Lucas, Jr. and Esteban Rossi-Hansberg. On the internal structure of cities. *Econometrica*, 70(4):1445–76, July 2002.
- [25] Charles F. Manski. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–42, July 1993.

- [26] Robert A. Moffitt. Policy interventions, low-level equilibria, and social interactions. In Steven N. Durlauf and H. Peyton Young, editors, *Social Dynamics*, chapter 3, pages 45–82. The MIT Press, Cambridge, Massachusetts, 2001.
- [27] W. Keith Nicholson. *Linear Algebra with Applications*. McGraw-Hill Education, Boston, fourth edition, 2002.
- [28] Hideaki Ogawa and Masahisa Fujita. Equilibrium land use patterns in a nonmonocentric city. *Journal of Regional Science*, 20(4):455–75, November 1980.
- [29] Thomas J. Sargent. *Macroeconomic Theory*. Academic Press, San Diego, second edition, 1987.
- [30] Ronald Shaw. *Linear Algebra and Group Representations*, volume 1. Academic Press, London, 1982.
- [31] T. P. Speed. What is an analysis of variance? *The Annals of Statistics*, 15(3):885–910, September 1987.
- [32] Piet Van Mieghem. *Graph Spectra for Complex Networks*. Cambridge University Press, Cambridge, United Kingdom, 2011.
- [33] Halbert White. *Asymptotic Theory for Econometricians*. Academic Press, Inc., Orlando, 1984.
- [34] Halbert White. *Asymptotic Theory for Econometricians*. Academic Press, San Diego, revised edition, 2001.