CAUSALITY AND AGGREGATION IN ECONOMICS: THE USE OF HIGH DIMENSIONAL PANEL DATA IN MICRO-ECONOMETRICS AND MACRO-ECONOMETRICS

A Dissertation

by

DAE-HEUM KWON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2007

Major Subject: Agricultural Economics

CAUSALITY AND AGGREGATION IN ECONOMICS:

THE USE OF HIGH DIMENSIONAL PANEL DATA IN

MICRO-ECONOMETRICS AND MACRO-ECONOMETRICS

A Dissertation

by

DAE-HEUM KWON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	David A. Bessler
Committee Members,	Oral Capps Jr.
	Dennis W. Jansen
	David J. Leatham
Head of Department,	John P. Nichols

December 2007

Major Subject: Agricultural Economics

ABSTRACT

Causality and Aggregation in Economics:
The Use of High Dimensional Panel Data in
Micro-Econometrics and Macro-Econometrics. (December 2007)
Dae-Heum Kwon, B.S., Korea University;
M.S., Korea University
Chair of Advisory Committee: Dr. David A. Bessler

This study proposes one plausible procedure to address two methodological issues, which are common in micro- and macro- econometric analyses, for the full realization of research potential brought by recently available high dimensional data. To address the issue of how to infer the causal structure from empirical regularities, graphical causal models are proposed to inductively infer causal structure from non-temporal and non-experimental data. However, the (probabilistic) stability condition for the graphical causal models can be violated for high dimensional data, given that close co-movements and thus near deterministic relations are oftentimes observed among variables in high dimensional data. Aggregation methods are proposed as one possible way to address this matter, allowing one to infer causal relationships among disaggregated variables based on aggregated variables. Aggregation methods also are helpful to address the issue of how to incorporate a large information set into an empirical model, given that econometric considerations, such as degrees-of-freedom and multicollinearity, require an economy of parameters in empirical models. However, actual aggregation requires legitimate classifications for interpretable and consistent aggregation.

Based on the generalized condition for the consistent and interpretable aggregation derived from aggregation theory and statistical dimensional methods, we propose plausible methodological procedure to consistently address the two related issues of causal inference and actual aggregation procedures. Additional issues for empirical studies of micro-economics and macro-economics are also discussed. The proposed procedure provides an inductive guidance for the specification issues among the direct, inverse, and mixed demand systems and an inverse demand system, which is statistically supported, is identified for the consumer behavior of soft drink consumption. The proposed procedure also provides ways to incorporate large information set into an empirical model with allowing structural understanding of U.S. macro-economy, which was difficult to obtain based on the previously used factor augmented vector autoregressive (FAVAR) framework. The empirical results suggest the plausibility of the proposed method to incorporate large information sets into empirical studies by inductively addressing multicollinearity problem in high dimensional data.

DEDICATION

To my parents,

Won-Dal, Kwon, Jeong-Im, Kim

and brothers,

Kyu-Heum, Kwon, Jeong-Heum, Kwon.

ACKNOWLEDGEMENTS

I express sincere gratitude to many people, since I could not complete this dissertation without them. I am very grateful for their encouragement and support.

First and foremost, I am greatly thankful for Dr. David A. Bessler, who is my mental mentor as well as academic advisor. I shall not forget his academic and mental guidance. He has not only helped to shape this research with his academic advice but also mentally supported me with his patience. He has been with me in every period of my studies and every step of this dissertation.

I would like thank to all of the committee members: Dr. David J. Leatham, Dr. Oral Capps Jr., and Dr. Dennis W. Jansen. They give me valuable comments on the dissertations. I also would like to thank all of the other teachers who have taught me. Through their insightful courses and personal dialogues, they taught me valuable knowledge and insights for every topic. I would like to thank my family. I do express my sincere appreciation to my parents and brothers. Without their support, I could not finish my studies at Texas A&M. I give special thanks to my wife, Eun-Sook Ahn, who always supports and encourages me with patience and love. With her love and prayers, she has shared all my frustrating periods and supported me to finish this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

CHAPTER I
INTRODUCTION

Recent advances in data processing capabilities have brought the possibility of analyzing larger numbers of detailed variables. In many areas of economics, high dimensional panel data are now available. For example, retail checkout scanner data are available for thousand of products at firm, regional and national levels at various frequencies. Central banks and statistical institutes produce a large number of macro-economic time series data. These data have brought forth research potentials for significant advances in the micro-econometric analysis of consumer behavior (Capps and Love, 2002) and the macro-econometric study of monetary policy effects (Stock and Watson, 2005). The availability of high dimensional data, however, raises several methodological issues for the full use of the research potentials brought by this large information set. An important methodological issue to be addressed is how to incorporate such available broad range of information set into empirical models, given that econometric considerations, such as degrees-of-freedom and multicollinearity, require an economy of parameters in empirical models. Another methodological issue is how to determine the causal structure to relate empirical regularities captured in a reduced form model to theoretical properties represented by a structural form model (identification problem). Given that identifying a system of equations means determining the causal structure, the identification problem arises from the facts that: (a) the causal structure is generally under-determined by the statistical properties of the data (induction problem). (b) theories are too heterogeneous to provide a conclusive causal structure or overall theories do not provide sufficient information to identify causal structure. A simple but fundamental version of this issue is how to relate correlation patterns to causal structures.

How to infer the causal structure from empirical regularities and how to incorporate the large information set into an empirical model are two important methodological issues, which bring a more fundamental methodological issue. Is there a specific correct aggregation level? To deal with these fundamental issues consistently, we interpret theory as an inductive causal averaging procedure that concentrates only on similar tendencies to highlight a few common factors by ignoring many more individual differences and idiosyncrasies. When we follow an inductive causal averaging procedure, we need to identify empirically justifiable conditions that allow us to legitimately define common tendencies and individual idiosyncrasies. This issue is

---

This dissertation follows the style of *American Journal of Agricultural Economics*.

studied in the context of aggregation theory and some generalized conditions for consistent aggregation are derived. Based on the derived generalized condition for consistent aggregation, we propose one possible methodological procedure to consistently address the two related issues of causal inference and actual aggregation procedures.

In chapter II, the general methodological issues are discussed and a plausible procedure is proposed for the full realization of the research potentials brought by high dimensional data. More specifically, first, we provide a brief outline of developments on these issues to motivate this study. Second, graphical causal models are discussed to address the causality issue of how to infer the causal structure to relate empirical regularities captured in a reduced form model to theoretical properties represented by the structural form model. A (probabilistic) stability condition, which is one of the fundamental assumptions of the graphical causal models, is discussed in the context of the use of a high dimensional data set. Third, aggregation theory is discussed to identify consistent aggregation conditions, under which the common tendencies and individual idiosyncrasies can be legitimately defined. A compositional stability condition, which is proposed as a generalized condition for consistent aggregation, is discussed to address the information issue of how to incorporate large information set into an empirical model. Index number theory and statistical dimensional reduction methods are then discussed in the context of generalized conditions of aggregation theory. The relationship between the (probabilistic) stability condition for the causality issue and the compositional stability condition for the information issue is discussed. Based on the generalized condition for the consistent aggregation, an inductive method to systematically address causality and aggregation issues is proposed for the full use of the research potentials brought by high dimensional data.

The proposed method is illustrated with retail checkout scanner data and macro-economic time series panel data as examples of two sets of high dimensional data. In chapter III, the proposed method is illustrated for micro-econometric analysis of consumer behavior. When it can be considered as one of the main objectives of the study of consumer behavior to understand and measure responsiveness of consumer behavior to changes in exogenous variables, the empirical measure of responsiveness relies on three specification choices in an empirical model. First, given that there are full spectrums of direct, inverse, and mixed demand systems and the general relationship between elasticity and flexibility is not yet established, the measure depends on the relative predeterminess among the price and quantity variables represented by dependent and explanatory variables in an empirical model of a specific commodity. Second, given that small departures from valid classification and/or aggregation can result in large mistakes in empirical results, the measure depends on the classification and aggregation to define price and quantity variables themselves. For example, the decision on classification and aggregation can substantially affect the conclusions about elasticity estimates in multi-stage budgeting approach, because cross-price elasticities or cross-quantity flexibilities among products in different groups are likely to be small by construction. Third, given that the different assumptions used to parameterize functional relationships have different implications, the measure depends on the functional form, which relates the dependent variable with explanatory variables. For example, there are four combinations of constant or variation assumptions for the income (or scale) coefficient and Slutsky (or Antonelli) coefficient in the differential functional form approach as captured in popular demand systems specifications.

In chapter III, we propose an inductive empirical method to address these three methodological issues in the study of consumer behavior based on the discussion on the causality and aggregation issues in chapter II. The way to incorporate theoretical implications into empirical model specifications through the functional forms and the way to compare different specifications of direct, inverse, and mixed demand functions are the additional issues to be addressed. More specifically, first, the specification choice issue among direct, inverse, and mixed demand functions is addressed by using the inductively inferred causal information based on the graphical causal models. Second, the classification and aggregation issue are addressed by the compositional stability conditions and index number theory. Third, the functional form issue is addressed by the synthetic model approach based on the differential functional form framework. The comparison of alternative specifications is conducted in terms of model selection framework. The proposed method is illustrated with an application for soft drink products using retail checkout scanner data.

In chapter IV, the proposed method is illustrated for macro-econometric analysis of the U.S. macro-economy. Two methodological issues for the full realization of the research potential brought by the available high dimensional data are discussed. One is the identification problem of how to infer the underlying causal structure from the data, given that the causal structure is generally underdetermined by the statistical properties of the data and theory does not provide sufficient causal information. Unlike the structural equation model (SEM) approach which requires too much causal information for the identification problem, the vector autoregressive (VAR) model approach provides the possibility of inferring causal information from statistical properties of the data without pretending to have too much *a priori* theory and/or without demanding too much information from the data. Although the structural VAR framework provides the possibility of inferring causal information from data, how to inductively infer the causal structure to relate empirical regularities captured in the reduced form model to theoretical properties represented by the structural form model remains an open methodological issue. The other methodological issue to be addressed is how to incorporate an available large information set into an empirical model, given that econometric considerations such as degrees-of-freedom and multicollinearity require the economy of parameters in empirical models. This information problem is important, since misspecification problems can exist due to the small information set usually incorporated in empirical macro-econometric models, given the observation that monetary authorities monitor a large number of economic variables and there can be many possible channels through which the monetary policy affects the economy.

In chapter IV, we propose inductive empirical methods to address these two methodological issues in the study of monetary policy effects based on the discussions on the causality and aggregation issues in chapter II. A method to infer the causal structures for the study of the monetary policy transmission mechanism and a method to incorporate a broad range of information into the empirical macro-model are the primary issues to be addressed. More specifically, first, the SEM and VAR approaches are compared in terms of the identification problem. The relative advantage of the VAR approach beyond the recursive Wold causal chain system and the possibility of an inductive inference on the causal structure are discussed. Second, possible misspecification problems due to the small information set incorporated in the standard VAR approach is discussed in the context of the monetary transmission mechanism literature. The possibility both to incorporate high dimensional macro-economic panel data into a standard VAR model and to infer a structural interpretation for this large information set is discussed based on the factor augmented vector autoregressive (FAVAR) framework and the compositional stability conditions. Third, an identification issue in the FAVAR model is addressed by using inductively inferred causal information based on the graphical causal models. The proposed methods are illustrated with the applications for the study of the monetary policy effects using macro-economic panel data.

In chapter V, the proposed methodological procedure is summarized and several research topics to be further studied are suggested as concluding remarks.

CHAPTER II
CAUSALITY AND AGGREGATION IN ECONOMICS

Recent advances in data processing capabilities have brought the possibility of analyzing larger numbers of detailed variables. In many areas of economics, high dimensional panel data are now available. For example, retail checkout scanner data are available for thousand of products at firm, regional and national levels at various frequencies. And central banks and statistical institutes produce a large number of macro-economic time series data. These data have brought forth research potentials for significant advances in the micro-econometric analysis of consumer behavior and the macro-econometric study of monetary policy effects. The availability of high dimensional data, however, raises several methodological issues for the full use of the research potentials brought by this large information set. An important methodological issue to be addressed is how to incorporate such available broad range of information set into empirical models, given that econometric considerations, such as degrees-of-freedom and multicollinearity, require an economy of parameters in empirical models.

Empirical studies in economics have been developed to unify the theoretical-quantitative approach with the empirical-statistical approach to identify either the structural parameters corresponding to the coefficients in the structural equation model (SEM) approach or the effects of structural economic shocks in the structural vector autoregressive (VAR) model approach. Given that identifying a system of equations means determining the causal structure, the identification problem arises from the following facts: (a) The causal structure is generally under-determined by the statistical properties of the data (induction problem). A simple but fundamental version of this induction problem is that correlation does not imply causation. (b) Theories are too heterogeneous to provide a conclusive causal structure or overall theories do not provide sufficient information to identify causal structure. In this respect, another methodological issue is how to determine the causal structure to relate empirical regularities captured in reduced form model to theoretical properties represented by the structural form model (identification problem). A simple but fundamental version of this issue is how to relate correlation pattern to causal structure.

How to infer the causal structure from empirical regularities and how to incorporate the large information set into an empirical model are two important issues, which bring a more fundamental methodological issue for the full use of the research potentials brought by high dimensional data. Is there a specific correct aggregation level? Where we should apply a theoretical model of rational behavior? To what level should the regularity assumptions associated with rationality be applied? Are these to be applied at the individual level, to reasonably homogeneous groups, or to entire economies? These questions have been discussed for a very long time and have turned out to be difficult to solve. It might only be properly addressed by manipulative (randomized) experimentations or more extensive empirical research than has been performed to date (Blundell and Stoker, 2005).

To deal with these fundamental issues consistently, we interpret theory as an inductive causal averaging procedure that concentrates only on similar tendencies to highlight a few common factors by ignoring many more individual differences and idiosyncrasies. For example, the theory of firm (or consumer) can be understood as an inductive model that does not describe the actual objective function and constraints of any particular firm (or consumer) but only what most firms (or consumers) have in common as a tendency. It comes from observing the behavior of many firms (or consumers) and, based on those observations, abstracting the basic elements common to most of those firms (or consumers). In this respect, theory is considered to be a foundation for developing a more realistic account of the firm (or consumer) under consideration (Davis, 1999).

When we follow an inductive causal averaging procedure that concentrates only on similar tendencies to highlight a few common factors by ignoring many more individual differences and idiosyncrasies, we need to identify empirically justifiable conditions that allow us to legitimately define common tendencies and individual idiosyncrasies. This issue can be addressed in the context of an aggregation theory and some generalized conditions for consistent aggregation. Based on the generalized condition for the consistent aggregation, we propose one possible methodological procedure to consistently address the two related issues of causal inference and actual aggregation procedures. More specifically, first, we provide a brief outline of developments on these issues to motivate this study. Second, graphical causal models are discussed to address the causality issue of how to determine the causal structure to relate empirical regularities captured in a reduced form model to theoretical properties represented by the structural form model. A (probabilistic) stability condition, which is one of the fundamental assumptions of the graphical causal models, is discussed in the context of the use of a high dimensional data set. Third, aggregation theory is discussed to identify consistent aggregation conditions, under which the common tendencies and individual idiosyncrasies can be legitimately defined. A compositional stability condition, which is proposed as a generalized condition for consistent aggregation, is discussed to address the information issue of how to incorporate large information set into an empirical model. Index number theory and statistical dimensional reduction methods are then discussed in the context of generalized conditions of aggregation theory. The relationship between the (probabilistic) stability condition for the causality issue and the compositional stability condition for the information issue is discussed. Based on the generalized condition for the consistent aggregation, an inductive method to systematically address causality and aggregation issues are proposed for the full use of the research potentials brought by high dimensional data.

Brief Survey

Empirical studies in economics have relied on economic theories or researchers' intuitions in order to identify either the structural parameters corresponding to the coefficients in the structural equation model (SEM) approach or the effects of structural economic shocks in the structural vector autoregressive (VAR) model approach. While the SEM approach emphasizes the relative importance of deductive information and proceeds from the deductive information to inductive information, the VAR approach emphasizes the relative importance of inductive information and proceeds from the inductive information to deductive information. In the SEM approach, the economic theory or intuitive knowledge specifies *a priori* the causal structure and then statistical methods are applied to measure the strength of the causal relations and the possibility is pursued to test the restrictions derived from theory. In the structural VAR, on the other hand, statistical properties of economic time series are summarized by the reduced form VAR and then the causal structures are used based on either the theoretical implications or institutional knowledge. The structural equations approach, especially the Cowles Commission approach, pursues both necessary and sufficient algebraic conditions that make a system of equations identified, emphasizing the role of economic theory in identification. On the other hand, the VAR approach is more data intense at least in the estimation step, arguing that the absence of purely exogenous variables in observational data impedes algebraic solution of the identification problem. The VAR approach pursues the possibility of (absolutely) inductive methods minimizing, or without using, the deductive *a priori* information to infer the underlying causal structures from the statistical observations.

Given that identifying a system of equations means determining the causal structure among variables in the system and theory does not provide sufficient or conclusive information about causal structure, several empirical methods for learning causal relationships from data have been pursued. Hume provides philosophical foundations for the causality issues in economics by providing following definitions of the causal relation: “We may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object has not been, the second never had existed (Hume, 2000, page 54).” While the first part of the Hume’s definition is related to the probabilistic approach, the second part of the definition is related to the counterfactual approach. Suppes (1970) elaborates the first part of the Hume’s definition as follows: an event A causes an event B if (a) The conditional probability of B given A is greater than B alone (prima facie causality), and (b) A occurs before B. Based on a similar idea, Granger (1980) proposes an operational definition as follows: a (time-series) variable A causes B, if the probability of B conditional on its own past history and the past history of A does not equal the probability of B conditional on its own past history alone. On the other hand, Lewis (1986) elaborates the second part of the Hume’s definition as follows: the event A causes the event B if and only if (abbreviated by iff hereafter in all the subsequent chapters) (a) Both A and B happen and (b) If A had not been, then B would not have happened. Holland (1986) describes a statistical approach to causal inference based on this idea.

Granger-causality has been used in macro-econometric models, especially in time-series approach, whereas the Holland’s method has been applied in micro-econometric models, especially in experimental settings. However, given that causality denotes the possibility of controlling one variable in order to influence another one (efficient cause), Granger-causality does not fully address the causal issue, since it is based on the incremental predictability rather than an efficient cause. And given that most economic data are generated from non-experimental settings and the randomized experiment method is not feasible in general, Holland’s method can not be used for empirical studies in general situations, since it is based on the counterfactuals which we cannot observe without experiments.

In this respect, the problem of differentiating between causal relations and empirical regularities has remained an open issue in the development of econometrics. However, the inductive methods of learning causal relationships from non-temporal and non-experimental data have been developed by mathematically connecting probabilistic dependencies to graphical concepts at three universities: UCLA, Carnegie Mellon University (CMU) and Stanford in mid-1980 (Pearl, 2000). Researchers at UCLA and CMU teams pursued an approach, where (a) The fragments of the underlying structure are identified by searching the data patterns of conditional independencies and (b) The identified fragments are logically combined together to form a coherent causal model or a set of such models (see Spirtes et al., 2000 and Pearl, 2000 for examples). On the other hand, researchers at Stanford University and a number of other teams pursued a Bayesian approach, where data are used to update the posterior probabilities assigned to the candidate causal structures. This Bayesian approach provides the basis for several graph-based learning methods (see Buntine, 1996 and Heckerman, 1996 for examples). While these graphical causal models or directed acyclic graph (DAG) approaches are gradually finding their way into economics, the graphical causal models are based on the Markov and stability conditions as the underlying assumptions. Given that the Markov condition is assumed in most empirical studies, the stability condition can be problematic and thus require careful checking in using these inductive causal inference methods for high dimensional data. These issues will be discussed later.

Empirical studies in economics have also relied on various forms of classification and aggregation, since econometric considerations, such as degrees-of-freedom and multicollinearity, require an economy of parameters in empirical models. The full review of these classification and/or aggregation issue is beyond the scope of this study, since separate fields follow very different paths with regard to these issues. However, identifying a legitimate, but less restrictive, condition for a consistent aggregation remains an open issue in general. For example, in the consumption area, where the aggregation issue has been intensively discussed due to its importance in both micro-economics and macro-economics, consistent aggregation conditions have been studied in terms of both commodity-wise and agent-wise aggregations. For the commodity-wise aggregation, even though the Hicks-Leontief composite commodity theorem and the homothetic or weak separability concepts have been discussed in empirical micro-economic studies, it has been demonstrated that these two types of conditions provide only restrictive possibilities for consistent aggregation in empirical applications. For agent-wise aggregation, the issue of aggregating from individual agents to an aggregate unit is oftentimes ignored in standard macro-economic models by assuming individuals behind the aggregation to be representative agents, even though it has been recognized that the changing composition of economic agents and their incomes have significant implications on the aggregation issue.

In a more general methodological setting, Theil argues that only very restrictive special conditions allow aggregate models to be consistent with disaggregate models and predictions through micro-equations yield more precise estimates of the aggregate dependent macro-variable than the corresponding macro-equations. Despite his generally negative conclusions for aggregation approaches, Theil's arguments provide a general methodological framework for the aggregation issue (Theil, 1954). This general framework has inspired a considerable amount of related research, much of which has attempted to identify less restrictive legitimate aggregation conditions. Furthermore, Griliches (1972) argue that different true models can exist at different aggregation levels and they can be related by both the aggregation rules and the properties of the distribution of the micro-variables. However, identifying generalized legitimate aggregation conditions remains an open issue in aggregation theory.

Another issue, which has been somewhat separately discussed from the issue of identifying generalized legitimate conditions for consistent aggregation, is how to actually represent original variables by aggregate variables or how to decide the weighting schemes in aggregating the disaggregated micro-variables into the aggregated macro-variables. Index number theory has been the main approach followed in the economic literature. On the other hand, principal component and factor analyses have been primary approaches in the statistical literature. Index number theory has been developed based on the dual pairs of information of prices and quantities from economic transactions and provided theoretical background for many statistical institutes to generate economic data. Different index formulas can be understood based on five different approaches: the fixed basket, differential, economic, stochastic and axiomatic approaches. Statistical dimensional reduction methods have been developed in more general settings. The standard factor model is introduced in economics, when it is used for study of the arbitrage pricing theory (APT) in financial economics. On the other hand, dynamic factor models have been developed in macro-economics recently, when they are used to allow distributed lag effects of factors on individual variables in a general dynamic setting. The relationship between factor analysis and principal component analysis has been established in both static and dynamic settings. It has been demonstrated that these two statistical dimensional reduction methods are useful to incorporate broad range of information into empirical models. However, given that these aggregation methods are oftentimes discussed without explicit linkage to legitimate aggregation conditions, there remains an open issue as to the conditions under which these aggregation methods can be used.

As we briefly discussed above, even though there have been significant advances, there remain several open issues in using the previously suggested methods to address causality and aggregation issues for empirical applications, especially with a high dimensional data set. Given that the advances for these issues have been developed separately, it is necessary to consistently connect the methodological developments related with causality and aggregation issues with some generalizations. The generalization of legitimate aggregation conditions can be the main element for the required procedures. As we will discuss subsequently, we propose one possible methodological procedure to consistently address the related issues of causality and aggregation for the full realization of the research potentials brought forth by high dimensional data.

Graphical Causal Model

How to infer the causal structure from the observational data has been a fundamental issue in empirical studies for a long time, given that the causal structure is generally underdetermined by the statistical properties of the data (induction problem). A simple but fundamental version of this issue is how to relate correlation pattern to causal structure. The graphical causal model explicitly aims to inductively infer the causal structure that generated statistical properties of the sample data. According to the graphical causal model, causality is based on the manipulative view under the modular situation, where a complex system can be built by combining simpler local parts. Given that each local causal relationship represents a stable and autonomous physical mechanism, it is possible to manipulate one such relationship without changing the others and to test whether the (marginal) distribution of B is sensitive to the interventions on A. This type of verification provides the semantic basis of the claim that variable A has a causal influence on another variable B. In this manipulative view of causality, the causal claims are based on the behavior of two variables A and B under the influence of a third variable C. When the causal structure implies some pattern of informational (in)dependencies among triplets, which is captured by the patterns of (un)conditional (in)dependencies, the criterion for causation between two variables A and B can be whether a third variable C exhibits a specific pattern of (in)dependency with A and B. The graphical causal model is based on the following propositions: (a) Not all but a certain pattern of (un)conditional (in)dependencies reveal underlying causal directionality empirically, and (b) By logically combining such partially revealed information on causal directionality, it is possible to infer coherent causal structures or a set of such structures under certain conditions.

To capture dependency patterns mentioned above, the graphical causal model introduces the concept of a dependency model. Let A, B, C, and D denote four disjoint subsets of variables in this chapter for notational consistency. When we can determine informational irrelevance as a local property, such as "A is independent of B given C" or "A and B interact only via C", we can define conditional independent statements $I(A,B|C)$ among triplets. And when we can determine whether $I(A,B|C)$ is true for all possible triplets in the model M, we can also define a dependency model M by using all independent statements $I(M)$ which are true among a set of variables in the model M. Given that dependency can be defined as the negation of independency, we can use $D(A,B|C) \equiv \sim I(A,B|C)$ and $D(M)$ for dependent statements of individual triplets and of dependency model M respectively. When two different dependency models M and M' have the same set of variables, certain relationships among dependency models M and M' can be defined. M' is an independence-map (I-map) of M if $I(M') \subseteq I(M)$ so $I(A,B|C)$ in M' implies $I(A,B|C)$ in M. This means that all the conditional independence statements derived from a dependency model M' also hold in another dependency model M. M' is a dependence-map (D-map) of M if $D(M') \subseteq D(M)$ so $D(A,B|C)$ in M' implies $D(A,B|C)$ in M. This means that all the conditional dependence statements derived from a dependency model M' also hold in another dependency model M. Note that a relation that M' is a D-map of M implies another relation that M is a I-map of M' and vice versa, because dependency is negation of independency. M' is an perfect-map (P-map) of M if M' is both I-map and D-map of M, which implies $I(M') = I(M)$ and $D(M') = D(M)$. This means that all the conditional independence and dependence statements derived from a dependency model M' also hold in another dependency model M and vice versa. (Bouckaert, 1993).

The graphical causal model introduces two types of dependency models. And the graphical causal model can be explained by the relationships among dependency models. A joint probability distribution can define a probabilistic dependency model M_P by using conditional independence criteria. On the other hand, a graph also can define a graphical dependency model M_G by using graphical separation criteria. A probabilistic dependency model M_P is introduced by following two main lines of reasoning. First, even the most assertive and exhaustive causal proposition is usually subject to exceptions, either because randomness occurs due to our ignorance of the underlying boundary conditions or because all nature's laws are inherently probabilistic. So causes tend to make their consequences more likely, but not absolutely certain. Probability theory allows us to focus on the main issue of causality by virtue of being equipped to tolerate unexplained exceptions. Second, empirical information becomes verifiable or falsifiable by statistical methods. Empirical knowledge can be encoded in conditional probability statements and a joint probability distribution is computed from those statements through Bayes's rule: $P(A,B) = P(A|B)P(B) = P(B|A)P(A)$, where $A|B$ stands for an event A in the context specified by B and $P(A|B) = P(A,B)/P(B)$ specifies the belief in A under the assumption that B is known with certainty. In this respect, $P(A|B)$ can also be read that B probabilistically causes A with the quantitative belief of $P(A|B)$. Conditional independency in a probabilistic dependency model M_P captures the informational irrelevance structure among disjoint subsets of variables. A is independent of B given C, written as $I(A,B|C)$, means that once we know C, knowledge of B does not provide additional information about A, and thus learning B would no longer influence our belief in A or the probability of A. More formally, A is conditionally independent of B given C, iff $P(A|B,C) = P(A|C)$ or $P(A,B|C) = P(A|C)P(B|C)$. The unconditional or marginal independence can be treated as a particular case of conditional independence such as $I(A,B|\varnothing)$, iff $P(A|B,\varnothing) = P(A|\varnothing) = P(A)$ or $P(A,B|\varnothing) = P(A)P(B)$.

To understand the graphical dependency model M_G, the following graphical concepts are introduced. A graph model consists of a set of vertices (or nodes) V corresponding to variables and a set of edges (or links or arcs) E that connect some pair of variables. Each edge can be either directed or undirected to denote a certain relationship in pairs of variables. A pair of nodes is adjacent if they are connected by either an undirected edge or a directed edge. A triple $< A, C, B >$ is unshielded iff A is adjacent to C, B is adjacent to C, but A is not adjacent to B. C is a collider of A and B if both A and B direct into C. Given that C is a collider of A and B, C is shielded-collider of A and B if A and B are also adjacent and C is an unshielded-collider of A and B if A and B are not adjacent. Two nodes are connected if a path exists between two nodes in a graph and they are disconnected otherwise, where a path is a sequence of consecutive edges of any directionality. When two sets of nodes A and B are connected or interact only via third set C, conditioning on C can be understood as a blocking those interactions. The (un)conditional independence in graphical dependency model M_G is characterized by (a) The lack of edges between nodes or lack of information flow between variables as well as (b) A graphical concept of separating the dependency between nodes or of blocking (or screening-off) the information flows between variables. An undirected graph has a simple definition of separation. Two sets of nodes A and B are separated by a third set C in undirected graph, iff every path between the nodes in A and B contains at least one node in C. In such a case, a set C is called as a Cutset separating A and B. A directed acyclic graph (DAG), which is a directed graph with an acyclic constraint, has a more complicated notion of separation in order to capture directionality. A set S_{AB} is said to d-separated (directionally separated) A and B iff S_{AB} blocks every path between A and B. More specifically, a path is said to be d-separated by a Sepset (separating set) S_{AB} in a DAG iff (1) a path contains $A \rightarrow C \rightarrow B$ or $A \leftarrow C \leftarrow B$ (causal chain) or $A \leftarrow C \rightarrow B$ (causal fork) such that the middle node C is in Sepset S_{AB} and (2) a path contains an $A \rightarrow C \leftarrow B$ (inverted fork, unshielded collider, or v-structure) such that the middle node C and any descendents of C are not in Sepset S_{AB}. Note that the acyclic constraint is needed to define the graphical dependency model when we use d-separation as a conditional independence criterion. While undirected graphs or Markov networks (Pearl, 1988) are used primarily to represent symmetrical relationships, Directed graphs, especially DAGs or Bayesian networks, (Pearl, 1985) have been used to represent asymmetrical causal relationships. Since the causality is the issue to be addressed in this study, our discussion of graphical dependency models M_G are restricted to the directed acyclic graph (DAG) not the undirected graph.

These two types of dependency models have distinctive features. A probabilistic dependency model M_P provides an empirical or statistical method to infer patterns of conditional independencies from observational or non-experimental sample data, which involves probabilistic calculations. On the other hand, a graphical dependency model M_G provides a logical method for a qualitative characterization of conditional independence pattern in terms of graphical topology, which does not involve numerical calculations. There exist relationships between the two dependency models under certain conditions. It has been demonstrated that when it is assumed that a probability distribution satisfies the Markov and stability conditions, DAG is a perfect map of a probabilistic dependency model for the continuous normal distribution (Pearl, 1988) and for the discrete multinomial distribution (Meek, 1995b). The Markov and stability conditions can be understood by representing a causal model as a set of equations in the form of $X_i = f_i(Pa_i, u_i)$, $\forall i = 1, \cdots, I$, where Pa_i (denoting parents) stands for the set of variables judged to be immediate causes of X_i and u_i represent errors due to omitted factors. If it is assumed that there are no cycles representing mutual causations or feedback processes (causal acyclic condition), then the corresponding model is called semi-Markovian. And in addition to the acyclic condition, if it is assumed that a set of measured variables in the model includes all the common causes of all the pairs of variables, so the error terms u_i are mutually independent (causal sufficiency condition), then the model is called Markovian. Note that the causal Markov condition is based on both acyclic and sufficiency conditions. Note also that these two conditions are assumed in most empirical studies in economics, although using these conditions can be problematic. The model is defined to be stable (Pearl and Verma, 1991) or faithful (Spirtes et al. 2000) or a DAG-isomorphism (Pearl 1988), if it is assumed that all the (un)conditional (in)dependencies are invariant to parametric changes represented by the functions $f_i(\cdot)$ and the distributions $P(u_i)$. This means that all the unconditional and conditional probabilistic structures are stable with respect to changes of their numerical values. This stability condition has following implications: (a) All the observed (un)conditional probabilistic structures are due to the underlying causal structures, not their special numerical values in probabilistic structures. (b) No (in)dependence in probability dependency model can be destroyed or induced by changing probabilistic parameter values. (c) It is possible to effectively and efficiently encode (un)conditional (in)dependencies structures into graphical dependency model without numerical probabilities. Thus, with the Markov condition, (d) It is possible to infer the underlying causal structures from the observed marginal and conditional probabilistic structures, where the observation is done through the statistical decisions based on either the Neyman-Pearson type statistical test (conditional independence test approach) or the Bayesian information criterion (goodness-of-fit scoring approach).

It has been mathematically demonstrated that a necessary and sufficient condition for a probability distribution to be Markov is that every variable be independent of all its nondescendants, conditional on its direct parents Pa_i (see Pearl and Verma, 1991 for example). This implies that (a) An effect is independent of its indirect causes conditional on its direct causes, and (b) Variables are independent conditional on their common causes. This implication of the Markov condition provides a meaningful causal interpretation for a certain dependency pattern, which is captured in the first part of the d-separation criteria. For the two types of causal structures of the causal chains ($A \to C \to B$ or $A \leftarrow C \leftarrow B$) and the causal fork ($A \leftarrow C \to B$), the two extreme variables A and B, which are unconditionally dependent, become independent once we conditioning on the middle variable C by the Markov condition. The Markov condition can be intuitively understood as a generalization of the Markov property, which is originated from probability theory, by expanding the concepts of the past, current, and future states. According to probability theory, a stochastic process has a Markov property if the conditional probability distribution of future states of the process depends only upon the current state and not on any past states. Only the current state gives information relevant to the future behavior of the process. Knowledge of the history or path of the process does not add any new information. So given the current state, the future state is conditionally independent of any of the past states. The above idea is captured in the first part of d-separation criteria, which states that for causal chain $A \to C \to B$ or $A \leftarrow C \leftarrow B$ and the causal fork $A \leftarrow C \to B$, the middle variable C should be in the Sepset S_{AB}, because the two extreme variables A and B, which are unconditionally dependent, become independent once we conditioning on the middle variable C. Note that this criterion of the Sepset in the DAGs or Bayesian networks is common to criterion of the Cutset in the undirected graphs or Markov networks.

Given the common graphical separation criterion for both the undirected graph and the directed graph, the unique separation criterion is the second condition of the d-separate criterion in DAG, which provides the "observational clue" for the causal directionality. It based on the following phenomenon known as the Berkson's paradox or selection bias in the statistical literature (Berkson 1946) and the explaining away effect in the artificial intelligence (Kim and Pearl 1983). Observation on a common consequence of (unconditionally) independent causes tends to make those causes dependent, because information about one of the causes tends to make the other more or less likely, given that the consequence is observed. So when it is found that the three variables exhibit intransitive pattern of dependencies such that (a) The variables A and B are each correlated with a third variable C but are independent of each other ($A - C - B$) and (b) The two extreme variables A and B, which are unconditionally independent, become dependent once we conditioning on the middle variable C, the only meaningful interpretation in terms of causal directionality is the middle variable C is the common effect of A and B (unshielded-collider, $A \to C \leftarrow B$). Intuitively this interpretation of intransitive triples involves a virtual control of the effect variable, whereas the randomized experiment involves an actual manipulation of the putative causes. That is, if we can find another means B of potentially controlling C without affecting A, we preclude C from being a cause of A. For example, one of the reasons people insist that rain (A) causes wet grass (C) and not the other way around is that they can easily find other means such as sprinkler (B) that are totally independent of the rain (A) to getting the wet grass (C) (Pearl, 2000). The above idea is captured in the second part of d-separation criteria, which states that for the causal inverted fork $A \rightarrow C \leftarrow B$, the middle variable C or any of its descendants should not be in the Sepset S_{AB}, because the two extreme variables A and B , which are unconditionally independent, become dependent once we conditioning on the middle variable C .

The causal structure is generally underdetermined by the statistical properties of the data (induction problem). A simple but fundamental version of the induction problem is that correlation does not imply causation. This induction problem, however, can be partially addressed by the full use of the maximum information of unconditional and conditional probabilistic structures of non-temporal and non-experimental data. Under certain conditions, the combinational information of unconditional and conditional independencies among all the possible pairs of variables provides “empirical clues” (a) to discriminate the true statistical relationships from spurious correlations without causal orientations and (b) to discriminate the unshielded-collider structure from the observational equivalent causal structures of causal chain and fork. While correlation does not imply causation in general, no causation does imply no correlation under the stability and Markov conditions. This proposition of no correlation without causation can be understood as follows: (a1) The stability condition implies that if two variables are statistically independent, then neither variable is a direct cause of the other. (a2) The Markov condition implies that if a pair of variables is statistically dependent, then one of the variables is a direct cause of the other. Note that the sufficiency condition embedded in the Markov condition allows discriminating the spurious correlation induced by the common cause. On the other hand, the stability condition, with the Markov condition, makes it possible to discriminate the possible unstable existence or nonexistence of spurious correlation, which is possibly induced by the numerical parameter values. Fundamentally this proposition allows for the possibility of an inductive inference of causal structures from the statistical observations. When three variables exhibit intransitive pattern of dependencies ($A - C - B$) such that (i) there exist non-spurious correlations between A and C and between B and C . (ii) A and B are independent, which is not induced by the numerical parameter values, there are following two possibilities: (b1) The two extreme variables A and B , which are unconditionally dependent, become independent once we conditioning on the middle variable C . (b2) The two extreme variables A and B, which become dependent once we conditioning on the middle variable C, are independent without conditioning any subset of variables. The first probabilistic structure, which is commonly implied by both causal chain ($A \rightarrow C \rightarrow B$ or $A \leftarrow C \leftarrow B$) and causal fork ($A \leftarrow C \rightarrow B$), provides a causal interpretation for the simple but fundamental version of induction problem that correlation does not imply causation. On the other hand, the second probabilistic structure, which is implied by the unshielded-collider ($A \rightarrow C \leftarrow B$), makes it possible to inductively infer C as the common effect of A and B. Note that this type of causal orientation is the only possible (truly inductive) causal inference based on the statistical observations. This is the reason why a third variable is needed to decide the causal direction between two variables.

The observed equivalence between causal chain and causal fork can not be discriminated based only on statistical observations without using non-observational extra causal information or manipulative (randomized) experimentation. However, the graph theory provides "logical clues" to partially address the observational equivalence problem. After the maximum information of unconditional and conditional probabilistic structures from data is obtained, (a) All the discriminative information between the true statistical relationships and spurious correlations among variables without causal orientations are summarized into the graph with undirected edges, named as the skeleton, and (b) All the discriminative information of the unshielded-collider structure from the observational equivalent causal structures of causal chain and fork are summarized into the partially oriented graph, named as the partially directed acyclic graph (PDAG) with causal orientations from independent causes to the common effect. By logically deciding causal directions for the remaining undirected edges in PDAG, the completed partially directed acyclic graph (completed PDAG or essential graph), which is maximally oriented PDAG, can be further inferred. The logical inferences about causal directions are based on the idea that orienting the remaining undirected edges in PDAG does not result in the causal structure which is inconsistent with the statistical observations, as long as the logically decided orientations do not create either new unshielded-collider structure or a cyclic causal structure. It is mathematically demonstrated that the following four rules are the maximally possible logical orientation rules for the remaining undirected edges in the partially directed acyclic graph (PDAG) (see Verma and Pearl, 1992, Meek, 1995a, and Pearl, 2000). (Rule 1) Orient $A \rightarrow B$ for the remaining undirected edges $A - B$ in PDAG, whenever there is an arrow $C \rightarrow A$ and that C and B are not adjacent. Rule 1 is based on the fact that the orientation $A \leftarrow B$ would create an empirically unsupported new unshielded-collider at A. (Rule 2) Orient $A \rightarrow B$ for the remaining undirected edges $A - B$ in PDAG, whenever there is a causal chain $A \rightarrow C \rightarrow B$. Rule 2 is based on the fact that the orientation $A \leftarrow B$ would create directed cyclic pattern which is impossible by the acyclic assumption. Note that rule 2 creates a collider at B but it is a shielded-collder not an unshielded-collider. So this rule does not result in an inconsistency with the statistical observations. (Rule 3) Orient $A \rightarrow B$ for the remaining undirected edges $A - B$ in PDAG, whenever there are two chains $A - C \rightarrow B$ and $A - D \rightarrow B$ and that C and D are not adjacent. (Rule 4) Orient $A \rightarrow B$ for the remaining undirected edges $A - B$ in PDAG, whenever there are two chains $A - C \rightarrow D$ and $C \rightarrow D \rightarrow B$ and that A and D are adjacent but B and C are not adjacent. Rules 3 and 4 is based on the fact that the orientation $A \leftarrow B$, by two applications of rule 2, would create empirically unsupported new unshielded-collider at A ($C \rightarrow A \leftarrow D$ for rule 3 and $B \rightarrow A \leftarrow C$ for rule 4). These four rules are illustrated by Figure 2.1.

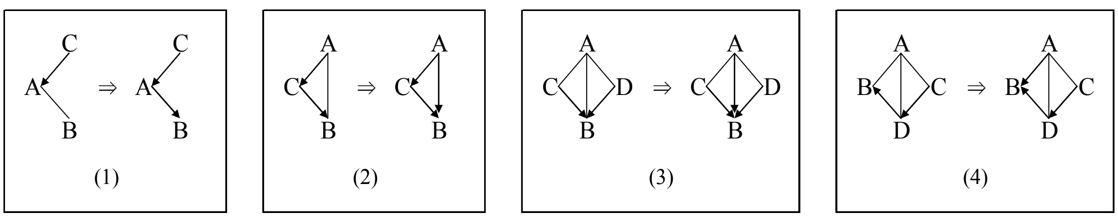


Figure 2.1. Logical Orientation Rules for Undirected Edges in PDAG

The graph theory, not only provides logical orientation rules to partially discriminate observational equivalent causal structures, but also allows the full use of the maximum information of unconditional and conditional probabilistic structures from data. Checking or searching all the relevant (un)conditional probabilistic structures among all the possible pairs of variables with respect to all possible combinations of other variables as the Sepset becomes feasible only by systematically and efficiently defining the relevant or entire search space, which consists of all possible causal hypotheses represented by DAGs. In graph theory, the relationships, which are used to relate the probabilistic dependency model M_P and the graphical dependency model M_G, are also used to define some relationships between two graphical dependency models. Two graphical dependency models of DAGs are perfect-map or observational equivalence for each other iff they have the same skeleton and the same unshieleded-colliders (Verma and Pearl 1990). This observational equivalence, which places a limit on the ability of the statistical approach to infer causal structure, provides logical background to systematically classify the search space by eliminating the problem of multiple searching for the statistically equivalent DAGs. The independent-map relationships are then used to efficiently connect the systematically classified search spaces or the equivalence classes of DAGs. The independent-map relationships relate each other by the natural relationship of whether one equivalent class E specifies more restrictions than the other E'. In particular, when one equivalent class E' is an independent-map of the other E, E imposes more independence constraint than E' and thus E' contains more edges than does E. Based on this fact, the whole search space can be systematically organized by a sequence of Independent-map relations between each equivalent class $E_o, E_1, \cdots, E_I$ such that E_i is an Independent-map of E_{i+1} and there is only one edge difference between them. Note that E_0 is a completely connected graph so is a trivial I-map of all DAGs and E_I is a completely disconnected graph so is a trivial D-map of all DAGs.

This idea can be illustrated for the three variable case by using the following Figure 2.2., which is adopted from Kocka et al. (2001) with some modifications. All the possible causal hypotheses except cyclic ones are represented by the DAGs. Each box represents an equivalence class of DAGs. For example, the equivalence of DAG (8)-(10) can be illustrated by applying Bayes's rule for factorization based on the DAG as well as the specified common conditional independence/dependence pattern. The joint distribution $P(A,B,C)$ can be factorized as follows: $P(C \mid A)P(A \mid B)P(B)$ for DAG (8), $P(B \mid A)P(C \mid A)P(A)$ for DAG (9) and $P(B \mid A)P(A \mid C)P(C)$ for DAG (10). The relationship $P(A \mid B)P(B) = P(A,B) = P(B \mid A)P(A)$ makes the two DAGs (8) and (9) equivalent and the relationship $P(C \mid A)P(A) = P(A,C) = P(A \mid C)P(C)$ makes the two DAGs (9) and (10) equivalent. So the DAGs (8)-(10) are equivalent in terms of factorization of joint distributions. Under the Gaussian and multinomial distributions, this independence equivalence become identical to distributional equivalence, which means that equivalence class of DAGs have the same probability distribution. Connections among boxes represent the sequence of independent-map relationships. For example, when DAG (8)-(10) are represented by equivalence class E_i, DAGs (1)-(6) are represented by equivalence class E_{i-1} and the union of two equivalence classes of DAGs (19)-(20) and DAG (21)-(22) is represented by equivalence

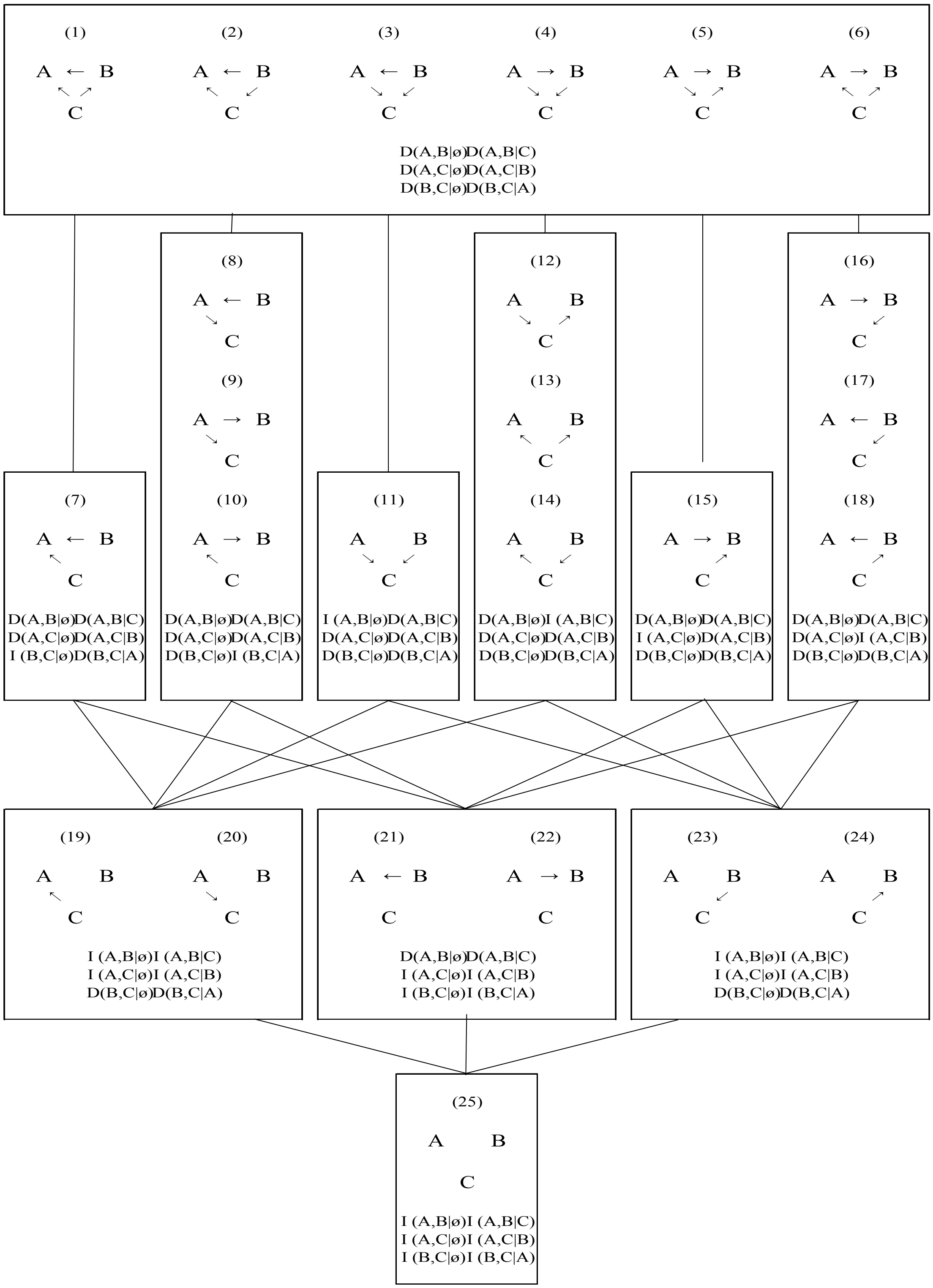


Figure 2.2. Search Space Defined by the Graph Theory

class E_{i+1}. Note that DAG (1)-(6) can be represented by equivalence class E_0 which is a trivial I-map of all DAGs and DAG (25) can be represented by E_I which is a trivial D-map of all DAGs. Note also that it is possible to travel or to search all equivalence classes of DAGs by a specific sequence of single edge modifications along these connections.

Many computer algorithms have been suggested to implement the logic of the graphical causal models for empirical studies. These algorithms can be classified as two types of approaches according to the two distinctive ways of the statistical observation, where the observation is done through the statistical decisions based on either the Neyman-Pearson type statistical test (conditional independence test approach) or the Bayesian information criterion (goodness-of-fit scoring approach). The first conditional independence test approach is based on the qualitative information about whether or not a particular individual local conditional independence constraint holds. On the other hand, the second goodness-of-fit scoring approach is based on the quantitative measure of how much the global independency patterns associated with an entire causal structure explain the data.

The conditional independence test approach starts by searching for a Sepset S_{AB} in all possible subsets of V\{A,B} such that $I(A,B|S_{AB})$ holds for each pair of variables A and B by applying local conditional independence tests on A and B conditional on S_{AB}. The categorical or qualitative decisions of such local tests are used to reconstruct topologies of the underlying DAG and to decide orientations based on the pattern of unshieleded-colliders. By using logical orientation rules, the partially directed acyclic graph (PDAG) is transformed into the completed partially directed acyclic graph (completed PDAG), which can be either a particular DAG or equivalent set of DAGs. The main task for this approach is to deal with the complexity and reliability problems in searching for the possible Sepsets. As this approach searches among all possible subsets in V\{A,B}, it involves a growing number of higher-order independence tests. As the number of variables increases, all the possible subsets rapidly increase, so the algorithm can become infeasible even when searching for the sparse true graphs. Furthermore, higher order conditional independence tests are generally less reliable than lower order independence tests (Spirtes et al., 2000).

There are several algorithms suggested to deal with this task. Among them, PC algorithm is used in this study, because it provides an efficient and reliable way of searching for Sepsets S_{AB}. The PC algorithm, named after its authors of Peter and Clark, is discussed in Spirtes et al. (2000). PC algorithm commences by forming a completely connected undirected graph. It then searches for the Sepsets S_{AB} of cardinality 0, then cardinality 1, and so on. The search for a Sepset S_{AB} is limited to variables that are adjacent to A and B at every stage. Edges are recursively removed from a complete graph as conditional independence is found. By this way, PC algorithm bounds the number of independence tests as $N^2(N-1)^{K-1}/(K-1)!$, where N is the number of variables and K is the highest number of adjacent variables in the graph. PC algorithm uses Neyman-Pearson type statistical tests of partial correlation for conditional independence test by assuming linear Gaussian distributions.

The goodness-of-fit scoring approach starts by logically defining the search space which consists of all possible causal hypotheses represented by DAGs. It then searches the DAG that best explains the data, where the explanation power of a given DAG at each search step is scored and compared by a goodness-of-fit measure. The main difficulty for this approach is that the number of possible hypothetic causal structures of DAGs rapidly increases as the number of variables N increases. It is demonstrated that the number of different DAG structures $r(N)$ is given by the recurrence formula $r(N)=\sum_{i=1}^{I}(-1)^{i+1}\binom{N}{i}2^{i(N-i)}r(N-i)$ (Robinson, 1977). This formula, for examples, gives $r(2)=3$, $r(3)=25$, $r(4)=543$, and $r(5)=29281$ as the number of possible DAGs for the number of variables 2, 3, 4, and 5 respectively. As the number of variables increases, all the possible DAGs rapidly increase so the algorithm can become infeasible even when searching for the sparse true graphs. This complexity problem suggests that it is needed to systematically represent the whole search space and to efficiently generate and evaluate neighbors for a particular state in the search.

There are several algorithms suggested to deal with this task. Among them, the two-phase Greedy Equivalence Search (GES) algorithm is used in this study, because it provides an efficient and optimal search algorithm. The GES algorithm is originated from Meek (1997) and its optimality is proved by Chickering (2002). Algorithmic logics are based on the results of graphical theory as follows: (a) The two-phase Greedy Equivalence Search (GES) algorithm greedily moves to equivalent classes of DAG as neighbors until it reaches the local maximum at each of the two phases of search procedure. This algorithmic logic relies on the result of graph theory that the whole search space can be systematically represented by the equivalence classes of DAG. (b) The two-phase GES algorithm restricts the neighbors of particular state of equivalent classes of DAG E_i as either E_{i-1} for first single edge addition phase or E_{i+1} for second single edge removal phase. This algorithmic logic relies on the result of graph theory that the whole search space can be efficiently searchable along the natural connections by the sequence of independent-map relations among equivalent classes $E_o, E_1, \cdots, E_I$ such that E_i is an independent-map of E_{i+1} and there is only one edge number difference between E_i and E_{i+1}. Note that when the algorithm considers the edge addition or removal, it also checks for the possible unshielded-colliders. For example using the above figure of three variable case, the current state E_{I-1} which consists of DAGs (19)-(20) is compared with four neighbors of E_{I-2} which consist of DAG (7) and DAG (11) as the possible unshielded-collider patterns as well as DAGs (8)-(10) and DAGs (12)-(14) in the first edge addition phase.

GES algorithm uses the Bayesian Information Criterion (BIC) as a measure of scoring goodness-fit of a given DAG G at each step of the search. The BIC is chosen as a goodness-fit score because (a) It is a consistent approximation of the Bayesian posterior probability under the Gaussian and multinomial distributions and (b) It has decomposability and equivalence properties that allow efficient scoring. BIC for a given DAG G of a set of variables $V = \{X_1, \cdots, X_N\}$ can be written as follows: $BIC(V,G) = \log P(V \mid G) - \dim(G) \cdot \log(T/2)$, where T is the sample size and $\dim(G)$ is the dimension or the number of parameters of DAG G and $\log P(V \mid G)$ is the log-likelihood function for a set of variables V given DAG G. For a given DAG G at each step of the search procedures, $\dim(G)$ is calculated by counting the number of edges in G and $\log P(V \mid G)$ is calculated by $\log P(X_1, \ldots, X_N) = \sum_n \log P(X_n \mid Pa_n)$, which has decomposable property and thus can be efficiently evaluated. Because the scoring function BIC is based on the factorization of the joint distribution by the DAG $P(X_1, \ldots, X_N) = \prod_n P(X_n \mid Pa_n)$ but the equivalence class of DAGs or the partially directed acyclic graph (PDAG) is used to represent each state, the PDAG is transformed into the completed partially directed acyclic graph (completed PDAG or essential graph) by using logical orientation rules at each step of the search procedures,. The property of equivalent BIC scores for members of an equivalence class comes from the fact that DAGs in an equivalence class have the same number of edges and a common factorization. Note that the BIC measure involves too many parameters for a completely connected graph, so the GES algorithm usually uses the completely unconnected graph as its initial PDAG. But it is possible to start the search with another PDAG based on other causal information such as theory and/or the completed PDAG, which can be resulted from the PC algorithm (Spirtes and Meek, 1995).

Two distinctive approaches to infer causal structures among variables represented by DAGs can be compared with respect to several aspects. Several other algorithms and their characteristics are discussed in Sangüesa and Cortés (1997). Among them, one comparison has an interesting feature in terms of using the logical orientation rule. In the PC algorithm, the logical orientation rule is used only after all the possible statistical information from data is obtained. On the other hand, the logical orientation rule is used at every step of the search procedures in the GES algorithm. This implies that separating the logical extension rule from the algorithms is relatively easy in the PC algorithm but relatively difficult in the GES algorithm. This different feature of two algorithms has implications for the purpose of relaxing the acyclic and sufficiency assumptions, given that the logical orientation rule relies on the Markov condition, which is based on the acyclic and sufficiency assumptions. In fact, the conditional independence test approach makes some progress for relaxing the acyclic or sufficiency assumptions. In particular, based on the PC algorithm, Richardson and Spirtes (1999) develop Cyclic Causal Discovery (CCD) algorithm to allow cyclic possibility and Spirtes et al. (2000) develop Fast Causal Inference (FCI) algorithm to relax sufficiency condition. These developments are not incorporated in this study, since it is still ambiguous how to distinguish between feedback and latent phenomena (Moneta and Spirtes, 2006). We hope that it is not too harmful to assume the acyclic and sufficiency conditions, given the observation that these two conditions are implicitly or explicitly assumed in most empirical studies in economics.

The other comparison has practical implications. The conditional independence test approach is based on the qualitative decision about local independence tests, so it is susceptible to incorrect qualitative local decisions, which can be sensitive to the chosen significant level. Based on the simulation results, it is recommended to systematically lower the significance level as the sample size increases. For example, 0.2 for the sample size less than 100 and 0.1 for the sample size between 100 and 300 are recommended as the significance level for local independence tests (Spirtes et al., 2000). However, it is still not easy to decide the appropriate significance level for the local tests, because the power of algorithm against alternatives is an extremely complex and unknown function of the power of the individual local test. The goodness-of-fit scoring approach does not require choosing a specific significance level, because it is based on the quantitative measure about how much the overall independence constraints associated with an entire causal structure are true. In this respect, it allows users to make finer distinctions among alternative causal structures or to combining them to better inferences by the model averaging process based on the quantitative measure such as BIC in GES algorithm.

The overall graphical causal models or DAG approaches can be also compared with the traditional structural equation model (SEM) approaches. To infer causal relationship between two variables A and B, the DAG use the criterion whether a third variable C exhibits a specific pattern of dependency with A and B. In this respect, the DAG approach can be compared with the SEM approach, where the simultaneous relationships of the j th endogenous variable (A) and other endogenous variables included in the j th equation (B) are discriminated (identification or induction problem) by the assumed exogenous variables (C) excluded from the j th equation as the additional third causal determinants or specific shifters for behavioral equations of other endogenous variables included in the j th equation. However, methods to address this induction problem are quite different.

In the SEM approach, the selection of exogenous variables is usually considered as maintained assumptions derived from the theory rather than something to be learned form data itself. Even when the hypothetical test approach is implemented based on regression framework, (a) The non-nested hypothetical test approaches oftentimes have the power problem related with the statistical hypotheses test, so that they have generally little power to discriminate competing specifications. (b) The nesting hypothetical test approaches based on variable selection methods also faces following issues: (b1) When the small explanatory variable set is initially assumed and then subsequently expanded into larger selected variable set (bottom-up approach), the omitted variable (especially common cause variable) problem in initial (or subsequent) small model can mislead the testing results. For example, if true causal structure is $y_t \leftarrow W_t \rightarrow x_t^1$ but the initial small model $y_t = a_1 x_t^1 + \varepsilon_t$ omits the common cause variable W_t, then hypothetic test of $H_0 : a_1 = 0$ can be rejected. (b2) When the large explanatory variable set is initially assumed and then subsequently reduced into smaller selected variable set (top-down approach), the included variable (especially common effect variable) problem in initial (or subsequent) large model can mislead the testing results. For example, if true causal structure is $y_t \rightarrow W_t \leftarrow x_t^1$ but the initial large model $y_t = a_1 x_t^1 + \beta W_t + \sum a_k x_t^k + \varepsilon_t$ includes the common effect variable W_t, then hypothetic test of $H_0 : a_1 = 0$ can be rejected. Note that these problems can arise, even though the causal sufficiency condition is assumed.

In the DAG approach, on the other hand, all the unconditional and conditional probabilistic structures among all the relevant combinations of variables are efficiently checked in search procedures to obtain the maximum information of specific pattern of dependencies among variables from data, where relevant search spaces are logically decided based on the graph theory. Note that checking or searching all the relevant (un)conditional probabilistic structures among all the possible combinations of variables becomes infeasible without systematically and efficiently defining the relevant or entire search space. The graph theory also provides logical orientation rules to discriminate observational equivalent causal structures, which can not be discriminated based on statistical properties only, without using non-observational extra causal information or manipulative (randomized) experimentation.

Graphical causal models or DAG approaches can be used for the empirical studies. Both PC and GES algorithms are implemented in Tetrad IV program. However, there are some caveats for their use in data analysis especially for the high dimensional data set. The graphical causal models are based on the Markov and stability conditions. Although the Markov condition is commonly assumed for most empirical studies and thus can be accepted, the Markov condition only makes the graphical dependency model as an independent-map of the probabilistic dependency model. This means that the underlying causal structure implies the probabilistic dependency pattern. On the other hand, the inductive inference of causal structure from the data is possible only when the probabilistic dependency model implies the underlying causal structure. In this respect, the stability condition needs to be further discussed to use the graphical causal models or DAG approach in empirical study, since the stability condition, with the Markov condition, makes a DAG as a perfect-map of (or equivalent to) a statistical dependency pattern. Recall that the stability condition implies that all the unconditional and conditional probabilistic structures are stable with respect to changes in their numerical values. This stability has following implications: (a) All the observed (un)conditional probabilistic structures are due to the underlying causal structures, not their special numerical values. (b) No spurious independence in probability dependency model can be destroyed or induced by changing probabilistic parameter values. (c) It is possible to effectively and efficiently encode (un)conditional (in)dependencies structures into graphical dependency model without numerical probabilities. Thus, with the Markov condition, (d) It is possible to infer the underlying causal structures from the observed marginal and conditional probabilistic structures, where the observation is done through the statistical decisions based on either the Neyman-Pearson type statistical test (conditional independence test approach) or the Bayesian information criterion (goodness-of-fit scoring approach).

There can be two circumstances where the stability condition can be violated, as discussed in the Tetrad II manual. One possible circumstance is that there may exist strict equality among products of parameters, so that a spurious independence in probability distribution can be destroyed or induced by changing underlying parameter values. For example, in the linear modeling of causal structure of $A = \lambda_1 B + \lambda_2 C + u_A$ and $C = \lambda_3 B + u_C$, the restriction of $\lambda_1 = -\lambda_2 \cdot \lambda_3$ can numerically induces independence between A and B, even if a structural dependence exists between A and B. It has been demonstrated that for the Gaussian distribution (Pearl and Verma, 1991) and multinomial distribution (Meek, 1995b), the strict equalities among products of parameters have very little possibility or Lebesgue measure of zero in any probability space in which parameters vary independently of one another. Note that parameters vary independently of one another under the modular situation, where a complex system can be built by combining simpler local parts and it is possible to manipulate one such relationship without changing the others.

The other possible circumstance is that there may exist deterministic or near deterministic relationships among variables so that any the statistically observed (un)conditional probabilistic structures are due to not only the underlying causal structures but also their special numerical values. According to Tetrad II manual, the Tetrad program should not be used for the following cases or these second cases should be practically addressed in empirical study, where (a) There are deterministic relationships among variables or (b) There are conditional probabilities very close to 1 in the discrete case or (c) There are correlations very close to 1 in the linear case. These restrictions for using the Tetrad program can be understood based on the following reasoning. If $P(A \mid B) \approx 1$, then $P(A \mid B, C) = P(A \mid B)$ can be hold for any set of variable C, regardless of the causal structures among them. So it is not possible to infer reliable causal structure from the probabilistic dependency pattern. For example from Tetrad II manual, when there are four variables A, B, C, and D such that (i) A, B, and C are independent each other. (ii) D is the common cause of A, B, and C, the near deterministic relationship between C and D such as $P(D \mid C) \approx 1$ can numerically induce independence between A and B by conditioning on C, instead of conditioning on D. For another example from empirical study, even for the same commodity, any causal relationships between price p_1 and quantity q_1 can be statistically broken, when another related commodity's price p_2 has a high co-movement with p_1. It is because high correlation between p_1 and p_2 can induce $P(p_1 \mid p_2, q_1) = P(p_1 \mid p_2)$ through $P(p_1 \mid p_2) \approx 1$. Note that this problem is similar to the multicollinearity problem, which makes it difficult to obtain precise estimates of the separate effects of the variables in regression methods. Given the observation that many variables in a high dimensional data set oftentimes move very closely, the use of the graphical causal model for the high dimensional data set can be problematic, since the stability condition can be violated in its applications for high dimensional data sets. One possible way to address this problematic situation is to use aggregation method. However, before using aggregation method, the legitimate aggregation condition should be empirically identified to consistently infer causal structures among disaggregated variables by using the aggregated variables as the legitimate representatives. This issue is closely related with the next topic to be discussed.

Aggregation Theory

Theil's aggregation theory is concerned with the transformation of individual relations (micro-relations) to a relation for the group as a whole (macro-relations) (Theil, 1971). It considers the possibility that micro-relations can be studied through the macro-relations, where micro-variables are grouped and represented by macro-variables. The main issue is to understand the general relationship between micro-parameters and macro-parameters. The ultimate goal is to identify conditions for the meaningful aggregation that makes it possible to represent micro-relations by macro-relations. Theil (1954) shows that macro-parameters generally depend upon complicated combinations of corresponding and non-corresponding micro-parameters. He, however, also identifies two special conditions for the possibility of meaningful aggregation, which are the micro-homogeneity and the compositional stability conditions. While the micro-homogeneity condition means that each of the micro-parameters is equal across all individual units, the compositional stability means that the ratios of micro-variables over macro-variables are constant over time (Monteforte, 2004). If one of these conditions is satisfied, then the aggregated macro-model is considered as a legitimate representative of the underlying disaggregated micro-model.

Theil's aggregation theory can be understood as follows. For a given T time period, each individual unit has its own linear behavioral relationship. That is, for each individual micro-unit ($n=1,.....,N$), an endogenous variable y_n linearly depends on K exogenous variables $x_n = [\ x_{1n},.....,x_{Kn}\ ]$ with corresponding micro-parameters $\beta_n = [\ \beta_{1n},.....,\beta_{Kn}\ ]'$. These relationships can be represented by following set of micro-equations.

(1) $\quad y_n = x_n \beta_n + u_n \qquad , \forall n = 1,.....,N$.

To study the general tendency of phenomena which are common to most of all $n=1,.....,N$ individual micro-unit behaviors, it is postulated that the relation between the aggregated dependent variable Y and aggregated predetermined variables $X = [\ X_1,.....,X_K]$ can be represented in the same linear form of micro-equations as the following macro-equation (2). And macro-parameters $\beta = [\ \beta_1,.....,\beta_K]'$ are estimated by the least-squares estimation method (4).

(2) $\quad Y = X\beta + U$ where

(3) $\quad Y \equiv \sum_{n=1}^{N} y_n$ and $X \equiv \sum_{n=1}^{N} x_n$.

(4) $\quad \hat{\beta} = (X'X)^{-1} X'Y$.

Theil studies this estimator's properties, especially in the context of the relationship between macro-parameters and micro-parameters. When micro-variables are represented by macro-variables through aggregation functions (3), the correct specification of the aggregated relation becomes following equation (5).

(5) $$Y\left(\equiv \sum_{n=1}^{N} y_n\right) = \sum_{n=1}^{N} x_n \beta_n + \sum_{n=1}^{N} u_n \ .$$

Note that the true aggregated equation (5) has the $K \cdot N$ explanatory variables, so it contains as detailed information as a set of individual micro-relations as a whole, except the loss of information due to using aggregated dependent variable. Note also that the aggregation function (3) defined as the simple sum can be generalized to the weighted average as (3') $Y' \equiv \sum W_n^y y_n$ and $X' \equiv \sum W_n^x x_n$. When the weighted average is used, the true aggregated relation can be written as follows $(5')$.

(5') $$Y'\left(\equiv \sum_{n=1}^{N} W_n^y y_n\right) = \sum_{n=1}^{N} W_n^y \left(x_n \beta_n + u_n\right) = \sum_{n=1}^{N} \left(W_n^x x_n\right)\left(\frac{W_n^y}{W_n^x}\beta_n\right) + \sum_{n=1}^{N}\left(W_n^y u_n\right) = \sum_{n=1}^{N} x_n' \beta_n' + \sum_{n=1}^{N} u_n' \ .$$

Since equation $(5')$ is fully equivalent to (5), the following discussion can be applied, mutatis mutandis, to the macro-parameters in the macro-equation by using (2') $Y' = X'\beta' + U'$ (Theil, 1954). Especially equation (5') is equivalent to (5), when we use the same weighting schemes for Y and X by $W_n^y = W_n^x$. Theil defines linear aggregation of economic relations as simple summation, simple average, and fixed weights average aggregations. The micro-homogeneity condition can be understood immediately as follows. When all micro-parameters are equal across all individual units, we can write $\beta_n = \beta$, $\forall n = 1,\ldots\ldots, N$ in the set of micro-equations (1). This implies that the macro-equation has a natural meaning such that all macro-parameters are equivalent to the common micro-parameters, because the true aggregate relation becomes $(5'')$.

(5'') $Y\left(\equiv \sum_{n=1}^{N} y_n\right) = \sum_{n=1}^{N} x_n\ \beta + \sum_{n=1}^{N} u_n = X\beta + U$, by assumption $\beta_n = \beta \ \ \forall n = 1,\ldots\ldots, N$.

This micro-homogeneity condition, however, might be a too restrictive condition to use for practical purposes, because it requires the complete knowledge of all micro-parameters. In this respect, we do not assume any restrictions on micro-parameters for each individual micro-unit in this study.

Using the true aggregated equation (5), the macro-parameter estimator can be written as follows.

(6) $\hat{\beta} = (X'X)^{-1}X'Y$, by the true aggregation (5) $Y = \sum_{n=1}^{N} x_n\ \beta_n + \sum_{n=1}^{N} u_n$

$$= \sum_{n=1}^{N} (X'X)^{-1}X'x_n\beta_n + (X'X)^{-1}X'\sum_{n=1}^{N} u_n \ .$$

To interpret this result, Theil postulates the following set of auxiliary equations (7), where exogenous micro-variables x_n are assumed to be linearly related with macro-variables X. When we assume that auxiliary-disturbances v_n are independent of exogenous macro-variables X and they have zero means, we can consistently estimate the coefficient A_n by the least-squares method (8). Note that in this study, $Cov(\ ) = 0$ is used to represent an independent relation, which is equivalent to no correlation under normal distribution with linearity.

(7) $x_n = \ X\ A_n + v_n$

$$\text{or } [\ x_{1n}, x_{2,n}, \cdots, x_{Kn}] = [\ X_1, X_2, \cdots, X_K\ ]\begin{bmatrix} a_{11,n} & a_{12,n} & \cdots & a_{1K,n} \\ a_{21,n} & a_{22,n} & \cdots & a_{2K,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1,n} & a_{K2,n} & \cdots & a_{KK,n} \end{bmatrix} + [\ v_{1,n}, \cdots, v_{K,n}] \ \ , \forall n = 1,\ldots\ldots, N$$

or $x_{k\,n} = \left(\sum_{j=1}^{K} X_j\, a_{j\,k,n} \right) + v_{k,n} = X_k a_{k\,k,n} + \left(\sum_{j \neq k}^{K} X_j\, a_{j\,k,n} \right) + v_{k,n}$ $\qquad$, $\forall k = 1,\ldots\ldots,K$, $\forall n = 1,\ldots\ldots,N$.

(8) $\hat{A}_n = (X'X)^{-1} X' x_n$ $\qquad$, where

$E\left(\hat{A}_n\right) = A_n + E\left[(X'X)^{-1} X' v_n\right]$ $\qquad$, by assumptions of $Cov(v_n, X) = 0$ and $E(v_n) = 0$

$= A_n$ $\qquad$, $\forall n = 1,\ldots\ldots,N$.

Note that equations (3) and (7) imply that the sum of coefficients becomes a $K \cdot K$ unit matrix and the sum of disturbances becomes a $T \cdot K$ zero matrix for the set of auxiliary equations. Because the coefficient A_n sums to 1 across micro-units, it can be used as the weighting scheme.

(10) $\sum_{n=1}^{N} A_n = I_{(K \times K)}$ and $\sum_{n=1}^{N} v_n = 0_{(T \times K)}$ $\qquad$, because $X \left(\equiv \sum_{n=1}^{N} x_n \right) = \sum_{n=1}^{N} \left(X\, A_n + v_n \right) = X \sum_{n=1}^{N} A_n + \sum_{n=1}^{N} v_n$.

Using the result (9) as well as the assumption of the correct specification of micro-equations, $y_n = x_n\, \beta_n + u_n$, which implies that micro-disturbances u_n are independent with exogenous macro-variables X and have zero means, Theil interprets the macro-parameter estimator $\hat{\beta}$ as a consistent estimator for $\sum A_n\, \beta_n$ as in (12).

(11) $\hat{\beta} = \sum_{n=1}^{N} (X'X)^{-1} X' x_n \beta_n + (X'X)^{-1} X' \sum_{n=1}^{N} u_n$ $\qquad$, by using result (8) $\hat{A}_n = (X'X)^{-1} X' x_n$

$= \sum_{n=1}^{N} \hat{A}_n \beta_n + (X'X)^{-1} X' \sum_{n=1}^{N} u_n$.

(12) $E\left(\hat{\beta}\right) = \sum_{n=1}^{N} E\left[\hat{A}_n\right] \beta_n + E\left[(X'X)^{-1} X' \sum_{n=1}^{N} u_n \right]$ $\qquad$, by using result (9) $E\left(\hat{A}_n\right) = A_n$

$= \sum_{n=1}^{N} A_n\, \beta_n + E\left[(X'X)^{-1} X' \sum_{n=1}^{N} u_n \right]$ $\qquad$, by assumptions of $Cov(u_n, X) = 0$ and $E(u_n) = 0$

$= \sum_{n=1}^{N} A_n\, \beta_n$.

Defining macro-parameters as mathematical expectation of its least-squares estimator, Theil (1954) concludes that macro-parameters generally depend upon complicated combinations of corresponding and non-corresponding micro-parameters as in (13). He then further decomposes corresponding micro-parameters into simple sum (if $c = 1$) or simple average (if $c = 1/N$) of corresponding micro-parameters and a deviation term from it. He labels the sum of this deviation term and the non-corresponding micro-parameters as the aggregation bias as in (14).

$$(13)\quad E\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_K\end{bmatrix} = \sum_{n=1}^{N}\begin{bmatrix}a_{11,n} & a_{12,n} & \cdots & a_{1K,n}\\ a_{21,n} & a_{22,n} & \cdots & a_{2K,n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{K1,n} & a_{K2,n} & \cdots & a_{KK,n}\end{bmatrix}\begin{bmatrix}\beta_{1,n}\\ \beta_{2,n}\\ \vdots\\ \beta_{K,n}\end{bmatrix} = \begin{bmatrix}\sum_{n=1}^{N} a_{11,n}\,\beta_{1,n}\\ \sum_{n=1}^{N} a_{22,n}\,\beta_{1,n}\\ \vdots\\ \sum_{n=1}^{N} a_{KK,n}\beta_{K,n}\end{bmatrix} + \begin{bmatrix}\sum_{n=1}^{N}\sum_{j\neq 1}^{K} a_{1j,n}\,\beta_{j,n}\\ \sum_{n=1}^{N}\sum_{j\neq 2}^{K} a_{2j,n}\,\beta_{j,n}\\ \vdots\\ \sum_{n=1}^{N}\sum_{j\neq K}^{K} a_{Kj,n}\,\beta_{j,n}\end{bmatrix}$$

$$= \sum_{n=1}^{N}\left(\begin{bmatrix}a_{11,n} & 0 & \cdots & 0\\ 0 & a_{22,n} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & a_{KK,n}\end{bmatrix} + \begin{bmatrix}0 & a_{12,n} & \cdots & a_{1K,n}\\ a_{21,n} & 0 & \cdots & a_{2K,n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{K1,n} & a_{K2,n} & \cdots & 0\end{bmatrix}\right)\begin{bmatrix}\beta_{1,n}\\ \beta_{2,n}\\ \vdots\\ \beta_{K,n}\end{bmatrix}$$

or $E\left(\hat{\beta}_k\right) = \sum_{n=1}^{N}\sum_{j=k}^{K} a_{kj,n}\beta_{j,n} = \sum_{n=1}^{N} a_{kk,n}\beta_{k,n} + \sum_{n=1}^{N}\sum_{j\neq k}^{K} a_{kj,n}\beta_{j,n} \qquad , \forall k = 1,\ldots\ldots, K$.

$$(14)\quad E\begin{bmatrix}\hat{\beta}_1\\ \hat{\beta}_2\\ \vdots\\ \hat{\beta}_K\end{bmatrix} = \sum_{n=1}^{N}\left(\begin{bmatrix}c & 0 & \cdots & 0\\ 0 & c & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & c\end{bmatrix} + \begin{bmatrix}a_{11,n}-c & 0 & \cdots & 0\\ 0 & a_{22,n}-c & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & a_{KK,n}-c\end{bmatrix} + \begin{bmatrix}0 & a_{12,n} & \cdots & a_{1K,n}\\ a_{21,n} & 0 & \cdots & a_{2K,n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{K1,n} & a_{K2,n} & \cdots & 0\end{bmatrix}\right)\begin{bmatrix}\beta_{1,n}\\ \beta_{2,n}\\ \vdots\\ \beta_{K,n}\end{bmatrix}$$

Note that Theil defines the true macro-parameters as either a simple sum of micro-parameters by using $c=1$ (Theil, 1954) or a simple average of micro-parameters by using $c=1/N$ (Theil, 1971). This choice of a constant c is arbitrary because it is not related to the weighting schemes used in the aggregation function of (3) or $(3')$, so it is not related to the correct specification of aggregated relation (5). For example, when we choose to use the same weighting schemes for Y and X by $W_n^y = W_n^x$ in $(3')$, the correct specification of aggregated relation (5') become exactly equivalent to (5), we can see that the choice of c does not depend on weighting schemes used in aggregation function and thus true macro-parameters defined based on the choice of c do not depend on the correct specification of aggregated relation.

Theil's conclusion summarized above has negative implications for the aggregate approach. Few economists will or can meaningfully interpret macro-parameters as complicated mixtures of heterogeneous components. However, meaningful aggregation can be possible based on a special case considered in Theil's discussions, which is the compositional stability condition. When each of macro-variable is composed of micro-variables of a homogeneous group with a constant compositional factor over time, the ratios of micro-variables over macro-variables becomes constant over time and the set of auxiliary equations (7) becomes equation $(7')$.

$$(7')\ \ [\, x_{1n}, x_{2n}, \cdots, x_{Kn}] = [\, X_1, X_2, \cdots, X_K \,] \begin{bmatrix} a_{11,n} & 0 & \cdots & 0 \\ 0 & a_{22,n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{KK,n} \end{bmatrix}$$

$$= [\, X_1 a_{11,n}, X_2 a_{22,n}, \cdots, X_K a_{KK,n} \,] \qquad , \forall\, n = 1,....., N$$

$$\text{or } \ x_{k\,n} = X_k\, a_{k\,k,n} \qquad , \forall\, k = 1,....., K \ , \forall\, n = 1,....., N \ .$$

This in turn implies that macro-parameters depend upon only the corresponding micro-parameters as in (13') , thus aggregated macro-parameters in macro-equations meaningfully and legitimately represent underlying homogeneous micro-parameters in micro-equations.

$$(13')\ \ E \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix} = \sum_{n=1}^{N} \begin{bmatrix} a_{11,n} & 0 & \cdots & 0 \\ 0 & a_{22,n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{KK,n} \end{bmatrix} \begin{bmatrix} \beta_{1,n} \\ \beta_{2,n} \\ \vdots \\ \beta_{K,n} \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} a_{11,n}\, \beta_{1,n} \\ \sum_{n=1}^{N} a_{22,n}\, \beta_{1,n} \\ \vdots \\ \sum_{n=1}^{N} a_{KK,n} \beta_{K,n} \end{bmatrix} \quad , \forall\, k = 1,....., K$$

$$\text{or } \ E\left(\hat{\beta}_k\right) = \sum_{n=1}^{N} a_{k\,k,n} \beta_{k,n} \qquad , \forall\, k = 1,....., K \ .$$

The homogeneity of micro-variables within a specific group is identified by the implied condition that micro-variables within the subset move absolutely synchronously and so have a correlation of one. By using the aggregation method that micro-variables are grouped and represented by macro-variables based on the condition that each macro-variable is composed of grouped micro-variables with a constant compositional factor $a_{k\,k,n}$ over time, (a) each macro-variable obtains a meaningful interpretation, since each macro-variable is composed of corresponding homogenous set of micro-variables measured by perfect correlation of one, and (b) each macro-parameter obtains a meaningful interpretation, since each macro-parameter is composed of only the corresponding homogeneous set of micro-parameters, not the non-corresponding micro-parameters. Note that this interpretation does not involve arbitrary choice of simple sum (if $c = 1$) or simple average (if $c = 1/N$).

This form of the compositional stability condition, however, requires a very strict condition that the variation in micro-variables within a group is strictly restricted by the equation of (7') $x_{k\,n} = X_k\, a_{k\,k,n}, \forall\, k = 1,....., K$ and $\forall\, n = 1,....., N$, without allowing any deviations from it. Obviously this condition is too restrictive to apply with real world data. In practice, the homogeneous group of micro-variables can only be identified through the certain group of micro-variables that are highly, but not perfectly, correlated, with the possibility that the aggregation bias in the aggregate model can be small as the specification error. In this respect, the strict form of compositional stability condition needs to be generalized for empirical applications. The strict proportionality condition for the postulated set of equations of micro-variables over macro-variables can be generalized for the less restrictive condition to obtain meaningful macro-parameters, which depend upon only the corresponding micro-parameters. When we decompose the set of auxiliary equations $x_n = X A_n + v_n$ into $x_n = X H_n + d_n$ as in (7'') and replace assumptions $Cov(v_n, X)=0$ and $E(v_n)=0$ with conditions $Cov(d_n, X)=0$ and $E(d_n)=0$ as in (8'), we obtain the same legitimate aggregation result as in (12'), by again using assumptions $Cov(u_n, X)=0$ and $E(u_n)=0$. Note that the conditions $Cov(u_n, X)=0$ and $E(u_n)=0$ are based on the assumptions $Cov(u_n, x_n)=0$ and $E(u_n)=0$, which are in turn from the background assumption of the correct specification of micro-equations (1) $y_n = x_n \beta_n + u_n$. Note again that in this study, $Cov(\ )=0$ is used to represent an independent relation, which is equivalent to no correlation under normal distribution with linearity.

(7'') $x_n = X A_n + v_n = X H_n + d_n$

$$\text{or } [x_{1n}, x_{2,n}, \cdots, x_{Kn}] = [X_1, X_2, \cdots, X_K] \begin{bmatrix} a_{11,n} & a_{12,n} & \cdots & a_{1K,n} \\ a_{21,n} & a_{22,n} & \cdots & a_{2K,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1,n} & a_{K2,n} & \cdots & a_{KK,n} \end{bmatrix} + [v_{1,n}, \cdots, v_{K,n}]$$

$$= [X_1, X_2, \cdots, X_K] \left\{ \begin{bmatrix} a_{11,n} & 0 & \cdots & 0 \\ 0 & a_{22,n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{KK,n} \end{bmatrix} + \begin{bmatrix} 0 & a_{12,n} & \cdots & a_{1K,n} \\ a_{21,n} & 0 & \cdots & a_{2K,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1,n} & a_{K2,n} & \cdots & 0 \end{bmatrix} \right\} + [v_{1,n}, \cdots, v_{K,n}]$$

$$= [X_1, X_2, \cdots, X_K] \begin{bmatrix} a_{11,n} & 0 & \cdots & 0 \\ 0 & a_{22,n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{KK,n} \end{bmatrix} + [d_{1,n}, d_{2,n}, \cdots, d_{K,n}],$$

where $[d_{1,n},\ d_{2,n},\ \cdots,\ d_{K,n}] = [\sum_{j\neq k}^{K} X_j a_{j1,n} + v_{1,n},\ \sum_{j\neq k}^{K} X_j a_{j2,n} + v_{2,n},\ \cdots,\ \sum_{j\neq k}^{K} X_j a_{jK,n} + v_{K,n}], \forall n=1,.....,N$.

(8') $\hat{H}_n = (X'X)^{-1} X' x_n$, where

$E(\hat{H}_n) = H_n + E\left[(X'X)^{-1} X' d_n\right]$, by assumption of $Cov(d_n, X) = 0$ and $E(d_n) = 0$

$= H_n$, $\forall n = 1, \ldots, N$.

(11') $\hat{\beta} = (X'X)^{-1} X'Y$, by (5) $Y = \sum_{n=1}^{N} x_n \beta_n + \sum_{n=1}^{N} u_n$

$= \sum_{n=1}^{N} (X'X)^{-1} X' x_n \beta_n + (X'X)^{-1} X' \sum_{n=1}^{N} u_n$, by (8') $\hat{H}_n = (X'X)^{-1} X' x_n$

$= \sum_{n=1}^{N} \hat{H}_n \beta_n + (X'X)^{-1} X' \sum_{n=1}^{N} u_n$.

(12') $E(\hat{\beta}) = \sum_{n=1}^{N} E\left[\hat{H}_n\right] \beta_n + E\left[(X'X)^{-1} X' \sum_{n=1}^{N} u_n\right]$, by assumption of $Cov(d_n, X) = 0$ and $E(d_n) = 0$

$= \sum_{n=1}^{N} H_n \beta_n + E\left[(X'X)^{-1} X' \sum_{n=1}^{N} u_n\right]$, by assumption of $Cov(u_n, X) = 0$ and $E(u_n) = 0$

$= \sum_{n=1}^{N} H_n \beta_n$

$$\text{or } E\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix} = \sum_{n=1}^{N} \begin{bmatrix} a_{11,n} & 0 & \cdots & 0 \\ 0 & a_{22,n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{KK,n} \end{bmatrix} \begin{bmatrix} \beta_{1,n} \\ \beta_{2,n} \\ \vdots \\ \beta_{K,n} \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} a_{11,n} \beta_{1,n} \\ \sum_{n=1}^{N} a_{22,n} \beta_{1,n} \\ \vdots \\ \sum_{n=1}^{N} a_{KK,n} \beta_{K,n} \end{bmatrix} .$$

This generalized form of the compositional stability requires the condition of $Cov(d_n, X) = 0$ in the set of equations $x_n = X H_n + d_n$. Hausman (1978) shows that this type of no regressor-error correlation condition can be empirically studied by using a statistical test of $H_0 : \gamma_n = 0$ in $x_n = X H_n + IV \cdot \gamma_n + \varepsilon_n^{IV}$, where IV are Instrumental Variables such that IV is closely correlated with regressor X (the relevance condition of IV) and independent of error d_n (the validity condition of IV). Based on this Hausman type misspecification testing method, we can empirically test the generalized form of the compositional stability condition, if we can find appropriate instrumental variables.

In terms of identifying the homogeneous group of micro-variables, it is also possible to generalize the strict requirement that micro-variables of all items within the subset move absolutely synchronously and have a correlation of one. The main feature of the compositional stability condition is that each macro-variable is composed of grouped micro-variables with a "stable" compositional factor over time, so the ratios of micro-variables over macro-variables are "near" constant with a certain stability over time. In this respect, the compositional stability condition can be generalized to require a less strict requirement. We can use the conditions that micro-variables within group are highly correlated but micro-variables across groups are only weakly correlated over time, instead of the strict requirement that micro-variables within group are perfectly correlated with correlation of one. Not only the degree of co-movement, but also the way to measure the co-movement can be generalized. While the strict form of the compositional stability condition requires that micro-variables within the subset move absolutely synchronously, the generalized form of the compositional stability condition can allow the possible lead and lag dependencies among micro-variables within a group, as long as micro-variables within the group are highly correlated but micro-variables across groups are only weakly correlated. While the standard static correlation only measures synchronous or contemporaneous co-movements between variables and requires an independence assumption over time, there are several alternative measurements of dependency allowing for possible leads and/or lags in dependency among the time-series data in a dynamic setting. Two of these are the co-integration and the cross correlation. Co-integration is designed to measure long-run co-movements, so it can be too restrictive to use for identifying mid-run or short-run or contemporaneous dependency patterns. The cross-correlation with some leads and lags can capture mid-run or short-run dependency by varying lead and lag parameters, but the choice of lead and lag parameters can be somewhat arbitrary. In this respect, we propose to use the standard static correlation as well as the dynamic correlation defined in (15) and (16) to measure the high co-movements of micro-variables within a group and near independences of micro-variables across groups.

(15) $\rho_{xy}(\lambda)=\dfrac{C_{xy}(\lambda)}{\sqrt{S_x(\lambda)\cdot S_y(\lambda)}}$ for frequency λ where $-\pi \le \lambda \le \pi$

(16) $\rho_{xy}(\Lambda)=\dfrac{\int_\Lambda C_{xy}(\lambda)d\lambda}{\sqrt{\int_\Lambda S_x(\lambda)d\lambda \cdot \int_\Lambda S_y(\lambda)d\lambda}}$ for frequency band $\Lambda=[\lambda_1,\lambda_2)$ where $0 \le \lambda_1 < \lambda_2 \le \pi$,

where x and y are two zero-mean real stochastic processes, $S_x(\lambda)$ and $S_x(\lambda)$ are the spectral density functions, and $C_{xy}(\lambda)$ is the co-spectrum of x and y.

The dynamic correlation, proposed from the frequency domain approach, has useful properties such as: (a) The dynamic correlation measures different degrees of co-movement which varies between -1 and 1 just as standard static correlation. (b) The dynamic correlation over the entire frequency band is identical to static correlation after suitable pre-filtering and it is also related to stochastic co-integration. (c) The dynamic correlation can be decomposed by frequency and frequency band, where the low or high frequency band in spectral domain have implication for the long-run or short-run in time domain respectively (Croux, Forni, and Reichlin, 2001).

This generalization of the compositional stability condition in terms of not only the degree of co-movement but also the way to measure the co-movement makes it possible to approximate the condition of $Cov(d_n, X)=0$ and $E(d_n)=0$ by the condition of $Cov(d_k, d_{k'})\leq\delta$, $\forall k \neq k'$ where δ is a small value. This approximate condition in turn implies a block-diagonal pattern of the covariance or correlation matrix among micro-variables as in (17). The correlation is measured by static correlation $Corr(\chi)$ or dynamic correlation $DynCorr(\chi)$, where χ is defined as follows. We first transpose $x_n = X A_n + v_n$ into $x_n^T = A_n^T X^T + v_n^T$, $\forall n=1,.....,N$. By expanding to incorporate all $K\cdot N$ micro-variables, we can write $x_n = X A_n + v_n$ as the matrix form $\chi = L\cdot\aleph+\upsilon$. Based on the logic of decomposition of set of auxiliary equations to derive generalized compositional stability condition, we decompose $\chi = L\cdot\aleph+\upsilon$ into $BDiag(L)\cdot\aleph+d$ as in (18), where the dimension of χ, υ and d are $(KN\times T)$, and L is of dimension $(KN\times K)$, $\aleph$ is of dimension $(K\times T)$, and $BDiag(L)$ denotes a block diagonal matrix of L. The equation (7'') is recalled to clarify the relationship, where $Diag(A_n)$ denotes a diagonal matrix of A_n.

(17) $\Sigma = Corr(\chi)$ or $DynCorr(\chi)$

$$= \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1K} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{K1} & \Sigma_{K2} & \cdots & \Sigma_{KK} \end{bmatrix} \quad \text{, by compositional stability assumption of } Cov(d_k, d_{k'})\leq\delta$$

$$\approx \begin{bmatrix} \Sigma_{11} & 0 & \cdots & 0 \\ 0 & \Sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{KK} \end{bmatrix} \quad \text{, where } \Sigma_{kk} = \begin{bmatrix} 1 & \rho_{k,12} & \cdots & \rho_{k,1N} \\ \rho_{k,21} & 1 & \cdots & \rho_{k,2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k,N1} & \rho_{k,N2} & \cdots & 1 \end{bmatrix}$$

(18) $x_n^{\,T} = A_n^{\,T} X^T + v_n^{\,T} \qquad , \forall n = 1,.....,N$

or $\begin{bmatrix} \chi_{1,n} \\ \chi_{2,n} \\ \vdots \\ \chi_{K,n} \end{bmatrix} = \begin{bmatrix} a_{11,n} & a_{21,n} & \cdots & a_{K1,n} \\ a_{12,n} & a_{22,n} & \cdots & a_{K2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1K,n} & a_{2K,n} & \cdots & a_{KK,n} \end{bmatrix} \begin{bmatrix} \aleph_1 \\ \aleph_2 \\ \vdots \\ \aleph_K \end{bmatrix} + \begin{bmatrix} \upsilon_{1n} \\ \upsilon_{2n} \\ \vdots \\ \upsilon_{Kn} \end{bmatrix} = \begin{bmatrix} l_{11,n} & l_{12,n} & \cdots & l_{1K,n} \\ l_{21,n} & l_{22,n} & \cdots & l_{2K,n} \\ \vdots & \vdots & \ddots & \vdots \\ l_{K1,n} & l_{K2,n} & \cdots & l_{KK,n} \end{bmatrix} \begin{bmatrix} \aleph_1 \\ \aleph_2 \\ \vdots \\ \aleph_K \end{bmatrix} + \begin{bmatrix} \upsilon_{1,n} \\ \upsilon_{2,n} \\ \vdots \\ \upsilon_{K,n} \end{bmatrix}$

or $\left[\begin{array}{c} \chi_{1,1} \\ \chi_{1,2} \\ \vdots \\ \chi_{1,N} \\ \hdashline \chi_{2,1} \\ \chi_{2,2} \\ \vdots \\ \chi_{2,N} \\ \hdashline \vdots \\ \hdashline \chi_{K,1} \\ \chi_{K,2} \\ \vdots \\ \chi_{K,N} \end{array}\right] = \left[\begin{array}{c:c:c:c} l_{11,1} & l_{12,1} & \cdots & l_{1K,1} \\ l_{11,2} & l_{12,2} & \cdots & l_{1K,2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{11,N} & l_{12,N} & \cdots & l_{1K,N} \\ \hdashline l_{21,1} & l_{22,1} & \cdots & l_{2K,1} \\ l_{21,2} & l_{22,2} & \cdots & l_{2K,2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{21,N} & l_{22,N} & \cdots & l_{2K,N} \\ \hdashline \vdots & \vdots & \ddots & \vdots \\ \hdashline l_{K1,1} & l_{K2,1} & \cdots & l_{KK,1} \\ l_{K1,2} & l_{K2,2} & \cdots & l_{KK,2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{K1,N} & l_{K2,N} & \cdots & l_{KK,N} \end{array}\right] \cdot \begin{bmatrix} \aleph_1 \\ \aleph_2 \\ \vdots \\ \aleph_K \end{bmatrix} + \left[\begin{array}{c} \upsilon_{1,1} \\ \upsilon_{1,2} \\ \vdots \\ \upsilon_{1,N} \\ \hdashline \upsilon_{2,1} \\ \upsilon_{2,2} \\ \vdots \\ \upsilon_{2,N} \\ \hdashline \vdots \\ \hdashline \upsilon_{K,1} \\ \upsilon_{K,2} \\ \vdots \\ \upsilon_{K,N} \end{array}\right]$

or $\chi_{k,n} = \left(\sum_{j=1}^{K} a_{jk,n} \aleph_j \right) + \upsilon_{k,n} = \left(\sum_{j=1}^{K} l_{kj,n} \aleph_j \right) + \upsilon_{k,n}$, where $\chi_{k,n} = x_{k,n}^{\,T}$, $\aleph_k = X_k^{\,T}$, and $\upsilon_{k,n} = v_{k,n}^{\,T}$

or $\chi = L \cdot \aleph + \upsilon = BDiag(L) \cdot \aleph + d$

or $\begin{bmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_K \end{bmatrix} = \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1K} \\ L_{21} & L_{22} & \cdots & L_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ L_{K1} & L_{K2} & \cdots & L_{KK} \end{bmatrix} \begin{bmatrix} \aleph_1 \\ \aleph_2 \\ \vdots \\ \aleph_K \end{bmatrix} + \begin{bmatrix} \upsilon_1 \\ \upsilon_2 \\ \vdots \\ \upsilon_K \end{bmatrix}$, where $\chi_k = \begin{bmatrix} \chi_{k,1} \\ \chi_{k,2} \\ \vdots \\ \chi_{k,N} \end{bmatrix}$, $L_{kj} = \begin{bmatrix} l_{kj,1} \\ l_{kj,2} \\ \vdots \\ l_{kj,N} \end{bmatrix}$, $\upsilon_k = \begin{bmatrix} \upsilon_{k,1} \\ \upsilon_{k,2} \\ \vdots \\ \upsilon_{k,N} \end{bmatrix}$

or $\begin{bmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_K \end{bmatrix} = \begin{bmatrix} L_{11} \aleph_1 \\ L_{22} \aleph_2 \\ \vdots \\ L_{KK} \aleph_K \end{bmatrix} + \begin{bmatrix} \sum_{j \neq 1}^{K} L_{1j} \aleph_j + \upsilon_1 \\ \sum_{j \neq 2}^{K} L_{2j} \aleph_j + \upsilon_2 \\ \vdots \\ \sum_{j \neq K}^{K} L_{Kj} \aleph_j + \upsilon_K \end{bmatrix}$, where $\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_K \end{bmatrix} = \begin{bmatrix} \sum_{j \neq 1}^{K} L_{1j} \aleph_j + \upsilon_1 \\ \sum_{j \neq 2}^{K} L_{2j} \aleph_j + \upsilon_2 \\ \vdots \\ \sum_{j \neq K}^{K} L_{Kj} \aleph_j + \upsilon_K \end{bmatrix}$

or $\begin{bmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_K \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & \cdots & 0 \\ 0 & L_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & L_{KK} \end{bmatrix} \begin{bmatrix} \aleph_1 \\ \aleph_2 \\ \vdots \\ \aleph_K \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_K \end{bmatrix}$, where $d_k = \begin{bmatrix} d_{k,1} \\ d_{k,2} \\ \vdots \\ d_{k,N} \end{bmatrix}$.

(7'') $x_n = X A_n + v_n = X \cdot Diag(A_n) + d_n = X H_n + d_n$

$$\text{or } [x_{1n}, x_{2,n}, \cdots, x_{Kn}] = [X_1, X_2, \cdots, X_K] \begin{bmatrix} a_{11,n} & a_{12,n} & \cdots & a_{1K,n} \\ a_{21,n} & a_{22,n} & \cdots & a_{2K,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1,n} & a_{K2,n} & \cdots & a_{KK,n} \end{bmatrix} + [v_{1,n}, \cdots, v_{K,n}]$$

$$= [X_1, X_2, \cdots, X_K] \begin{bmatrix} a_{11,n} & 0 & \cdots & 0 \\ 0 & a_{22,n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{KK,n} \end{bmatrix} + [d_{1,n}, d_{2,n}, \cdots, d_{K,n}],$$

where $[d_{1,n},\ d_{2,n},\ \cdots,\ d_{K,n}] = [\sum_{j \neq k}^{K} X_j a_{j1,n} + v_{1,n}, \sum_{j \neq k}^{K} X_j a_{j2,n} + v_{2,n}, \cdots, \sum_{j \neq k}^{K} X_j a_{jK,n} + v_{K,n}], \forall n = 1,....., N$.

Note that the equivalence between $x_n = X A_n + v_n$ and $\chi = L \cdot \aleph + \upsilon$ through $x_n^T = A_n^T X^T + v_n^T$ implies the equivalence between $x_n = X \cdot Diag(A_n) + d_n = X H_n + d_n$ and $\chi = BDiag(L) \cdot \aleph + d$. Given that the strict form of compositional stability condition $x_n = X \cdot Diag(A_n)$ implies the block diagonal structure in the standard correlation matrix $\Sigma = Corr(\chi)$, we can infer the approximate form of compositional stability condition $x_n = X\ Diag(A_n) + d_n = X H_n + d_n$ with $Cov(d_k, d_{k'}) \leq \delta$, $\forall k \neq k'$ by identifying the approximate block diagonal structure in static or dynamic correlation matrix $\Sigma = Corr(\chi)$ or $DynCorr(\chi)$. Note that the generalized form of the stability condition $Cov(d_n, X) = 0$ and $E(d_n) = 0$ is approximated by the condition of $Cov(d_k, d_{k'}) \leq \delta$, $\forall k \neq k'$ in the equation $x_n = X A_n + v_n = X \cdot Diag(A_n) + d_n = X H_n + d_n$.

This approximate form of the compositional stability condition can also be used to search for a specific homogeneous group to define an interpretable macro-variable, which is composed of highly correlated micro-variables with stable compositional factor. In this case, we can use an index k as micro-variables' group index that should be empirically identified, instead of using k as an index for pre-determined classes of exogenous variables. The problem of forming suitable partitions before conducting any empirical test to justify those classifications has relied on researchers' intuition rather than empirical data patterns. For example from demand analysis, intuitive partitions are formed based on several reference variables such as animal origin, product quality etc., which hopefully proxy consumers' unobservable marginal utility structures. This intuition-based approach has an ambiguous aspect, since alternative choices of reference variables may result in several different classifications. More fundamentally, such intuitive partitions based on the subjective reasoning are only a small set of possible partitions among an extremely large number of possible partitions. Thus when classification is empirically rejected, it can be simply because of researchers' unsuccessful identification of the partition itself, not because of non-existence of legitimate classification itself. Given the empirical implausibility of attempting all possible partitions, it is better to pursue inductive partitions related with legitimate aggregation conditions based on the data pattern itself. The approximate form of the compositional stability condition can be used for searching for a specific homogeneous group, which is composed of highly correlated micro-variables with a stable compositional factor, so it allows us to define an interpretable macro-variable based on empirical data patterns. For this purpose, we choose to use the modified k-nearest neighbor algorithm based on Wise's pseudo-color map code in this study. This algorithm keeps track of changes of correlation matrix, when it reorders the variables in the correlation matrix to sort highly correlated variables near each other along the main diagonal. On the other hand, other standard clustering methods, such as hierarchical algorithm and k-mean algorithm, use the correlation matrix as only an initial input of similarity measures and thus it is not easy to keep track of changes of correlation matrix (Xu and Wunsch, 2005). For example, based on the same correlation matrix from macro-economic data used in preliminary study, the modified k-nearest neighbor algorithm, which returns an intuitively interpretable reordered final correlation matrix as a final result, provides a meaningful clustering result, whereas the hierarchical algorithm, which returns a not-easy-to-interpret dendogram as a final result, only provides an ambiguous final clustering result based on either the intuitive reasoning or the correlation matrix.

Theil reaches his generally negative conclusion for aggregation based on two kinds of main assumptions. One is $Cov(u_n, X)=0$ and $E(u_n)=0$, which is related with the background assumption of correctly specified micro-equations. The other is $Cov(v_n, X)=0$ and $E(v_n)=0$, which is the primary assumption that makes it possible to relate the macro-parameters to the micro-parameters. By replacing these primary assumptions with the testable condition of $Cov(d_n, X)=0$ and $E(d_n)=0$, we reach a generalized form of the compositional stability condition for the positive possibility of legitimate aggregation. This generalized condition is, however, involved with the difficult search for instrumental variables in a Hausman-type misspecification test in the set of equations $x_n = X H_n + d_n$. When appropriate instrumental variables are not available, we can use the approximate form of the compositional stability condition based on the empirically identifiable pattern of $Cov(d_k, d_{k'}) \leq \delta$ through the implied block-diagonal pattern in a static or dynamic correlation matrix among micro-variables. This approximate form of the compositional stability condition can also be used for searching specific homogeneous groups of original variables to form an initial partitioning.

Index Number Theory

Heretofore, we have explored the possibility for legitimate aggregation in generalized forms of the compositional stability condition based on Theil's aggregation theory. Given that Theil's theory is valid for the weighted average aggregation, mutatis mutandis, as mentioned in $(5')$, one of the remaining issues is how to decide the weighting schemes in aggregating micro-variables into macro-variables. This issue has been studied under the Index number theory, which is based on distinct features of economic phenomena, especially in the area of micro-economic. All economic transactions on N commodities reveal dual pairs of information of prices $p = [p_1, p_2, \cdots, p_N]$ and quantities $q = [q_1, q_2, \cdots, q_N]$ such that total sum of each product of individual price and quantity equals the total value (V) of N commodities. There have been many different index formulas suggested to represent these dual pairs of individual information by a pair of aggregate price index P and aggregate quantity index Q such that the product of the price index and the quantity index equals the total value of N commodities. In this context, the index number problem can be understood to find P and Q for given p, q and V as in (19)

(19) $$p \cdot q^T \equiv \sum_{n=1}^{N} p_n q_n \equiv \sum_{n=1}^{N} v_n \equiv V = P \cdot Q$$

However, it turns out that it is mathematically impossible to determine functional forms of aggregate price and quantity variables, when (a) both the price p and quantity q vector are regarded as independent variables and (b) aggregate price P and quantity Q variables have a positivity property (Eichhorn, 1978). Many distinct index formulas suggested are based on the some variants of equation (19) as explained below as $(19')$, $(19'')$, or $(19''')$ i.e. instead of decomposing total value level into price and quantity level, the alternative forms of the decomposition of total value change over time into the product of the price change component and the quantity change component, which uses the relative price and relative quantity to define the aggregate index (Diewert, 2001). Many different index formulas can be understood based on five different approaches, which are fixed basket, differential, economic, stochastic and axiomatic approaches. Note that if the price index is determined, then the quantity index may be implicitly decided using the product rule ($V^1 / V^0 = P \cdot Q$), or vice versa. Thus discussions can be focused on the price index.

The fixed basket approach tries to decompose total value ratio over time into aggregate price and quantity components as in $(19')$. The price index is defined as the value ratio for the price changes to purchase a fixed reference basket of quantities $m(q)$ as in (20). Different price indexes can be derived, depending on how one chooses the fixed basket as a common reference commodity bundle $m(q)$ representing the two periods. Choosing $m(q) = q^0$ or $m(q) = q^1$ results in the Laspeyres or Paasche index, respectively. Choosing annual base year quantities for $m(q)$ results in the Lowe index, which is used by many statistical institutes to produce monthly data in timely fashion. If we choose the geometric average of $m(q) = \sqrt{q^0 \cdot q^1}$ for the reference basket or take the geometric average of the Laspeyres or Paasche indexes $\sqrt{P_{Laspeyres} \cdot P_{Paasche}}$, we get the Walsh or Fisher index, respectively (Diewert, 2001).

$$(19')\ \frac{\sum_{n=1}^{N} p_n^1 q_n^1}{\sum_{n=1}^{N} p_n^0 q_n^0} \equiv \frac{V^1}{V^0} = P \cdot Q \qquad (20)\ P \equiv \frac{\sum_{n=1}^{N} p_n^1 \cdot m(q)}{\sum_{n=1}^{N} p_n^0 \cdot m(q)} \quad \text{and } Q = \frac{V^1}{V^0} \Big/ P$$

In the Divisia differential approach, the observed price, quantity, and value are regarded as continuous functions of (continuous) time. By taking differentials with respect to time, the logarithmic rate of changes of total value is decomposed into logarithmic rate of changes of price and quantity as in $(19'')$. This approach treats price and quantity indexes symmetrically. Different price indices can be derived, depending on how one makes discrete approximations to the continuous time index (21). If we take the arithmetic average of $\left(s_n^{\ 0} + s_n^{\ 1}\right)/2$ for numerical approximation or assume the most regular path of monotone paths or constant growth rate paths for line integrals in the absence of additional information, we get the Tornquivist-Theil price index (Hillinger, 2002).

$$(19'')\ p(t) \cdot q(t)^T \equiv \sum_{n=1}^{N} p_n(t) q_n(t) \equiv \sum_{n=1}^{N} v_n(t) \equiv V(t) = P(t) \cdot Q(t)$$

$$\Rightarrow \sum_{n=1}^{N} \frac{\partial V}{\partial p_n} dp_n + \sum_{n=1}^{N} \frac{\partial V}{\partial q_n} dq_n \equiv dV = \frac{\partial V}{\partial P} dP + \frac{\partial V}{\partial Q} dQ$$

$$\Rightarrow \sum_{n=1}^{N} \frac{q_n p_n}{V} \frac{dp_n}{p_n} + \sum_{n=1}^{N} \frac{p_n q_n}{V} \frac{dq_n}{q_n} \equiv \frac{dV}{V} = \frac{QP}{V} \frac{dP}{P} + \frac{PQ}{V} \frac{dQ}{Q}$$

$$\Rightarrow \sum_{n=1}^{N} s_n d\ln p_n + \sum_{n=1}^{N} s_n d\ln q_n \equiv d\ln V = d\ln P + d\ln Q,$$

because $q_n = \dfrac{\partial V}{\partial p_n}$ and $p_n = \dfrac{\partial V}{\partial q_n}$ from $\sum_{n=1}^{N} p_n(t)q_n(t) \equiv V(t)$, $s_n \equiv \dfrac{v_n}{V} \equiv p_n q_n \Big/ \sum_{n=1}^{N} p_n q_n$,

$Q = \dfrac{\partial V}{\partial P}$ and $P = \dfrac{\partial V}{\partial Q}$ from $V(t) = P(t)\cdot Q(t)$, and $\dfrac{dz}{z} = d\ln z$.

(21) $\ln\dfrac{p^1}{p^0}\left(\equiv \ln P_0^1\right) = \int_0^1 \sum_{n=1}^{N} s_n(\tau)\,\dfrac{d\ln p_n(\tau)}{d\tau}d\tau$ and $\ln\dfrac{q^1}{q^0}\left(\equiv \ln Q_0^1\right) = \int_0^1 \sum_{n=1}^{N} s_n(\tau)\,\dfrac{d\ln q_n(\tau)}{d\tau}d\tau$

In the economic approach, observed quantity is regarded as the solution of an individual's optimization decision, given price data. This approach explicitly uses functional relations between quantity and price by assuming that the consumer (producer) is maximizing a utility (production) function subject to a budget constraint or minimizing cost function subject to a given utility (output) level as in (19'''). The price index or cost of living index is defined as the ratio of minimum cost for the price changes to achieving the common reference utility (production) level representing two periods as in (22). Different price indices can be derived, depending on how one chooses both the reference utility (production) level and the functional form of utility (production) function $u(\cdot)$, the cost function $C(\cdot)$ or Mckenzie expenditure function $M(\cdot)$ (Balk, 2005). If we choose the geometric average of $u(q)=\sqrt{u^0u^1}$ for the reference utility level and the translog functional form for the quadratic approximation to arbitrary cost function $C(p,u)$, economic price index or cost of living index becomes the Tornquivist-Theil price index (Diewert, 2001).

(19''') $p(t)\cdot q(t)^T \equiv M\big(p(t),q(t),t\big) \equiv C\big(p(t),u(q,t),t\big) \equiv \underset{q}{Min}\left\{ p\cdot q^T \mid u(q,t) \ge u(q(t),t)\right\}$

$$\equiv V(t) = P(t)\cdot Q(t).$$

(22) $P \equiv \dfrac{C\big(p^1,u(q)\big)}{C\big(p^0,u(q)\big)}$ and $Q = \dfrac{V^1}{V^0}\Big/P$.

Note that the economic approach has a similar idea with the fixed basket approach in using common reference vector representing standard of living in two periods. While the fixed basket approach uses common reference commodity bundles to represent the two periods, the economic approach uses common reference utility (production) level to represent the two periods. The economic approach can also be understood in the connection to the Divisia differential approach by using differential property of the Mckenzie-expenditure function as in (23) (Balk, 2005).

$$(23)\ \left\langle dV \equiv \sum_{n=1}^{N} \frac{\partial V}{\partial p_n} dp_n + \sum_{n=1}^{N} \frac{\partial V}{\partial q_n} dq_n \right\rangle = \sum_{n=1}^{N} q_n dp_n + \sum_{n=1}^{N} p_n dq_n = \left\langle \sum_{n=1}^{N} \frac{\partial M}{\partial p_n} dp_n + \sum_{n=1}^{N} \frac{\partial M}{\partial q_n} dq_n \equiv dM \right\rangle$$

because $q_n = \frac{\partial V}{\partial p_n}$ and $p_n = \frac{\partial V}{\partial q_n}$ using $\sum_{n=1}^{N} p_n(t) q_n(t) \equiv V(t)$

and $q_n = \frac{\partial M}{\partial p_n}$ and $p_n = \frac{\partial M}{\partial q_n}$ using $p(t) \cdot q(t)^T \equiv M\left(p(t), q(t), t\right)$

In the stochastic approach, each of the observed N price relatives or some transformation of price relatives is regarded as an estimate of a common inflation rate with an idiosyncratic error term as in (24), whose variability decreases as the representative value share increases, i.e. as the commodity becomes more important in the budget. This approach can be used to derive a standard error of the index number. Different price indices can be derived by applying Generalized Least Squares method, depending on how the functions $f(\cdot)$ and $m(\cdot)$ are chosen as in (25). The choice of natural logarithm for $f(\cdot)$ and the arithmetic average for $m(s_n) = \left(s_n^{\,0} + s_n^{\,1}\right)/2$ results in the Tornquivist-Theil price index (Selvanathana and Prasada Rao, 1994).

$$(24)\ f\left(\frac{p_n^1}{p_n^0}\right) = f(P) + \varepsilon_n \qquad , \forall n = 1, \ldots, N \text{, where } \varepsilon_n \sim \left(0, \frac{\sigma^2}{m(s_n)}\right).$$

$$(25)\ \sqrt{m(s_n)}\, f\left(\frac{p_n^1}{p_n^0}\right) = \sqrt{m(s_n)}\, f(P) + \sqrt{m(s_n)}\varepsilon_n \qquad , \forall n = 1, \ldots, N \text{, where } \sqrt{m(s_n)}\varepsilon_n \sim \left(0, \sigma^2\right)$$

$$\text{so } \hat{f}(P) = \frac{\sum_{n=1}^{N} \sqrt{m(s_n)} \cdot \sqrt{m(s_n)}\, f\left(\frac{p_n^1}{p_n^0}\right)}{\sum_{n=1}^{N} \left(\sqrt{m(s_n)}\right)^2} = \sum_{n=1}^{N} f\left(\frac{p_n^1}{p_n^0}\right) \cdot m(s_n) \text{, using restriction of } \sum_{n=1}^{N} m(s_n) = 1.$$

There have been many index number formulas suggested, so it is useful to be able to evaluate various index number formulas in terms of their mathematical properties. In the axiomatic approach, it is attempted to determine whether a formula is consistent with reasonable properties. For example, good index number formulas should be invariant to changes in commodity ordering and measurement unit (Invariance test) and should become reciprocal to changes in time ordering (Time reversal test). A good price (or quantity) index should also be proportional to current period price (or quantity) vector p^1 (or q^1) and inverse proportional to base period price (or quantity) vector p^0 (or q^0) (Homogeneity tests). Note that properties derived from or imposed on the price index can be transferred to quantity index by using the product rule, and vice versa. The difficulty in this axiomatic approach is the fact that there is no universal agreement on what the best set of reasonable axioms is (Diewert, 2004). For example, The Walsh index is considered as a good index based on the time reversal test and invariance test within the average basket approach. The Fisher index is considered as a good index from the axiomatic approach in the framework of $P(p^0, p^1, q^0, q^1)$ based on list of 20 properties. The Tornqvist-Theil index is regarded as a good index from the axiomatic approach in the framework of $P(p^1/p^0, v^0, v^1)$ based on a similar list of properties.

We choose to use the Tornqvist-Theil index in this study, although it has been argued that the Tornqvist-Theil, Walsh, and Fisher indexes are approximately equivalent as the class of superlative indexes. The preference toward the Tornqvist-Theil index, rather than the Fisher index, is due to following facts: (a) Although almost all of index number formulas suggested can be derived from any of five approaches by making different choices, the Tornqvist-Theil index is easily justified from any of four approaches, because it can be derived from almost all of approaches to index number theory with a reasonable choice within each approach. (b) Unlike the Fisher index, the Tornqvist-Theil index does not invoke the problematic assumption of a homothetic or linear homogeneous utility function.

The class of superlative indexes and their relations with the homothetic assumption can be understood as follows. If we assume the utility (production) function is linearly homogeneous in quantities, then the cost function can be decomposed into a utility (production) level times a unit cost function, which is linearly homogeneous in prices (26). In this case, the cost of living index becomes a unit cost ratio which is independent of the reference quantity vector and the (implicit) quantity index becomes a utility ratio which is also independent of the reference price vector (27).

(26) $C\big(p, u(q)\big) = c(p) \cdot u(q)$,

where $c(p)$ is linearly homogeneous unit cost function and $C(\cdot)$ is cost function,

when $u(q)$ is linearly homogeneous utility (or production) function

(27) $P \equiv \dfrac{C\big(p^1, u(q)\big)}{C\big(p^0, u(q)\big)} = \dfrac{c(p^1) \cdot u(q)}{c(p^0) \cdot u(q)} = \dfrac{c(p^1)}{c(p^0)}$ and $Q \equiv \dfrac{V^1}{V^0} \Big/ P = \dfrac{c(p^1) \cdot u(q^1)}{c(p^0) \cdot u(q^0)} \Big/ \dfrac{c(p^1)}{c(p^0)} = \dfrac{u(q^1)}{u(q^0)}$

Using the fact that any arbitrary (twice continuously differentiable) linear homogeneous function can be approximated to the second order by the quadratic mean of order r function or the flexible function (28), Diewert uses flexible functional form for approximating the linearly homogenous utility (production) or unit cost function to define second-order approximate indexes for price and quantity index. Note that the utility (production) function determines the unit cost function, and vice versa, due to the duality theorem.

(28) $c(p)\ or\ u(q) \approx\ f^r(z) = \left[\sum_{m=1}^{N}\sum_{n=1}^{N}\alpha_{m,n}\cdot z_m^{\ r/2}\cdot z_n^{\ r/2}\right]^{/r}$, where $r \neq 0$, $\alpha_{mn} = \alpha_{nm}$,

and $z = p\ or\ q$ for unit cost and utility functions respectively.

Diewert argues that all of approximate indexes or superlative indexes, depending on the choice of value r, approximate each other to the second order, either when it is estimated at the point where prices and quantities are equal over time ($p^1 = p^0$ and $q^1 = q^0$) or when prices and quantities move exactly proportionally ($p^1 = \mu \cdot p^0$ and $q^1 = \varphi \cdot q^0$, $\forall \mu, \varphi > 0$). After showing that the superlative index become the Walsh index or the Fisher index when $r = 1$ or $r = 2$ respectively and as r tends to 0, a limiting case of superlative index become the Tornqvist-Theil index, Diewert argues that the standard superlative indexes such as the Tornqvist-Theil ($r \to 0$), Walsh ($r = 1$), Fisher ($r = 2$) indexes and other infinite number of higher order r superlative indexes will all give the same answer to a reasonably high degree of approximation and concludes that the choice among superlative indexes does not matter much in empirical applications.

Diewert's conclusion is based on the homothetic utility function and proportionality assumptions. It is interesting that these two assumptions are related with two approaches for legitimate aggregation condition. In economics, especially micro-economics, legitimate aggregation conditions or valid classification conditions have been studied or identified based on either pattern of variables or pattern of preference (or technology). While the homothetic condition of preference (or technology) patterns is related with the separability condition, the proportionality condition of variable patterns is related with the Hick-Leontief composite commodity theorem. In terms of preference (or technology) pattern, it is argued that there can be group demand functions, when a structural property of preference (or technology) reveals a pattern such that the marginal rate of substitution of all pairs of items within the subset is homogenous of degree zero in the quantities of items within the subset and is also independent of the quantities of all items outside the subset. While both conditions are required for homothetic separability, the latter condition is required for weak separability. Although the weakly separable condition implies only quantity aggregates not price aggregates, both of which are required for conducting consistent two-stage budgeting (Shumway and Davis, 2001). However, the homothetic assumption might be problematic due to its implication of unrealistic unitary income elasticities. The separability assumption implies rather strong condition, is difficult to test powerfully, and requires group price indexes that depend on the parameters of the individual utility (production) function (Lewbel, 1996). Separability condition is tested by estimating models for individual goods without imposing separability, and then testing whether the required elasticity restrictions such that the ratio of compensated cross-price elasticities of two commodities within the same group with respect to a third commodity in another group is equal to the ratio of their expenditure elasticities are satisfied. The problem is that without separability, each demand equation must include all the related individual prices. Even when enough degrees of freedom are available to estimate these models, the multicollinearity among the prices as well as the relatively complicated cross equation parameter restrictions causes the resulting tests to have little or no power. In a Monte Carlo study Barnett and Choi (1989) find that all of the standard tests fail to reject separability much of the time, even with data constructed from utility functions that are far from separable. Even though this "difficulty to reject" may be one reason why separability is so commonly assumed in practice, separability is often empirically rejected (see Diewert and Wales, 1995, for example). Although progress has been made in relaxing its restrictions (see Blundell and Robin, 1995 and Blackorby et al., 1995, for examples), even weak forms of separability impose very strong elasticity equality restrictions among every good in every group.

While the homothetic assumption is not easy to empirically test as discussed above, it can also be problematic in the context of index number theory, since it is challenged by the recent findings of index number theories. For example, Hill (2006) shows that although Diewert's approximation result is mathematically valid and has convenient implications for practical purposes, superlative indexes with higher order values of r do not necessarily empirically approximate the standard superlative indexes very closely, using two sets of empirical data. Hillinger (2002) further demonstrates that the Fisher index is not a quadratic approximation to the true index in the general non-homothetic case, while Tornqvist-Theil index is very accurate, using simulation data set generated by the simple non-homothetic form of the Stone/Geary utility function. Hillinger's simulation result is consistent with Samuelson and Swamy (1974, page 585)'s conclusion that "it is evident that the ideal (Fisher) index cannot give high-powered approximation to the true index in the general non-homothetic case." In general, the difficulties in empirical applications of the separability condition can be understood as the separability concept requires the complete knowledge of all micro-parameters. Similarly, the micro-homogeneity condition also requires a similar degree of information of all micro-parameters to check the equality of micro-parameters across all individual units. This requirement of the complete knowledge of all micro-parameters, which for instance is not easy to be estimated consistently due to multicollinearity problem, can be too restrictive to use for practical purposes. For this reasons, we do not assume any restrictions on micro-parameters based on either the micro-homogeneity condition for each individual micro-unit or the homothetic or weakly separability condition for the utility (production) function in this study.

While the homothetic and separability conditions and the related micro-homogeneity condition are based the complete knowledge of all micro-parameters, the Hicks or Leontief composite commodity theorem, Lewbel's generalized composite commodity theorem and the compositional stability condition are based on patterns of micro-variables within the subset category without requiring any knowledge of micro-parameters. The Hicks or Leontief composite commodity theorem is based on patterns of the prices or quantities of all items within the subset category respectively. It is argued that there can be composite commodities, when the ratios of the prices (quantities) of individual commodities to composite commodity price (quantity) are strictly equal to constant proportional factors. A more formal argument of Hicks' composite commodity theorem can be summarized as follows. If all the prices of commodities within group A (p_A) move in exact proportion to a certain common representative price (P_A) with fixed vector of constant (μ), in other words, the variation in the price vector within group is restricted by the equation of $p_A = \mu \cdot P_A$, even though P_A and p_B may vary in an arbitrary, then (a) an aggregated macro-utility function defined over composite commodity can be derived from disaggregated micro-utility functions as $U_\mu(Q_A, q_B) \equiv \max_{q_A}\{\, U(q_A, q_B) \;\mid\; \mu \cdot q_A \leq (y_A / P_A) \,\}$, which has similar properties corresponding to micro-utility functions such as continuity, monotonicity, and quasi-concavity in its arguments, (b) the corresponding property of the continuity from above in both micro- utility and macro-utility functions guarantee the existence of solutions to both micro-optimization and corresponding macro-optimization problems, and (c) the optimization problem based on disaggregated micro-utility functions as $\max_{q_A,\, q_B}\{\, U(q_A, q_B) \mid p_A q_A + p_B q_B \leq y \,\}$ is equivalent to the optimization problem based on aggregated macro-utility function as $\max_{Q_A,\, q_B}\left\{ U_{\mu}\left(Q_A, q_B\right) \mid P_A Q_A + p_B q_B \leq y \right\}$ in terms of equivalence with adjustment by constant proportional factor (μ) between micro-optimization solution of $\left(q_A^*, q_B^*\right)$ and macro-optimization solution of $\left(Q_A^*, q_B^*\right)$ where $Q_A^* = \mu \cdot q_A^* = y_A^* / P_A$. Thus the composite commodity can be defined as either the weighted average of micro-commodities with the vector of proportional factors as weighting scheme or the real expenditure for group commodities deflated by the representative group price index. While the formal proofs for Hicks composite commodity theorem in the consumer context and its application in the producer context can be found in Diewert (1978), this result of Hicks composite commodity theorem can be intuitively understood based on the relationship of $p_A \cdot q_A = \left(\mu \cdot P_A\right) \cdot q_A = P_A \cdot \left(\mu \cdot q_A\right) \equiv y_A \equiv P_A \cdot Q_A$. Similarly the Leontief-composite commodity theorem can also be understood by starting with quantity-proportionality $q_A = (1/\mu) \cdot Q_A$ instead of price-proportionality $p_A = \mu \cdot P_A$ and the intuitive relationship of $p_A \cdot q_A \equiv y_A \equiv P_A \cdot Q_A$ through $p_A \cdot q_A = p_A \cdot \left\langle \, (1/\mu) \cdot Q_A \, \right\rangle = \left\langle \, (1/\mu) \cdot p_A \, \right\rangle \cdot Q_A = P_A \cdot Q_A$.

We can see that the condition of Hicks-Leontief composite commodity theorem $p_A = \mu \cdot P_A$ and/or $q_A = (1/\mu) \cdot Q_A$ is equivalent to the strict form of compositional stability condition (7') $x_{k\,n} = X_k \, a_{k\,k,n}$, $\forall\, k = 1,.....,K$ and $\forall\, n = 1,.....,N$, where either price variables or quantity variables are generalized to any explanatory or right-hand side variables. Given the equivalence between the conditions of Hick-Leontief composite commodity theorem and the strict form of the compositional stability condition, the generalized form of the compositional stability condition can be regarded as a generalization of the conditions of Hick-Leontief composite commodity theorem. In this respect, the generalized form of the compositional stability condition can be compared with Lewbel's generalized composite commodity theorem, which can be regarded as the alternative generalization of Hick composite commodity theorem. The generalized form of compositional stability condition allows some deviations from the strict form of compositional stability condition, as long as such deviation does not cause inconsistency for estimating H_n in $x_n = \; X \, H_n + d_n$. While this generalization maintained non-randomness of proportionality factors $a_{k\,k,\,n}$ $\forall\, k = 1,.....,K$ and $\forall\, n = 1,.....,N$, Lewbel (1996) argues that the differences of the prices of individual commodities and composite commodity price can be allowed to vary as long as these differences are independent of composite commodity price or general rate of inflation of the group. This generalized composite theorem is based on the idea that the differences between individual commodity prices and the aggregate commodity price can be regarded as the aggregation errors and the estimated aggregated parameters can be consistent if these aggregation errors are well behaved so that they can be either included in the intercept term or absorbed into the error term. Lewbel's generalized composite commodity theorem can be understood in the context of Theil's aggregation theory and the compositional stability condition. While Lewbel's theorem requires that macro-variables X be independent of d_n^{Lewbel}, which is defined by further decomposing $x_n = X A_n + v_n$ into $x_n = X + d_n^{Lewbel}$ rather than $x_n = X H_n + d_n$. Or if we assume that either the proportionality factor $a_{kk,n} = 1$ or the constant $c = 1$ in (14) which implies *a priori* condition that the true macro-parameter is a simple sum of the corresponding micro-parameters, then we can obtain Lewbel's consistent aggregation condition from the Theil's aggregation theory framework. This further decomposition as in (7''') makes it possible for us to easily define $d_n^{Lewbel} \equiv x_n - X$ and allows us to avoid difficulty involved in searching for instrumental variables in empirically testing the compositional stability condition of $Cov\left(d_n, X\right) = 0$ in $x_n = X H_n + d_n$.

(7''') $x_n = X A_n + v_n \qquad , \forall n = 1,\ldots\ldots, N ,$

$$\text{where } A_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} + \begin{bmatrix} a_{11,n}-1 & 0 & \cdots & 0 \\ 0 & a_{22,n}-1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{KK,n}-1 \end{bmatrix} + \begin{bmatrix} 0 & a_{12,n} & \cdots & a_{1K,n} \\ a_{21,n} & 0 & \cdots & a_{2K,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1,n} & a_{K2,n} & \cdots & 0 \end{bmatrix}$$

$$\text{so } x_{k,n} = X + X_k (a_{kk,n} - 1) + \sum_{j \neq k}^{K} X_j a_{jk,n} = X + d_{k,n}^{Lewbel} \qquad , \forall k = 1,\ldots\ldots, K \;\; \forall n = 1,\ldots\ldots, N .$$

Lewbel's theorem, however, has following ambiguities: One ambiguity in Lewbel's theorem is how to deal with fact that the Hick-Leontief composite commodity theorem is based on non-randomness of proportionality factors $a_{kk,n}$. Lewbel deals with this difficulty either (a) By restricting his generalized theorem into log-linear model which should absorb non-random part of $d_{k,n}^{Lewbel} \equiv X_k (a_{kk,n} - 1) + \sum_{j \neq k}^{K} X_j a_{jk,n} + v_{k,n}$ into an intercept term in macro-parameter vector of β or (b) By allowing the differences of the prices of individual commodities to the composite commodity price to vary and be absorbed into the random error term. If the first assumption is taken, the macro-model should always have a significant intercept term, which is a complicated mixture of heterogeneous components and thus is difficult to be meaningfully interpreted. If the second assumption is taken, the intuitive rationale of a constant or stable budget constraint condition within each commodity group for the Hick-Leontief composite commodity theorem is lost. Another ambiguity in Lewbel's theorem is that it has the same arbitrariness for the choice of constant $c = 1$ as in Theil's case discussed above (14), because there is no reason not to choose $c = 1/N$, for example. There are no *a priori* reasons that the ratio of observed micro-variables to true macro-variable should be restricted to one. Note that the differences are transformed into ratios in Lewbel's log-linear model. It is convenient either to define aggregation bias as the difference between micro-variables and macro-variables or to avoid the difficulty involved in searching for instrumental variables in empirically testing the compositional stability condition. However, it is restrictive because it implies that the true macro-parameters should be a simple sum of micro-parameters. There is no a prior reason that the true macro-parameters can not be a simple average of micro-parameters, for example. The other ambiguity in Lewbel's theorem is in interpretation of empirically test result of no correlation or no cointegration as independence condition between a pair of two variables, where one is the composite commodity price or the general rate of inflation of the group and the other is the difference between individual commodity prices and the aggregate commodity price or the aggregation bias. Lewbel's theorem is applied for empirical study based on the following basic logic: (a) If two variables are stationary, then a correlation test is conducted, (b) If both variables are nonstationary, a cointegration test is conducted, (c) If one is stationary but the other is nonstationary, then no test is conducted with conclusion that they are not cointegrated because the stationary series can not be cointegrated with the non-stationary series by the algebra of cointegration. If two variables are uncorrelated or not cointegrated, then they are interpreted as independent. Lewbel's empirical testing strategy has following difficulties: (a) Correlation and cointegration are designed for testing linear dependencies. Thus even if independence is not rejected by these two tests, it is still possible that there remains some non-linear dependency, (b) Cointegration is designed for testing dependencies in the long-run. Thus even if cointegration is rejected by either empirical cointegration test or the algebra of cointegration, it is still possible that there remain some mid/short-run and/or contemporaneous dependencies, (c) When micro-variables are nonstationary, it is conceivable that the macro-variable, which is required to be representative of micro-variables and thus closely related to micro-variables, is also nonstationary such that $d_n^{Lewbel} \equiv x_n - X$ is stationary. In this case, an empirical testing strategy based on cointegration might have a tendency to accept the independent condition of Lewbel's theorem by construction.

Compared with the Lewbel's consistent aggregation condition, the generalized form of the compositional stability condition maintains (a) The non-randomness of proportionality factors and thus the intuitive rationale of Hick-Leontief composite commodity theorem and (b) It does not have *a priori* restrictions for true macro-parameters such as simple sum or simple average of micro-parameters in the context of Theil's aggregation theory. (c) It does not invoke ambiguities involving the use of correlation or cointegration test results as an independence test, as in empirical application of Lewbel's theorem, although empirical application of it requires a difficult search for instrumental variables in Hausman type misspecification test of $Cov(d_n, X)=0$ in $x_n = X H_n + d_n$. In this respect, based on the generalized form of the compositional stability condition among disaggregated micro-variables, we can rely on index number theory to decide the proper weighting schemes in aggregation of micro-variables into macro-variables when we have dual pairs of information.

Principal Component Method

The index number approach for deciding weighting schemes in aggregating micro-variables into macro-variables has the advantage that the resulting index number formula does not require parameter estimates. The index number approach, however, requires dual pairs of information and these dual pairs are not always available in all areas of study. For example, even though there exist some efforts to use the Tornqvist-Theil index to obtain monetary aggregates (Barnett, 1984), it is not easy to get this kind of dual pairs in other macro-economic areas. An alternative way to get weighting schemes for dimensional reduction without invoking parameter estimates is to use the multivariate statistical method of principal component analysis.

Principal component analysis has been a major statistical tool to condense large dimensional data into a small number of aggregate variables with as little loss of information as possible in the mean squared error sense. It seeks to reduce the dimension of the data by finding a few linear combinations or principal components of original variables that successively have maximum variance, subject to the restriction that successive principal component are uncorrelated with previous principal components as in (29)

(29) $PC_k \equiv W_k \chi$, where $W_k = \arg\max Var\{W \cdot \chi\}$ $s.t. W \cdot W^T = 1$

and $W_k = \lfloor w_{1k}, w_{2k}, \cdots, w_{Kk}, \cdots w_{KNk} \rfloor, \forall k = 1, 2, \cdots, K, \cdots, KN$,

where χ is $KN \times T$ matrix defined as in (15).

It has been demonstrated that solving such a successive maximizing problem is equivalent to applying the approximations to the second-order summary matrix Σ of data such as a covariance or correlation matrix, which is decomposed by the singular value decomposition theorem. There are several useful properties in this method. (a) When we get as many principal components as the number of the original variables, the total variation of original variables is equal to the total variation of principal components, which is equal to the sum of the eigen-values of the covariance matrix.

(30) $PC_{KN} \equiv Q_{KN}^{T}\chi$, where $\Sigma = Q_{KN} \cdot \Lambda_{KN} \cdot Q_{KN}^{T}$,

$\Lambda_{KN} = diag(\lambda_1, \lambda_2, \cdots, \lambda_K, \cdots, \lambda_{KN})$ where λ_k are descending ordered eigen-values,

$Q_{KN} = diag(e_1, e_2, \cdots, e_K, \cdots, e_{KN})$ where e_k are (KN×1) corresponding eigen-vectors,

and $\sum_{k=1}^{KN} Var(PC_k) = \sum_{k=1}^{KN} \lambda_k = Trace(\Lambda_{KN}) = Trace(\Sigma) = \sum_{k=1}^{KN} Var(\chi_k)$.

(b) The first K principal components can explain most of variance of the original variables so that the rest can be disregarded with minimum loss of information, when the last $K \cdot N - K$ eigen-values are insignificant, i.e. $\lambda_1 > \lambda_2 > \cdots > \lambda_K >> \lambda_{K+1} > \cdots > \lambda_{KN}$. When this is the case, the cumulative proportion of the variance explained by the first K principal components can be calculated by $\sum_{k=1}^{K} \lambda_k \big/ Trace(\Sigma)$.

(31) $PC \equiv Q^T \chi$, where for $\lambda_1 > \lambda_2 > \cdots > \lambda_K >> \lambda_{K+1} > \cdots > \lambda_{KN}$,

$\Sigma = Q_{KN} \cdot \Lambda_{KN} \cdot Q_{KN}^{T} = Q \cdot \Lambda \cdot Q^T + \varepsilon \approx Q \cdot \Lambda \cdot Q^T$,

$\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_K)$ where λ_k are descending order eigen-values,

$Q = diag(e_1, e_2, \cdots, e_K)$ where e_k are (KN×1) corresponding eigen-vectors,

And $\sum_{k=1}^{K} Var(PC_k) \Big/ \sum_{k=1}^{KN} Var(\chi_k) = \sum_{k=1}^{K} \lambda_k \Big/ Trace(\Sigma) = Trace(\Lambda) / Trace(\Sigma)$.

The possibility of the dimensional reduction can be understood as follows using $\Sigma = Q_{KN} \cdot \Lambda_{KN} \cdot Q_{KN}^{T} = Q \cdot \Lambda \cdot Q^T + \varepsilon \approx Q \cdot \Lambda \cdot Q^T$. The first equality of $\Sigma = Q_{KN} \cdot \Lambda_{KN} \cdot Q_{KN}^{T}$ is an application of the singular value decomposition theorem to the positive matrix of second-order data summary matrix Σ such as the covariance or correlation matrix just as in (30). The second equality $Q_{KN} \cdot \Lambda_{KN} \cdot Q_{KN}^{T} = Q \cdot \Lambda \cdot Q^T + \varepsilon$ represents the following further matrix decomposition of the resulting first decomposed matrix by the singular value decomposition theorem. When the last $K \cdot N - K$ eigen-values are insignificant, i.e. $\lambda_1 > \lambda_2 > \cdots > \lambda_K >> \lambda_{K+1} > \cdots > \lambda_{KN}$, the corresponding ε matrix can not be too large to ignore. The third equality $Q \cdot \Lambda \cdot Q^T + \varepsilon \approx Q \cdot \Lambda \cdot Q^T$ represents this approximation where the amount of information loss is represented by the ε matrix. The degree of dimensional reduction from $K \cdot N$ to K depends on the eigen-value structure of $\lambda_1 > \lambda_2 > \cdots > \lambda_K >> \lambda_{K+1} > \cdots > \lambda_{KN}$ i.e. how insignificant of the last $K \cdot N - K$ eigen-values, where the last insignificant $K \cdot N - K$ eigen-values guarantee the amount of information loss ε to be small.

$$\Sigma = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1K} & e_{1\,K+1} & \cdots & e_{1\,KN} \\ e_{21} & e_{22} & \cdots & e_{2K} & e_{2\,K+1} & \cdots & e_{2\,KN} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{K1} & e_{K2} & \cdots & e_{KK} & e_{K\,K+1} & \cdots & e_{K\,KN} \\ e_{K+1\,1} & e_{K+1\,2} & \cdots & e_{K+1\,K} & e_{K+1\,K+1} & \cdots & e_{K+1\,KN} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{KN\,1} & e_{KN\,2} & \cdots & e_{KN\,K} & e_{KN\,K+1} & \cdots & e_{KN\,KN} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \lambda_{K+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \lambda_{KN} \end{bmatrix} \cdot \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{K1} & e_{K+1\,1} & \cdots & e_{KN\,1} \\ e_{12} & e_{22} & \cdots & e_{K2} & e_{K+1\,2} & \cdots & e_{KN\,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{1K} & e_{2K} & \cdots & e_{KK} & e_{K+1\,K} & \cdots & e_{KN\,K} \\ e_{1\,K+1} & e_{2\,K+1} & \cdots & e_{K\,K+1} & e_{K+1\,K+1} & \cdots & e_{KN\,K+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{1\,KN} & e_{2\,KN} & \cdots & e_{K\,KN} & e_{K+1\,KN} & \cdots & e_{KN\,KN} \end{bmatrix}$$

$$= \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1K} \\ e_{21} & e_{22} & \cdots & e_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ e_{K1} & e_{K2} & \cdots & e_{KK} \\ e_{K+1\,1} & e_{K+1\,2} & \cdots & e_{K+1\,K} \\ \vdots & \vdots & \ddots & \vdots \\ e_{KN\,1} & e_{KN\,2} & \cdots & e_{KN\,K} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix} \cdot \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{K1} & e_{K+1\,1} & \cdots & e_{KN\,1} \\ e_{12} & e_{22} & \cdots & e_{K2} & e_{K+1\,2} & \cdots & e_{KN\,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{1K} & e_{2K} & \cdots & e_{KK} & e_{K+1\,K} & \cdots & e_{KN\,K} \end{bmatrix} + \begin{bmatrix} e_{1\,K+1} & e_{1\,K+2} & \cdots & e_{1\,KN} \\ e_{2\,K+1} & e_{2\,K+2} & \cdots & e_{2\,KN} \\ \vdots & \vdots & \ddots & \vdots \\ e_{K\,K+1} & e_{K\,K+2} & \cdots & e_{K\,KN} \\ e_{K+1\,K+1} & e_{K+1\,K+2} & \cdots & e_{K+1\,KN} \\ \vdots & \vdots & \ddots & \vdots \\ e_{KN\,K+1} & e_{KN\,K+2} & \cdots & e_{KN\,KN} \end{bmatrix} \cdot \begin{bmatrix} \lambda_{K+1} & 0 & \cdots & 0 \\ 0 & \lambda_{K+2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{KN} \end{bmatrix} \cdot \begin{bmatrix} e_{1\,K+1} & e_{2\,K+1} & \cdots & e_{K\,K+1} & e_{K+1\,K+1} & \cdots & e_{KN\,K+1} \\ e_{1\,K+2} & e_{2\,K+2} & \cdots & e_{K\,K+2} & e_{K+1\,K+2} & \cdots & e_{KN\,K+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{1\,KN} & e_{2\,KN} & \cdots & e_{K\,KN} & e_{K+1\,KN} & \cdots & e_{KN\,KN} \end{bmatrix}$$

$$= \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1K} \\ e_{21} & e_{22} & \cdots & e_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ e_{K1} & e_{K2} & \cdots & e_{KK} \\ e_{K+1\,1} & e_{K+1\,2} & \cdots & e_{K+1\,K} \\ \vdots & \vdots & \ddots & \vdots \\ e_{KN\,1} & e_{KN\,2} & \cdots & e_{KN\,K} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix} \cdot \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{K1} & e_{K+1\,1} & \cdots & e_{KN\,1} \\ e_{12} & e_{22} & \cdots & e_{K2} & e_{K+1\,2} & \cdots & e_{KN\,2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{1K} & e_{2K} & \cdots & e_{KK} & e_{K+1\,K} & \cdots & e_{KN\,K} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1K} & \varepsilon_{1\,K+1} & \cdots & \varepsilon_{1\,KN} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2K} & \varepsilon_{2\,K+1} & \cdots & \varepsilon_{2\,KN} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{K1} & \varepsilon_{K2} & \cdots & \varepsilon_{KK} & \varepsilon_{K\,K+1} & \cdots & \varepsilon_{K\,KN} \\ \varepsilon_{K+1\,1} & \varepsilon_{K+1\,2} & \cdots & \varepsilon_{K+1\,K} & \varepsilon_{K+1\,K+1} & \cdots & \varepsilon_{1\,KN} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{KN\,1} & \varepsilon_{KN\,2} & \cdots & \varepsilon_{KN\,K} & \varepsilon_{KN\,K+1} & \cdots & \varepsilon_{KN\,KN} \end{bmatrix}$$

(c) The subspace spanned by the first K eigen-vectors has the smallest mean square deviation from original data matrix among all subspaces of dimension of K. (d) If the sample size T is large, then sample eigen-values are consistent estimates of the population eigen-values and sample eigen-vectors and principal components are consistent estimates of orthogonal transformations of their population counterparts, when variable number M ($= K \cdot N$ in our study) is fixes. Heaton and Solo (2006) also show that in a large-M and large-T framework, this conclusion is still valid by showing that the convergence rate is $\sqrt{T}$, which is independent of M. They emphasize that (a) There is no requirement of growing gaps between eigen-values and (b) Increasing variable numbers M does not imply improving estimates.

When we impose certain structure on ε by assuming $Cov\left(\varepsilon_{n}, \varepsilon_{n'}\right)=0$, $\forall n \neq n'$ as in (32), we find that principal components analysis is equivalent with factor analysis, which is another popular multivariate statistical method of factor analysis and whose framework will be summarized in (33) in the connection to the (32).

(32) $\Sigma = Q \cdot \Lambda \cdot Q^{T} + \varepsilon = Q \cdot \Lambda \cdot Q^{T} + \Psi = L \cdot L^{T} + \Psi$

$$
= \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1K} \\ e_{21} & e_{22} & \cdots & e_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ e_{K1} & e_{K2} & \cdots & e_{KK} \\ e_{K+11} & e_{K+12} & \cdots & e_{K+1K} \\ \vdots & \vdots & \ddots & \vdots \\ e_{KN1} & e_{KN2} & \cdots & e_{KNK} \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix} \cdot \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{K1} & e_{K+11} & \cdots & e_{KN1} \\ e_{12} & e_{22} & \cdots & e_{K2} & e_{K+12} & \cdots & e_{KN2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ e_{1K} & e_{2K} & \cdots & e_{KK} & e_{K+1K} & \cdots & e_{KNK} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1K} & \varepsilon_{1K+1} & \cdots & \varepsilon_{1KN} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2K} & \varepsilon_{2K+1} & \cdots & \varepsilon_{2KN} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ \varepsilon_{K1} & \varepsilon_{K2} & \cdots & \varepsilon_{KK} & \varepsilon_{KK+1} & \cdots & \varepsilon_{KKN} \\ \varepsilon_{K+11} & \varepsilon_{K+12} & \ddots & \varepsilon_{K+1K} & \varepsilon_{K+1K+1} & \ddots & \varepsilon_{1KN} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ \varepsilon_{KN1} & \varepsilon_{KN2} & \cdots & \varepsilon_{KNK} & \varepsilon_{KNK+1} & \cdots & \varepsilon_{KNKN} \end{bmatrix}
$$

$$
= \begin{bmatrix} \sqrt{\lambda_1}e_{11} & \sqrt{\lambda_2}e_{12} & \cdots & \sqrt{\lambda_K}e_{1K} \\ \sqrt{\lambda_1}e_{21} & \sqrt{\lambda_2}e_{22} & \cdots & \sqrt{\lambda_K}e_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_1}e_{K1} & \sqrt{\lambda_2}e_{K2} & \cdots & \sqrt{\lambda_K}e_{KK} \\ \sqrt{\lambda_1}e_{K+11} & \sqrt{\lambda_2}e_{K+12} & \cdots & \sqrt{\lambda_K}e_{K+1K} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_1}e_{KN1} & \sqrt{\lambda_2}e_{KN2} & \cdots & \sqrt{\lambda_K}e_{KNK} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\lambda_1}e_{11} & \sqrt{\lambda_1}e_{21} & \cdots & \sqrt{\lambda_1}e_{K1} & \sqrt{\lambda_1}e_{K+11} & \cdots & \sqrt{\lambda_1}e_{KN1} \\ \sqrt{\lambda_2}e_{12} & \sqrt{\lambda_2}e_{22} & \cdots & \sqrt{\lambda_2}e_{K2} & \sqrt{\lambda_2}e_{K+12} & \cdots & \sqrt{\lambda_2}e_{KN2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{\lambda_K}e_{1K} & \sqrt{\lambda_K}e_{2K} & \cdots & \sqrt{\lambda_K}e_{KK} & \sqrt{\lambda_K}e_{K+1K} & \cdots & \sqrt{\lambda_K}e_{KNK} \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_K & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & \psi_{K+1} & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \psi_{KN} \end{bmatrix}
$$

$$
= \begin{bmatrix} L_1 & L_2 & \cdots & \cdots & \cdots & \cdots & L_K \end{bmatrix} \cdot \begin{bmatrix} L_1^T \\ L_2^T \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ L_K^T \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_K & 0 & \cdots & 0 \\ 0 & 0 & \ddots & 0 & \psi_{K+1} & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \psi_{KN} \end{bmatrix}
$$

Factor analysis model is closely related with Theil's aggregation theory. When we use index k and K as micro-variables' group index and total number of groups that should be empirically identified, we see that factors F and residuals ε are equivalent to macro-variables $\aleph$ and disturbances υ by comparing (33) and (18). This is a reason to keep notation $K \cdot N = M$ for the number of original variables and to use the same notation for original data matrix χ and factor loadings L as defined in (18). They are actually the same matrix. We also see that, except for $Cov\left(\varepsilon_{n}, \varepsilon_{n'}\right)=0$ which will be generalized, the equivalence of assumptions between the two methods, because the assumptions of $Cov(\varepsilon, F)=0$ and $E(\varepsilon)=0$ in factor analysis are equivalent to the primary conditions of $Cov\left(v_{n}, X\right)=0$ and $E\left(v_{n}\right)=0$ in Theil's aggregation theory, given the assumptions of $E(F)=0$ and $Cov(F)=I$ can be interpreted as normalizing assumptions. The reason to use different notation for factors F and macro-variables $\aleph$ for the same matrix is to emphasize that factor analysis and Theil's aggregation theory have been developed separately. However, they are closely related with each other and we will show that the possible condition of getting interpretable principal components is also closely related with the compositional stability condition in aggregation theory.

(33) $\chi = L \cdot F + \varepsilon$,

with assumptions of $Cov(\varepsilon, F) = 0$, $E(\varepsilon) = 0$,

$Cov(F) = E(F \cdot F^T) = I$, $E(F) = 0$, and $Cov(\varepsilon) = E(\varepsilon \cdot \varepsilon^T) = \Psi$ where Ψ is diagonal matrix,

so that $Cov(\chi) = \Sigma = L \cdot E(FF^T) \cdot L^T + E(\varepsilon\varepsilon^T) + L \cdot E(F\varepsilon^T) + E(\varepsilon F^T) \cdot L^T = L \cdot L^T + \Psi$.

(18) $\chi = L \cdot \aleph + \upsilon$,

with assumptions of $Cov(\upsilon, \aleph) = 0$ and $E(\upsilon) = 0$,

where χ and υ are $KN \times T$ matrix and L is $KN \times K$ and $\aleph$ is $K \times T$ matrix.

Factor analysis is based on the idea that when there are co-movements among original variables, it is conceivable that this co-movement is due to their partial dependences on the common latent components such that common factors can capture all the dependence among variables, leaving no cross correlations in the residuals. Standard factor analysis is explicitly based on this structural assumption so that the data admit a factor structure or a common-idiosyncratic decomposition among original variables. While factor analysis based on the maximum likelihood estimation method or state space method requires parameter estimation, principal component analysis based on the singular value decomposition theorem has the advantage that it does not require such parameter estimation. In this respect, the possibility of relaxing the assumption of $Cov(\varepsilon_n, \varepsilon_{n'}) = 0$, $\forall n \neq n'$ and of connecting principal component analysis to factor analysis has been studied. Chamberlain and Rothchild (1983) and Connor and Korajczyk (1986) introduce the approximate factor model to allow a non-diagonal covariance matrix such that $Cov(\varepsilon_n, \varepsilon_{n'}) < \delta$ where δ is a small value and show that the principal component method is equivalent to factor analysis when the number of variables M increases to infinity. Note that the standard and approximate factor model also assumes that factors affect individual variables at contemporaneous time only. To relax this rather strong assumption for time-series data, the distributed lag effect of factors on individual variables is also introduced. In this dynamic setting, two approaches, commonly called as the dynamic factor model, are suggested to generalize the standard covariance or correlation matrix. While Forni et al (2000) use the spectral density matrix in a frequency-domain framework, Stock and Watson (2002) use cross-covariance matrix, which includes auto-covariance matrix in a time-domain framework. Since both approaches apply the singular value decomposition theorem to their generalized covariance or correlation matrix to derive eigen-vectors as weighting schemes, the dynamic factor model can be understood as the generalized approximate factor model based on the generalized principal component method. Forni and Lippi (2001), similar to Chamberlain and Rothchild (1983) but in the dynamic setting, shows that K -factor representation exists iff the first K eigen-values of the spectral density matrix are unbounded, while other eigen-values are bounded as the number of variables M increases to infinity. Stock and Watson (2002) also shows that principal component of the covariance matrix converge in probability to the true factors up to a sign change. In terms of bounding condition of cross-correlation of residuals $Cov\left(\varepsilon_n, \varepsilon_{n'}\right) < \delta$ for the equivalence of principal component method to factor analysis method, Heaton and Solo (2006) shows that while the condition of Chamberlain and Rothchild (1983) or Forni and Lippi (2001) is the bounding condition of maximum eigen-value of residual covariance matrix in the static or dynamic setting respectively, the condition of Stock and Waston (2002), Bai and Ng (2002) and Bai (2003) is the bounding condition of maximum row sum of residual covariance matrix. They also demonstrate that these bounding conditions can be allowed to relaxed, provided that the growth rate of maximum eigen-value is $M^{1-\alpha}$, where $0 \leq \alpha \leq 1$ and the growth rate of maximum row sum is strictly less than M , where M is the number of original variables. Given that the maximum eigen-value is always less than or equal to the maximum row sum of residuals, this means that the sample principal components estimator converge to latent population factors, as long as the number of strongly correlated residuals grows at a rate strictly less than the number of original variables, although the higher is the growth rate, the slower is the convergent rate. Based on these result, we can interpret principal component analysis as one factoring method of the covariance or correlation matrix for the factor analysis model in general conditions.

However, as Heaton and Solo (2006) emphasize, not only the number of variables but also the data structure itself should be the primary issue in using principal component analysis. The importance of the data structure can be understood based on following two extreme cases, whose principal components are expressed in simple and extreme forms (Johnson and Wichern, 1988).

$$(34)\ \Sigma^0 = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{KN}^2 \end{bmatrix} \text{or} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (35)\ \Sigma^H = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} \text{or} \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$$

When the original variables are perfectly uncorrelated with each other, so the covariance or correlation matrix is the diagonal matrix Σ^0 as in (34), eigen-values and eigen-vector become all equal as in $(34')$ and thus the corresponding eigen-vector as a weighting scheme results in just the original set of variables. So there is nothing to gain by using principal component method in terms of dimensional reduction.

$$(34')\ \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

When original variables are equally correlated with each other, so covariance or correlation matrix has the specific structure Σ^H as in (35), (a) The first eigen-value becomes $\lambda_1 = 1-(M-1)\rho = M\rho + (1-\rho)$ with eigen-vector $e_1^T = \left\lfloor 1/\sqrt{M}, 1/\sqrt{M}, \cdots, 1/\sqrt{M} \right\rfloor$ and the remaining eigen-values become $\lambda_2 = \lambda_3 = \cdots \lambda_M = 1-\rho$ with some convenient choice for eigen-vectors as in $(35')$, (b) The first principal component becomes proportional to the simple sum of the original variables with proportional factor of $1/\sqrt{M}$, and (c) The first principal component explains the total variation of original variables by the following proportion: $\dfrac{Var(PC_1)}{\sum_{k=1}^{M} Var(\chi_k)} = \dfrac{\lambda_1}{\sum_{k=1}^{M}\lambda_k} = \dfrac{\lambda_1}{M\rho + M(1-\rho)} = \dfrac{\lambda_1}{M} = \dfrac{M\rho+(1-\rho)}{M} = \rho + \dfrac{1-\rho}{M} \approx \rho$. When the equal correlation ρ is close to 1 or the variable number M is large, the first principal component explains almost all the variation of original variables. So the first principal component is the perfect representative aggregate in terms of dimensional reduction purpose.

$$(35')\quad \begin{bmatrix} 1 & \rho & \cdots & \rho & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 & \rho & \cdots & \rho \\ \rho & \rho & \cdots & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & \rho & \rho & \cdots & 1 \end{bmatrix} = \left[\begin{array}{c:c:c:c:c:c:c} 1/\sqrt{M} & 1/\sqrt{1\cdot 2} & 1/\sqrt{2\cdot 3} & \cdots & 1/\sqrt{(k-1)\cdot k} & \cdots & 1/\sqrt{(M-1)\cdot M} \\ 1/\sqrt{M} & -1/\sqrt{1\cdot 2} & 1/\sqrt{2\cdot 3} & \cdots & \vdots & \cdots & 1/\sqrt{(M-1)\cdot M} \\ \vdots & 0 & -1/\sqrt{2\cdot 3} & \cdots & 1/\sqrt{(k-1)\cdot k} & \cdots & \vdots \\ 1/\sqrt{M} & 0 & 0 & \cdots & -1/\sqrt{(k-1)\cdot k} & \cdots & \vdots \\ \vdots & 0 & \vdots & \cdots & 0 & \cdots & \vdots \\ 1/\sqrt{M} & \vdots & \vdots & \cdots & \vdots & \cdots & 1/\sqrt{(M-1)\cdot M} \\ 1/\sqrt{M} & 0 & 0 & 0 & 0 & \cdots & -1/\sqrt{(M-1)\cdot M} \end{array}\right]$$

$$\times \begin{bmatrix} M\cdot\rho+1-\rho & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1-\rho & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-\rho & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1-\rho & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1-\rho \end{bmatrix} \times \left[\begin{array}{ccccccc} 1/\sqrt{M} & 1/\sqrt{M} & \cdots & 1/\sqrt{M} & \cdots & 1/\sqrt{M} & 1/\sqrt{M} \\ \hdashline 1/\sqrt{1\cdot 2} & -1/\sqrt{1\cdot 2} & 0 & 0 & \cdots & 0 & 0 \\ \hdashline 1/\sqrt{2\cdot 3} & 1/\sqrt{2\cdot 3} & -1/\sqrt{2\cdot 3} & 0 & \cdots & 0 & 0 \\ \hdashline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hdashline 1/\sqrt{(k-1)\cdot k} & \cdots & 1/\sqrt{(k-1)k} & -1/\sqrt{(k-1)k} & 0 & \cdots & 0 \\ \hdashline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hdashline 1/\sqrt{(M-1)M} & 1/\sqrt{(M-1)M} & \cdots & \cdots & \cdots & 1/\sqrt{(M-1)M} & -1/\sqrt{(M-1)M} \end{array}\right]$$

These two extreme cases imply that not only the number of variables but also the data structure itself should be the primary issue to be considered in using principal component approach. For example, when we add a sufficient number of idiosyncratic variables which are not correlated with each other as well as with previously formed homogeneous groups of variables, i.e. add variables with data structure of Σ^0 to the variables with data structure of Σ^H, we can create a situation where more data, through the increasing number of variables, might be undesirable, because the average common component will become smaller and/or the residual cross-correlation will eventually become larger. This implication is consistent with the Boivin and Ng (2003)'s simulation and empirical results that expanding the dataset by adding more variables without considering data structure can be not always desirable in terms of forecasting performance of dynamic factor model.

It have been demonstrated that the approximate factor model, especially the dynamic factor model can improve forecasting performance in many economic areas (see Bai, 2003 and references in there for examples). Although it might be not important to obtain interpretable principal components for forecasting purposes, interpretation of principal components has been major issue in the multivariate statistical analysis. Traditional approaches for the interpretation of extracted principal components use either factor loading of components for original variables or correlation between original variables and components. The extracted principal components are interpreted based on the original variables with high loadings or high correlation values. Although large loadings and large correlations often go together, this is not necessarily true (Al-Kandari and Jolliffe, 2001).

Choosing a subset of the original variables that best approximate the information in the extracted principal components and using such a subset to interpret the extracted principal components is another way of interpretation, which dates back at least to Jolliffe (1972). In this respect, Al-Kandari and Jolliffe (2001, 2005) review various methods, including McCabe (1984)'s principal variables approach as well as traditional procedures, for choosing subsets of original variables to approximate and interpret the extracted principal components, using real data sets from various areas as well as simulation data sets that are generated such that the variables are allocated to a few clusters with various strengths of correlations between clusters and different factor loading structures at each level of correlation between clusters. After evaluating various procedures in terms of various efficiency criteria, they conclude that (a) The traditional procedure in interpreting a principal component in terms of only those variables that have high loadings in the component is not always successful in retaining the best variables for the purpose of reducing the dimensionality, or aiding interpretation of the component of interest. (b) The method for retaining the best subsets is often the cluster criterion, which is mainly based on allocating the original variables to clusters using the average-linkage method and then retaining one variable from each cluster. Although they choose an original variable rather than a principal component to represent each cluster and their results can vary depending on the choice of different clustering algorithms, their results imply that we need to use some grouping method before extracting principal components, rather than using traditional method based on factor loadings after extracting principal components from the entire dataset.

The fundamental motive of seeking interpretable principal components can be understood by the following explanation of Johnson and Wichern (1988) with some modification of sentence orders for clarification. "A principal component analysis is concerned with explaining the variance-covariance structure through a few linear combinations of the original variables. Its general objectives are (1) data reduction, and (2) interpretation. … Analyses of principal components are more of a means to an end … because they frequently serve as intermediate steps in much larger investigations. For example, principal components may be inputs to a multiple regression. … The essential purpose of factor analysis is to describe, if possible, the covariance relationships among many variables in terms of a few underlying, but unobservable, random quantities called factors. … Factor analysis can be considered as an extension of principal component analysis. Both can be viewed as attempts to approximate the covariance matrix. However, the approximation based on the factor analysis model is more elaborate. … Basically the factor model is motivated by the following argument. Suppose variables can be grouped by their correlations. That is, all variables within a particular group are highly correlated among themselves but have relatively small correlations with variables in a different group. It is conceivable that each group of variables represents a single underlying construct, or factor, that is responsible for the observed correlations. … For example, correlations from the group of test scores in classics, french, english, mathematics, and music collected by Spearman suggested an underlying intelligence factor. A second group of variables, representing physical-fitness scores, if available, might correspond to another factor. It is this type of structure that factor analysis seeks to confirm. The primary question in factor analysis is whether the data are consistent with a prescribed structure (Johnson and Wichern, 1988, page 340 and 378-379).” Given that modern origins of principal component and factor analysis lie in the early twentieth-century attempts of Karl Pearson, Charles Spearman and others to define and measure intelligence for the subsequence structural analysis, the fundamental purpose is to get the interpretable common latent factors among original variables by using dimensional reduction method of principal component estimator. And its possible condition can be the special type of correlation structure such that all variables within a particular group are highly correlated among themselves but have relatively small correlations with variables in different groups.

This special structure of covariance or correlation can be also understood as the approximate combinations of the two extreme correlation structures discussed in (34) and (35). If variables can be grouped based on their correlations such that variables in different groups have the first extreme type of perfectly uncorrelated structure Σ^0 and variables within a particular group have the second extreme type of equally correlated structure Σ^H as in (36), then it is possible to extract the almost perfect representatives and the meaningfully interpretable aggregates by applying principal component method to each of homogeneous group separately, rather than applying it to the entire group of heterogeneous variables as in (37).

$$(36)\ \Sigma = Corr(\chi) \text{ or } DynCorr(\chi) = \begin{bmatrix} \Sigma_{11}^H & 0 & \cdots & 0 \\ 0 & \Sigma_{22}^H & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{KK}^H \end{bmatrix} \approx \begin{bmatrix} \Sigma_{11} & 0 & \cdots & 0 \\ 0 & \Sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{KK} \end{bmatrix},$$

$$\text{where } \Sigma_{kk}^H = \begin{bmatrix} 1 & \rho_k & \cdots & \rho_k \\ \rho_k & 1 & \cdots & \rho_k \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k & \rho_k & \cdots & 1 \end{bmatrix} \text{ and } \Sigma_{kk} = \begin{bmatrix} 1 & \rho_{k,1\,2} & \cdots & \rho_{k,1\,N} \\ \rho_{k,2\,1} & 1 & \cdots & \rho_{k,2\,N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k,N\,1} & \rho_{k,N\,2} & \cdots & 1 \end{bmatrix},\ \forall k = 1,.....,K .$$

$$(37)\quad \begin{bmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_K \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & \cdots & 0 \\ 0 & L_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & L_{KK} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_K \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_K \end{bmatrix} \text{ or } \chi_k = L_{kk} \cdot F_k + d_k \text{ , } \forall k = 1,.....,K \text{ .}$$

Obviously these two extreme types and the combinations of them are too strong for the real world data. However, it is conceivable that this combination of two extreme types of correlation structures can be approximated by the special type of structure mentioned above. If the original variables can be grouped into this kind of special data pattern empirically, principal components applied to each homogeneous group separately can be an almost perfect representative in terms of dimensional reduction purpose. This implies for the principal component approach that when the approximate block diagonal structure in static or dynamic correlation matrix is identified, estimating the principal components from each homogenous group of variables $\chi_k = L_{kk} \cdot F_k + d_k$, $\forall k = 1,.....,K$ can be better than estimating those from the entire data set $\chi = L \cdot F + \varepsilon$ to attain the dimensional reduction purpose with less information loss as well as to obtain the meaningfully interpretable aggregate variables. The near homogeneity of original variables within a specific group makes it possible to provide a meaningful interpretation to this near perfect representative aggregate. Since the main difficulty of interpreting principal components is due to the fact that each of principal components is a linear combination of "all" original variables, using cluster method to define homogeneous subset of variables before extracting principal components is an intuitive solution to achieve interpretable principal components.

The subsequent analyses of studying relationship among aggregate variables also can be justified to understand the relationship among disaggregate variables, since (a) The estimated principal components extracted from each homogenous group of variables can be legitimate representative for the disaggregate variable, and (b) The special type of a block diagonal correlation structure derived from statistical dimensional reduction methods is equivalent to the approximate form of the compositional stability condition obtained from Theil's aggregation theory. To clarify this relationship, the equations $(7'')$, (18), and (33) are recalled.

(33) $\chi = L \cdot F + \varepsilon$,

with assumptions of $Cov(\varepsilon, F) = 0$, $E(\varepsilon) = 0$,

$Cov(F) = E(F \cdot F^T) = I$, $E(F) = 0$, and $Cov(\varepsilon) = E(\varepsilon \cdot \varepsilon^T) = \Psi$ where Ψ is diagonal matrix,

so that $Cov(\chi) = \Sigma = L \cdot E(FF^T) \cdot L^T + E(\varepsilon\varepsilon^T) + L \cdot E(F\varepsilon^T) + E(\varepsilon F^T) \cdot L^T = L \cdot L^T + \Psi$.

(18) $\chi = L \cdot \aleph + \upsilon$,

with assumptions of $Cov(\upsilon, \aleph) = 0$ and $E(\upsilon) = 0$,

where χ and υ are $KN \times T$ matrix and L is $KN \times K$ and $\aleph$ is $K \times T$ matrix.

or $\chi = L \cdot \aleph + \upsilon = BDiag(L) \cdot \aleph + d$

or $\begin{bmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_K \end{bmatrix} = \begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1K} \\ L_{21} & L_{22} & \cdots & L_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ L_{K1} & L_{K2} & \cdots & L_{KK} \end{bmatrix} \begin{bmatrix} \aleph_1 \\ \aleph_2 \\ \vdots \\ \aleph_K \end{bmatrix} + \begin{bmatrix} \upsilon_1 \\ \upsilon_2 \\ \vdots \\ \upsilon_K \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & \cdots & 0 \\ 0 & L_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & L_{KK} \end{bmatrix} \begin{bmatrix} \aleph_1 \\ \aleph_2 \\ \vdots \\ \aleph_K \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_K \end{bmatrix}.$

(7'') $x_n = X A_n + v_n = X \cdot Diag(A_n) + d_n = X H_n + d_n$

or $[x_{1n}, x_{2,n}, \cdots, x_{Kn}] = [X_1, X_2, \cdots, X_K] \begin{bmatrix} a_{11,n} & 0 & \cdots & 0 \\ 0 & a_{22,n} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{KK,n} \end{bmatrix} + [d_{1,n}, d_{2,n}, \cdots, d_{K,n}]$.

Given that the factors F and residuals ε are equivalent to macro-variables $\aleph$ and disturbances υ respectively in (33) and (18), the equivalence between $x_n = X A_n + v_n$ and $\chi = L \cdot \aleph + \upsilon$ implies the equivalence between $x_n = X \cdot Diag(A_n) + d_n = X H_n + d_n$ and $\chi = BDiag(L) \cdot \aleph + d$ in the equation (18) and (7'') , where $Diag(A_n)$ denotes a diagonal matrix of A_n and $BDiag(L)$ denotes a block diagonal matrix of L. The strict form of compositional stability condition $x_n = X \cdot Diag(A_n)$ implies the block diagonal structure in standard correlation matrix $\Sigma = Corr(\chi)$. This suggests that by identifying the approximate block diagonal structure in static or dynamic correlation matrix $\Sigma = Corr(\chi)$ or $DynCorr(\chi)$, we can infer the approximate form of compositional stability condition $x_n = X\ Diag(A_n) + d_n = X H_n + d_n$ with $Cov(d_k, d_{k'}) \leq \delta$, $\forall k \neq k'$, which is equivalent to $\chi_k = L_{kk} \cdot F_k + d_k$, $\forall k = 1, \ldots\ldots, K$ in terms of the factor analysis framework.

Based on the special block diagonal correlation matrix, an interpretable principal component can be obtained by applying principal component approach onto each of homogenous group of variables. Given the equivalence between the principal component approach and the factor analysis method, which in turn is equivalent to the auxiliary equations in the Theil's aggregation theory framework, the approximate form of the compositional stability condition provides not only the possibility of obtaining common principal component or macro-variable as the representative aggregate of homogeneous micro-variables but also the possibility of getting interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for the subsequence analysis. In this respect, we can rely on the principal component method to decide the proper weighting schemes in aggregation of micro-variables into macro-variables with as little loss of information as possible in the mean squared error sense, when we do not have dual pairs of information, which the index number theory is based on to derive the proper weighting schemes in aggregating variables.

Summary and Proposed Method

At the beginning of this study, we suggest to interpret theory as an inductive causal averaging procedure to deal with two methodological issues of how to infer the causal structure from empirical regularities and how to incorporate the large information set into empirical model. When we follow an inductive causal averaging procedure that concentrates only on similar tendencies to highlight a few common factors by ignoring many more individual differences and idiosyncrasies, we need to identify empirically justifiable conditions that allow us to legitimately define common tendencies and individual idiosyncrasies. Based on the generalized condition for the consistent aggregation, we propose one possible methodological procedure to consistently address the two related issues of causal inference and actual aggregation procedures for the full use of research potentials brought by high dimensional data.

To address the issue of how to infer the causal structure from empirical regularities, the graphical causal models, which are empirically implemented by using either PC algorithm or GES algorithm, can be used to inductively infer causal structure from non-temporal and non-experimental data. However, the (probabilistic) stability condition for the graphical causal models can be violated for high dimensional data, when close co-movements and thus near deterministic relations exist among variables in high dimensional data. One possible way to address this issue is using aggregation methods to infer causal relationship among disaggregate variables based on aggregated variables. The aggregation method is also helpful to address another issue of how to incorporate the large information set into empirical model, given that econometric considerations, such as degrees-of-freedom and multicollinearity, require the economy of parameters in empirical models. The weighting schemes to aggregate disaggregate micro-variables into aggregate macro-variable can be empirically decided, based on either index number theory or principal component approach. However, the actual aggregation procedures or decisions on weighting schemes require the legitimate classifications or sufficient conditions for the interpretable and consistent aggregation. In this respect, identifying legitimate aggregation conditions is the main topic to be discussed for both causal inference and actual aggregation.

We studied possible legitimate conditions for the interpretable and consistent aggregation based on both aggregation theory framework and statistical dimensional reduction methods with minimizing any deductive assumptions such as micro-homogeneity of micro-parameters, separability, and homogeneity of utility (production) function. From both the aggregation theory and the statistical dimensional reduction methods, we identify the same generalized forms of the compositional stability condition. When generalized forms of the compositional stability condition can be identified in data set by grouping micro-variables based on their correlation or covariance matrix, there exist not only the possibility of obtaining interpretable common factors or macro-variables as the representative aggregate of homogeneous micro-variables but also the possibility of getting interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for the subsequence analysis. This means that when the micro-variables can be legitimately grouped and represented by macro-variables, it is possible to use aggregation methods to capture micro-relations through macro-relations as the legitimate representatives, where micro-relation or macro-relation can be causal relations. In this respect, we argue that the (probabilistic) stability condition for an "inductive causal" procedure requires the compositional stability condition for an "inductive averaging" procedure.

More specific procedure we propose is as follows; (a) Both standard static correlation matrix and dynamic correlation matrix over identified frequency band are used to measure co-movement among original variables. Based on these similarity measure of disaggregate micro-variables, the modified k-nearest neighbor algorithm is used to sort the variables such that the highly correlated variables are near each other along the main diagonal in reordered correlation matrix. The block-diagonal pattern of reordered or sorted static and dynamic correlation matrixes are used to identify homogeneous group of variables, based the approximate form of the compositional stability condition. (b) Based on identified classifications of original variables, index number theory or statistical dimensional reduction methods are used for actual aggregation procedure to decide weighting schemes for aggregating disaggregated micro-variables into representative macro-variables within each identified group. When we have dual pairs of price and quantity or analogues information, we can use the index number theory to decide the weighting schemes. When such dual pairs of data are not available, principal component method applied onto each of groups is used as the best dimensional reduction method with as little loss of information as possible in the mean squared error sense. (c) The identified classification and aggregation of micro-variables into macro-variables can be tested, as long as appropriate instrumental variables can be identified. The Hausman type misspecification test of $H_0 : \gamma_n = 0$ in $x_n = X H_n + IV \cdot \gamma_n + \varepsilon_n^{IV}$, where x_n and X are micro- and macro-variables respectively and IV are Instrumental Variables such that IV is closely correlated with X and independent of d_n, provides statistical test framework for the generalized form of the compositional stability condition of independence between d_n and X in the set of equations $x_n = X H_n + d_n$. (d) Given the observed phenomena of close co-movements and thus near deterministic relations among variables in high dimensional data, it is conceivable and oftentimes observed that the (probabilistic) stability condition for the graphical causal models is violated for using high dimensional data in empirical study. When this is the case, based on the compositional stability condition, it is still possible to infer causal structures among micro-variables through relationships among representative aggregated macro-variables. It is possible because micro-relations including causal relationships can be legitimately captured by the macro-relations incorporated by the aggregation methods as long as the compositional stability conditions hold among micro-variables. The PC algorithm or GES algorithm is used to infer causal structures among macro-variables as the legitimate representative causal relationships among micro-variables are used for the subsequent analysis.

The inductively inferred causal structures is crucial for subsequent empirical studies, since causal structures are underdetermined by empirical-statistical properties (induction problem) and theory often-times does not provide sufficient or conclusive information for this induction problem. Subsequent analyses are sensitive to the causal structure in the form of pre-classification of dependent and independent variables and other forms of identification problem. The empirically justifiable classification and aggregation are also important for the full use of research potentials brought by high dimensional data in the subsequent empirical studies, given that econometric considerations, such as degrees-of-freedom and multicollinearity, require the economy of parameters in empirical models. Note that inductive properties are emphasized in every sequence of the proposed method, since any types of deductive properties can bring subjectivities or ambiguities into the empirical results. While theory as the inductive causal averaging procedure can allow some deductive elements in its developments, empirical methodologies need to be based more on inductive properties to maintain their objectivity. The remaining subjectivities in our proposed method are left as further research issues, with the hope that the remaining subjectivities bring fewer ambiguities relative to previously used methods. The proposed method is illustrated with the applications for retail checkout scanner data and macro-economic time series panel data as examples of two sets of high dimensional data.

CHAPTER III
USE OF HIGH DIMENSIONAL PANEL DATA IN MICRO-ECONOMETRICS

The study of consumer behavior has a long history and is one of the most studied areas in economics. The demand analysis has significantly advanced from both theoretical and empirical perspectives. However, there remain three methodological issues in applying the micro-economic consumer theory for empirical study of consumer behavior, especially using the retail checkout scanner data. When to understand and measure responsiveness of consumer behavior to changes of exogenous variables can be considered as one of the main objectives of the study of consumer behavior, the empirical measure of responsiveness of consumer behavior to changes in exogenous variables relies on three specification choices in an empirical model. First, given that there are full spectrums of direct, inverse, and mixed demand systems and the general relationship between elasticity and flexibility is not yet established, the measure depends on the relative predeterminess among the price and quantity variables represented by dependent and explanatory variables in an empirical model of a specific commodity. Second, given that small departures from valid classification and/or aggregation can result in large mistakes in empirical results, the measure depends on the classification and aggregation to define price and quantity variables themselves. For example, the decision on classification and aggregation can substantially affect the conclusions about elasticity estimates in multi-stage budgeting approach because cross-price elasticity or cross-quantity flexibility between products in different groups is likely to be small by construction. Third, given that the different assumptions used to parameterize functional relationships have different implications, the measure depends on the functional form to relate dependent variable with explanatory variables. For example, there are four combinations of constant or variation assumptions for the income (or scale) coefficient and Slutsky (or Antonelli) coefficient in the differential functional form approach as captured in Rotterdam, LA/AIDS, CBS, NBR specifications.

In this chapter, we propose an inductive empirical method to address these three methodological issues in the study of consumer behavior based on the discussion on the causality and aggregation issues in chapter II. The way to incorporate theoretical implications into empirical model specifications through the functional forms and the way to compare different specifications of direct, inverse, and mixed demand functions are the additional issues to be addressed. More specifically, first, the specification choice issue among direct, inverse, and mixed demand functions is addressed by using the inductively inferred causal information based on the graphical causal models. Second, the classification and aggregation issue are addressed by the compositional stability conditions and index number theory. Third, the functional form issue is addressed by the synthetic model approach based on the differential functional form framework. And the comparison of alternative specifications is conducted in terms of model selection framework. The proposed method is illustrated with the applications for soft drink products using retail checkout scanner data.

Theoretical Considerations

Causality in Study of Consumer Behavior

One of the main objectives of the study of consumer behavior is to understand and measure responsiveness of consumer behavior to changes in exogenous variables. Responsiveness is measured by elasticities or flexibilities, where the elasticity (or flexibility) is defined by the percentage change in quantity demanded (or willingness to pay) resulting from a 1-percent increase in an exogenous variable. Elasticities are directly measured based on the direct demand function, expressing quantities as a function of price. On the other hand, flexibilities are directly measured based on the inverse demand function, expressing (normalized) prices as a function of quantities. Given that the general relationship between elasticity and flexibility is not yet established, the empirical measure of responsiveness of consumer behavior to changes in an exogenous variable relies on the relative predeterminess among the price and quantity variables represented by dependent and explanatory variables in an empirical model of a particular commodity. In many empirical studies of consumer behavior, the choice of individual direct or inverse demand function is usually based on researchers' intuition about product properties or market characteristics of a specific commodity. A typical argument for predeterminess of price relies on price-taking agent assumption, short-run fixity in prices, or administratively setting of price in publicly offered goods. A typical argument for predeterminess of quantity relies on fixed biological lags in production and non-storable fixed supply of commodities in agricultural commodities, or Bertrand type strategic pricing rules of suppliers in differentiated good.

In general, the choice of quantity-dependent demand function relies on the elastic supply condition and the choice of price-dependent demand function relies on the inelastic supply condition. In this respect, the choice issue of direct or inverse demand function can be addressed by using the full simultaneous equations approach, where demand and supply equations are simultaneously estimated and each is identified by the appropriate instrumental variables such as demand and supply shifters. However, this approach is rarely pursued in empirical work, due to major difficulties to find appropriate instrumental variables needed to identify demand and supply equations of all the related commodities. Furthermore as Thurman (1986) argues, the practical equivalence of the two demand specifications of direct or inverse demand function in a simultaneous equations model does not carry over to models which are not fully simultaneous. In particular, he argues that the choice of dependent variable is crucial to econometric estimation and to economic interpretation in models where demand adjusts to current shocks but supply does not.

Instead of using full simultaneous equations approach, the system-wise approach has been widely used to study interrelationship among related commodities demanded. However, given that most empirical specification of demand systems constitute a monotone set of either direct or inverse demand equations, the commonly used (monotonic) system-wise approach has some limitations, since it might be too restrictive to assume *a priori* that all of related goods have the same characteristics. Depending on the market characteristics of a particular commodity, some demand functions need to be specified as quantity-dependent and others as price-dependent. In this respective, the mixed demand system, expressing demand relationships as a function of mixed set of prices and quantities can be used to provide a flexible way to incorporate the possible combination of quantity-dependent and price-dependent specifications within a system. It is also argued that the mixed demand system also provide the possibility of sidestepping the estimation of both demand and supply functions in a full simultaneous equation framework (Moschini and Vissa, 1993).

The mixed demand function is first proposed in the context of studying market equilibrium with some rationed commodities (Samuelson, 1965). It was then theoretically elaborated by in the context of demand theory by showing the equivalence between the compensated mixed demand function and the compensated rationed demand function (Chavas, 1984). While it has been empirically used in a Rotterdam functional form (Moschini and Vissa, 1993), it is also extended to a generalized Rotterdam functional form (Matsuda, 2004). The mixed demand function not only provides an alternative way to study interrelationship among related commodities demanded without sacrificing the theory of consumer behavior, but also makes it possible to derive some relationships between elasticity and flexibility by extending arguments of Moschini and Vissa (1993). Given that the relationship between elasticity and flexibility is not yet established in general, these relationships can be helpful to understand different implications of direct, inverse, and mixed demand systems.

The three specifications of direct, inverse, and mixed demand functions are rarely discussed in one place, so it is worthwhile to summarize these in terms of the properties of each demand system. Let the set of commodities of interest $A \cup B = \{1, \cdots, m, m+1, \cdots, N\}$ be divided into quantity-dependent $A = \{1, \cdots, m\}$ and price-dependent $B = \{m+1, \cdots, N\}$ commodity groups. The subscripts $(n, n') \in A \cup B$, $(i, j) \in A$, and $(k, r, s) \in B$ are used to denote whole and each group of commodities respectively. Total expenditure and the normalized prices can be represented by $y \equiv P \cdot Q \equiv P_A \cdot Q_A + P_B \cdot Q_B$ and $\pi_n = p_n / y$ respectively. The superscript c is used for compensation and D, I, and M are used for direct, inverse, and mixed demand systems respectively. Following functions play a crucial role in consumer theory as Chavas (1984) summarizes.

· The direct utility function $U(q)$, which is continuous, increasing and quasi-concave in q.

· The indirect utility function $V(p, y)$, which is continuous, decreasing and quasi-convex in p.

· The cost or expenditure function $C(p,u)$, which is continuous, increasing in u,

and increasing, linear homogenous and concave in p.

· The distance or transformation function $D(q,u)$, which is continuous, increasing in u,

and increasing, linear homogenous and concave in q.

· The restricted or rationed cost function $C^R(p_A, q_B, u)$, which is continuous, increasing in u,

increasing and concave in p_A, decreasing and convex in q_B, linear homogenous in p_A.

Note that these functions have the duality relationships, so it is possible to construct any one of the four functions from any other function.

These functions and their properties are used to derive direct, inverse, and mixed demand functions and their properties.

· $C(\pi, V) = 1 \rightarrow V(\pi, 1)$ and $D(q, U) = 1 \rightarrow U(q)$

implying that the indirect and direct utility function can be obtained by inverting the cost function and distance function respectively. Each direct, inverse, and mixed demand system can be derived as follows.

$\cdot V(\pi,1) = \underset{q}{Max}\{U(q) \mid \pi \cdot q = 1\}$ or $V(p,y) = \underset{q}{Max}\{U(q) \mid p \cdot q = y\}$,

where the solution $q(\pi,1)$ or $q(p,1)$ is the vector of uncompensated direct demand functions.

$\cdot U(q) = \underset{\pi}{Min}\{V(\pi,1) \mid \pi \cdot q = 1\}$,

where the solution $\pi(q,1)$ is the vector of uncompensated (normalized) inverse demand functions.

$\cdot V^M(p_A, q_B, y) = \underset{q_A,\, p_B}{Max}\{U(q_A, q_B) - V(p_A, p_B, y) \mid p_A q_A + p_B q_B = y\}$

or $V^M(\pi_A, q_B, 1) = \underset{q_A,\, \pi_B}{Max}\{U(q_A, q_B) - V(\pi_A, \pi_B) \mid \pi_A q_A + \pi_B q_B = 1\}$,

where two solutions $q_A(p_A, q_B, y) = q_A(\pi_A, q_B, 1)$ and $\pi_B(\pi_A, q_B, 1) \cdot y = p_B(p_A, q_B, y)$ are the uncompensated quantity-dependent and price-dependent mixed demand functions respectively.

$\cdot C(\pi,u) = \underset{q}{Min}\{\pi \cdot q \mid U(q) = u\}$ or $C(p,u) = \underset{q}{Min}\{p \cdot q \mid U(q) = u\}$,

where the solution $q^c(\pi,u)$ or $q^c(p,u)$ is the vector of compensated direct demand functions.

$\cdot D(q,u) = \underset{\pi}{Min}\{\pi \cdot q \mid V(\pi) = u\}$,

where the solution $\pi^c(q,u)$ is the vector of compensated (normalized) inverse demand functions.

$\cdot C^R(p_A, q_B, u) = \underset{q_A}{Min}\{p_A \cdot q_A \mid U(q_A, q_B) = u\}$,

where solution $q_A^c(p_A, q_B, u)$ is the vector of compensated rationed demand, which is equal to the compensated quantity-dependent mixed demand. The negative of the compensated shadow or virtual prices is $-\left(\partial C^R / \partial q_k\right) = p_k^C(p_A, q_B, u)$, which are the compensated price-dependent mixed demand functions (Chavas, 1984). Thus mixed cost function can be defined as follows.

$$\cdot C^M(p_A, q_B, u) \equiv p_A \cdot q_A^c(p_A, q_B, u) + p_B^c(p_A, q_B, u) \cdot q_B = \sum_{i=1}^{m} p_i \cdot \left(\frac{\partial C^R}{\partial p_i}\right) + \sum_{k=m+1}^{N} \left(-\frac{\partial C^R}{\partial q_k}\right) \cdot q_k$$

which implies following derivative properties of mixed demand functions

$$\cdot \ \frac{\partial C^M}{\partial p_j} = \frac{\partial C^R}{\partial p_j} + \sum_{s=m+1}^{N} \frac{\partial p_s}{\partial p_j} q_s = q_j^c + \sum_{s=m+1}^{N} \frac{\partial p_s^c}{\partial p_j} q_s$$

$$\cdot \ \frac{\partial C^M}{\partial q_k} = \frac{\partial C^R}{\partial q_k} + \sum_{s=m+1}^{N} \frac{\partial p_s^c}{\partial q_k} q_s + p_k = -p_k + \sum_{s=m+1}^{N} \frac{\partial p_s^c}{\partial q_k} q_s + p_k = \sum_{s=m+1}^{N} \frac{\partial p_s^c}{\partial q_k} q_s$$

Note that no disequilibrium occurs in mixed demand, because the prices of commodities in fixed supply adjust at the shadow prices to clear market, while some markets do not clear in rationed demand (Moschini and Vissa, 1993).

It have been demonstrated, based on the envelope theorem, that the following Roy, Wold, Hotelling-Shephard, Shephard-Hanoch, and Samuelson theorems are useful to derive each of direct, inverse, and mixed demand functions respectively.

- $q_n(\pi,1)=\dfrac{\partial V/\partial\pi_n}{\sum_{n'=1}^{N}(\partial V/\partial\pi_{n'})\cdot\pi_{n'}}$ (Roy) · $\pi_n(q,1)=\dfrac{\partial U/\partial q_n}{\sum_{n'=1}^{N}(\partial U/\partial q_{n'})\cdot q_{n'}}$ (Wold)
- $q_n^c(p,u)=\dfrac{\partial C(p,u)}{\partial p_n}$ (Hotelling-Shephard) · $\pi_n^c(q,u)=\dfrac{\partial D(q,u)}{\partial q_n}$ (Shephard-Hanoch)
- $q_i^C(p_A,q_B,u)=\dfrac{\partial C^R}{\partial p_i}$ and $p_k^C(p_A,q_B,u)=-\dfrac{\partial C^R}{\partial q_k}$ (Samuelson's envelope theorems)

Note that the inverse demand functions can also be derived from the optimization problem for direct demand functions of $\underset{q}{Max}\left\{U(q)|\sum_{n=1}^{N}p_nq_n=y\right\}$ = $\underset{q,\lambda}{Max}\left\{U(q)+\lambda\cdot\left(y-\sum_{n=1}^{N}p_nq_n\right)\right\}$, which implies relations of $\dfrac{\partial U}{\partial q_n}=\lambda\cdot p_n$ as first order conditions. From this result and using the result $\lambda=\sum_{n'}^{N}\dfrac{\partial U}{\partial q_{n'}}\cdot q_{n'}\Big/y$ implied by $\sum_{n'}^{N}\dfrac{\partial U}{\partial q_{n'}}\cdot q_{n'}=\sum_{n'}^{N}\lambda\cdot p_{n'}\cdot q_{n'}=\lambda\cdot\sum_{n'}^{N}p_{n'}\cdot q_{n'}=\lambda\cdot y$, we can get following relation of the unnormalized and normalized inverse demand functions $p_n=\dfrac{\frac{\partial U}{\partial q_n}}{\lambda}=\dfrac{\frac{\partial U}{\partial q_n}}{\lambda}=\left(\dfrac{\frac{\partial U}{\partial q_n}}{\sum_{n'=1}^{N}\frac{\partial U}{\partial q_{n'}}q_{n'}}\right)\cdot y=p_n(q,y)=\pi_n(q)\cdot y$. The last equality holds by the Wold's theorem $\pi_n(q,1)=\dfrac{\partial U/\partial q_n}{\sum_{n'=1}^{N}(\partial U/\partial q_{n'})\cdot q_{n'}}$. We can see that the unnormalized inverse demand function is linear homogeneous in y, which implies that flexibility defined from either unnormalized or normalized inverse demand function has the same implication. This implication is explained as in $f_{n,n'}\equiv\dfrac{\partial p_n(q,y)}{\partial q_{n'}}\dfrac{q_{n'}}{p_n(q,y)}=\dfrac{\partial p_n\left(q,\frac{y}{y}\right)\frac{1}{y}}{\partial q_{n'}}\dfrac{q_{n'}}{p_n\left(q,\frac{y}{y}\right)\frac{1}{y}}=\dfrac{\partial p_n(q,1)}{\partial q_{n'}}\dfrac{q_{n'}}{p_n(q,1)}=\dfrac{\partial\pi_n(q)}{\partial q_{n'}}\dfrac{q_{n'}}{\pi_n(q)}$.

From an empirical perspective, consumer theory is considered as properties of the demand system of equations such as homogeneity, symmetry, negativity, adding-up, and relation of compensated and uncompensated demand functions (Barten, 1993). The first three properties for direct, inverse, and mixed demand functions can be derived from the properties of cost, distance, and restricted cost functions using Hotelling-Shephard, Shephard-Hanoch, and Samuelson's envelope theorems respectively. The Euler and Young theorem are used to derive properties of homogeneity and symmetry. While the Euler's theorem states that when $f(z)$ is r-th degree homogenous in z, then $g(z)=\partial f(z)/\partial z_{n'}$ is (r-1)-th degree homogenous in z and $\sum_{n'=1}^{N}\frac{\partial g(z)}{\partial z_{n'}}z_{n'}=(r-1)\cdot g(z)$, the Young's theorem states that when $f(z)$ is continuous function in z, then $\frac{\partial f(z)}{\partial z_n \partial z_{n'}}=\frac{\partial f(z)}{\partial z_{n'} \partial z_n}$, where $z=(z_1,\cdots,z_n,\cdots,z_N)$. The adding-up property of direct, inverse, and mixed demand functions can be derived from the budgetary identity equation or budgetary share equations. The main issue has been to derive relation of compensated and uncompensated demand functions. The Slutsky equation for direct demand is derived from the identity between compensated and uncompensated direct demands $q^c(p,u)\equiv q(p,y)\equiv q[p,C(p,u)]$. The Antonelli equation for inverse demand is derived from (normalized) inverse demand and direct utility function $\pi_n = f^n(q)= f^n(k\cdot q^*)= g^n(k,\, q^*)$ and $u=U(q)=U(k\cdot q^*)=U^*(k,\, q^*)$ in terms of scale parameters $q=k\cdot q^*$ where k is scalar and q^* is reference vector. The decomposition for mixed demand is derived from two identity equations between compensated and uncompensated mixed demands $q_i^c(p_A,q_B,u)\equiv q_i\ [p_A,q_B,C^M(p_A,q_B,u)]$ and $p_k^c(p_A,q_B,u)\equiv p_k[p_A,q_B,C^M(p_A,q_B,u)]$. The resulting theoretical implications can be summarized as follows, where $\varepsilon_{n,n'}\equiv\left(\frac{\partial q_n}{\partial p_{n'}}\frac{p_{n'}}{q_n}\right)$ and $\varepsilon_n\equiv\left(\frac{\partial q_n}{\partial y}\frac{y}{q_n}\right)$ denote price and expenditure elasticities from direct demand, $f_{n,n'}\equiv\left(\frac{\partial \pi_n}{\partial q_{n'}}\frac{q_{n'}}{\pi_n}\right)=\left(\frac{\partial p_n}{\partial q_{n'}}\frac{q_{n'}}{p_n}\right)$ and $f_n\equiv\left(\frac{\partial \pi_n}{\partial k}\frac{k}{\pi_n}\right)$ denote quantity and scale flexibilities from inverse demand, and $\varepsilon_{i,j}\equiv\left(\frac{\partial q_i}{\partial p_j}\frac{p_j}{q_i}\right)$, $f_{k,s}\equiv\left(\frac{\partial p_k}{\partial q_s}\frac{q_s}{p_k}\right)$, $q_{i,k}\equiv\left(\frac{\partial q_i}{\partial q_k}\frac{q_k}{q_i}\right)$, $p_{k,i}\equiv\left(\frac{\partial p_k}{\partial p_j}\frac{p_j}{p_k}\right)$, $\varepsilon_i\equiv\left(\frac{\partial q_i}{\partial y}\frac{y}{q_i}\right)$, and $f_k\equiv\left(\frac{\partial p_k}{\partial y}\frac{y}{p_k}\right)$ denote price and/or quantity and expenditure elasticities from mixed demand. The derivations of all theoretical properties or restrictions used for direct, inverse, and mixed demand system are explained in Appendix A. Since it is useful to express theoretical properties as elasticity or flexibility forms as well as derivative properties, especially for the differential demand systems, theoretical properties are summarized in both derivative and elasticity (or flexibility) forms.

Theoretical implications for direct demand systems

(a) Homogeneity: $\sum_{n'=1}^{N} \frac{\partial q_n^c}{\partial p_{n'}} p_{n'} = 0$ or $\sum_{n'=1}^{N} \varepsilon_{n,n'}^c = 0$ or $\sum_{n'=1}^{N} w_n \varepsilon_{n,n'}^c = 0$.

(b) Symmetry: $\frac{\partial q_n^c}{\partial p_{n'}} = \frac{\partial q_{n'}^c}{\partial p_n}$ or $w_n \varepsilon_{n,n'}^c = w_{n'} \varepsilon_{n',n}^c$.

(c) Slutsky equation: $\frac{\partial q_n^c}{\partial p_{n'}} = \frac{\partial q_n}{\partial p_{n'}} + \frac{\partial q_n}{\partial y} q_{n'}$ or $\varepsilon_{n,n'}^c = \varepsilon_{n,n'} + \varepsilon_n w_{n'}$ or $\varepsilon_{n,n'} = \varepsilon_{n,n'}^c - \varepsilon_n w_{n'}$.

(d) Adding-up: $\sum_{n=1}^{N} \varepsilon_n w_n = 1$ or $\sum_{n'=1}^{N} \varepsilon_{n,n'} = -\varepsilon_n$ or $\sum_{n=1}^{N} w_n \varepsilon_{n,n'} = -w_{n'}$.

Theoretical implications for inverse demand systems

(a) Homogeneity: $\sum_{n'=1}^{N} \frac{\partial \pi_n^c}{\partial q_{n'}} q_{n'} = 0$ or $\sum_{n'=1}^{N} f_{n,n'}^c = 0$ or $\sum_{n'=1}^{N} w_n f_{n,n'}^c = 0$.

(b) Symmetry: $\frac{\partial \pi_n^c}{\partial q_{n'}} = \frac{\partial \pi_{n'}^c}{\partial q_n}$ or $w_n f_{n,n'}^c = w_{n'} f_{n',n}^c$.

(c) Antonelli equation: $\frac{d\pi_n^c}{dq_{n'}^*} = \frac{\partial \pi_n}{\partial q_{n'}} - \frac{\partial \pi_n}{\partial k} \pi_{n'}$ or $f_{n,n'}^c = f_{n,n'} - f_n \cdot w_{n'}$ or $f_{n,n'} = f_{n,n'}^c + f_n \cdot w_{n'}$.

(d) Adding-up: $\sum_{n=1}^{N} w_n f_n = -1$ or $\sum_{n'=1}^{N} f_{n,n'} = f_n$ or $\sum_{n=1}^{N} f_{n,n'} w_n = -w_{n'}$.

Theoretical implications for mixed demand systems

(a) Homogeneity: $\sum_{j=1}^{m} \frac{\partial q_i^c}{\partial p_j} p_j = 0$ or $\sum_{j=1}^{m} \varepsilon_{i,j}^c = 0$ or $\sum_{j=1}^{m} w_i \cdot \varepsilon_{i,j}^c = 0$ and

$\sum_{j=1}^{m} \frac{\partial p_k^c}{\partial p_j} p_j = p_k^c$ or $\sum_{j=1}^{m} p_{k,j}^c = 1$ or $\sum_{j=1}^{m} w_k \cdot p_{k,j}^c = w_k$.

(b) Symmetry: $\frac{\partial q_i^c}{\partial p_j} = \frac{\partial q_j^c}{\partial p_i}$ or $w_i \cdot \varepsilon_{i,j}^c = w_j \cdot \varepsilon_{j,i}^c$,

$\frac{\partial p_k^c}{\partial q_s} = \frac{\partial p_s^c}{\partial q_k}$ or $w_k \cdot f_{k,s}^c = w_s \cdot f_{s,k}^c$, and

$-\frac{\partial p_k^c}{\partial p_i} = \frac{\partial q_i^c}{\partial q_k}$ or $-w_k \cdot p_{k,i}^c = w_i \cdot q_{i,k}^c$.

(c) Slutsky equation:

$$\frac{\partial q_i^c}{\partial p_j} = \frac{\partial q_i}{\partial p_j} + \frac{\partial q_i}{\partial y}\left[q_j^c + \sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial p_j} q_r \right], \qquad \frac{\partial q_i^c}{\partial q_s} = \frac{\partial q_i}{\partial p_j} + \frac{\partial q_i}{\partial y}\left[\sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial q_s} q_r \right],$$

$$\frac{\partial p_k^c}{\partial p_j} = \frac{\partial p_k}{\partial p_j} + \frac{\partial p_k}{\partial y}\left[q_j^c + \sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial p_j} q_r \right], \text{ and } \frac{\partial p_k^c}{\partial q_s} = \frac{\partial p_k}{\partial q_s} + \frac{\partial p_k}{\partial y}\left[\sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial q_s} q_r \right] \text{ or}$$

$$\varepsilon_{i,j}^c = \varepsilon_{i,j} + \varepsilon_i \left\lfloor w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right\rfloor \qquad \text{or } \varepsilon_{i,j} = \varepsilon_{i,j}^c - \varepsilon_i \left\lfloor w_j + \sum_{k=m+1}^{N} w_k \cdot p_{k,j}^c \right\rfloor,$$

$$q_{i,s}^c = q_{i,s} + \varepsilon_i \left\lfloor \sum_{r=m+1}^{N} w_r \cdot f_{r,s}^c \right\rfloor \qquad \text{or } q_{i,s} = q_{i,s}^c - \varepsilon_i \left\lfloor \sum_{r=m+1}^{N} w_r \cdot f_{r,s}^c \right\rfloor,$$

$$p_{k,j}^c = p_{k,j} + f_k \left\lfloor w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right\rfloor \qquad \text{or } p_{k,j} = p_{k,j}^c - f_k \left\lfloor w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right\rfloor, \text{ and}$$

$$f_{k,s}^c = f_{k,s} + f_k \left\lfloor \sum_{r=m+1}^{N} w_r \cdot f_{r,s}^c \right\rfloor \qquad \text{or } f_{k,s} = f_{k,s}^c - f_k \left\lfloor \sum_{r=m+1}^{N} w_r \cdot f_{r,s}^c \right\rfloor.$$

(d) Adding-up: $\sum_{i=1}^{m} w_i \varepsilon_i + \sum_{k=m+1}^{N} w_k f_k = 1$, $\sum_{i=1}^{m} w_i \varepsilon_{i,j}^c = 0$ and $\sum_{i=1}^{m} w_i q_{i,s}^c = -w_s$,

or $\varepsilon_i = -\sum_{j=1}^{m} \varepsilon_{i,j}$ and $f_k = 1 - \sum_{j=1}^{m} p_{k,j}$.

The choice of direct or inverse demand function is not trivial in empirical modeling to measure consumers' responsiveness, since it has been demonstrated that the flexibility (or elasticity) matrix has not the simple matrix inversion relation with the elasticity (or flexibility) matrix estimated from the direct (or inverse) demand functions (for example, Schultz, 1938, Houck, 1966, and Huang, 1996). From an econometrical perspective, the reason why an inverse relationship between elasticities and flexibilities does not hold can be understood by a following simple illustration of single demand equations with only one independent variable. For simplification, let p_t and q_t denote logarithmic transformation of price and quantity variables, so that α and β are the price elasticity and quantity flexibility respectively as in (a) $q_t = \alpha p_t + u_t^q$ and $p_t = \beta q_t + u_t^p$ and assume that direct least squares estimates are used as in (b) $\hat{\alpha} = \frac{\sum p_t q_t}{(\sum p_t)^2} = \frac{Cov(p_t, q_t)}{Var(p_t)}$ and $\hat{\beta} = \frac{\sum p_t q_t}{(\sum q_t)^2} = \frac{Cov(p_t, q_t)}{Var(q_t)}$. We can derive two kinds of relationships between elasticity and flexibility as follows: (c) $\hat{\alpha} \cdot \hat{\beta} = R_{p,q}^2$, where $R_{p,q}^2$ is the squared sample correlation of p and q. This relationship is based on the following relationship between estimators $\hat{\alpha}$ and $\hat{\beta}$. $\hat{\alpha} = \frac{Cov(p_t, q_t)}{Var(p_t)} = \left[\frac{Cov(p_t, q_t)}{Var(q_t)} \right]^{-1} \left[\frac{Cov(p_t, q_t)}{\sqrt{Var(p_t) Var(q_t)}} \right]^2 = \hat{\beta}^{-1} \cdot R_{p,q}^2$

(Schultz, 1938). (d) $\hat{\alpha} \cdot \hat{\beta} \leq 1$, where equality hold if and only if $\lambda_p p + \lambda_q q = 0$ for real scalars λ_p and λ_q, which is due to Cauchy-Schwarz inequality of $(p' \cdot q)^2 \leq (p' \cdot p)(q' \cdot q)$ for real column vector p and q (Huang, 1996). We can see that even in this simplest setting, the inverse relationship between two direct least squares estimates of $\hat{\alpha}$ and $\hat{\beta}$ does not hold in general. It holds only extremely special cases, where the squared sample correlation ($R^2_{p,q}$) is 1 as in (c) or price is exactly proportional to quantity ($p_t = (-\lambda_q / \lambda_p) \cdot q_t = \lambda \cdot q_t$) as in (d).

Given the fact that the general relationship between elasticity and flexibility is not yet well established, it is also worthwhile to derive some functional relationships among direct, inverse, and mixed demand systems. The relationships between elasticity and flexibility can be derived based on the mixed demand framework by extending the argument of Moschini and Vissa (1993). While they use a set of identity equations relating direct function to mixed function, there is another set of identity equations relating inverse function to mixed function. Using both sets of identity, we can also derive some relationship between direct and inverse demand, based on the mixed demand framework. Following notation is introduced. $E^D_{A,A} \equiv \lfloor \varepsilon_{i,i} \rfloor$, $E^D_{B,B} \equiv \lfloor \varepsilon_{k,s} \rfloor$, $E^D_{A,B} \equiv \lfloor \varepsilon_{i,k} \rfloor$, $E^D_{B,A} \equiv \lfloor \varepsilon_{k,j} \rfloor$, $E^D_A \equiv [\varepsilon_i]$, and $E^D_B \equiv [\varepsilon_k]$ are submatrices from direct demand, $F^I_{A,A} \equiv \lfloor f_{i,i} \rfloor$, $F^I_{B,B} \equiv \lfloor f_{k,s} \rfloor$, $F^I_{A,B} \equiv \lfloor f_{i,k} \rfloor$, $F^I_{B,A} \equiv \lfloor f_{k,j} \rfloor$, $F^I_A \equiv [f_i]$, and $F^I_B \equiv [f_k]$ are submatrices from inverse demand, and $E^M_{A,A} \equiv \lfloor \varepsilon_{i,i} \rfloor$, $F^M_{B,B} \equiv \lfloor f_{k,s} \rfloor$, $Q^M_{A,B} \equiv \lfloor q_{i,k} \rfloor$, $P^M_{B,A} \equiv \lfloor p_{k,j} \rfloor$, $E^M_A \equiv [\varepsilon_i]$, and $F^M_B \equiv [f_k]$ are submatrices from mixed demand. As Moschini and Vissa (1993) demonstrated, the direct demand system is related to the mixed demand system through the identities $q^D_A[p_A, p^M_B(p_A, q_B, y), y] \equiv q^M_A(p_A, q_B, y)$ and $q^D_B[p_A, p^M_B(p_A, q_B, y), y] \equiv \overline{q^M_B}$. By applying a similar logic, the inverse demand system is related to the mixed demand system through the following identities $p^I_A[q^M_A(p_A, q_B, y), q_B, y] \equiv \overline{p_A}$ and $p^I_B[q^M_A(p_A, q_B, y), q_B, y] \equiv p^M_B(p_A, q_B, y)$ which are implied by $\pi^I_A[q^M_A(\pi_A, q_B, 1), q_B, 1] \equiv \overline{\pi_A}$ and $\pi^I_B[q^M_A(\pi_A, q_B, 1), q_B, 1] \equiv \pi^M_B(\pi_A, q_B, 1)$ through $\pi^I_A[q^M_A(\pi_A, q_B, 1), q_B, 1] \cdot y \equiv \overline{\pi_A} \cdot y$ and $\pi^I_B[q^M_A(\pi_A, q_B, 1), q_B, 1] \cdot y \equiv \pi^M_B(\pi_A, q_B, 1) \cdot y$. From the resulting two kinds of relationships, other implied relationships can also be derived among direct, inverse, and mixed demand systems. Note that these relationships are based on the partitioning quantity-dependent and price-dependent groups of commodities or the legitimate mixed demand system. Note also that the scale flexibility is defined as responsiveness of (normalized) inverse demand with respect to scale parameter not with respect to expenditure variable. Derivations of following relationships are explained in Appendix B. The resulting relationships among direct, inverse, and mixed demand functions are summarized as follows:

Theoretical relation of direct elasticities to mixed elasticities.

$$E_{AA}^{D} = E_{AA}^{M} - Q_{AB}^{M} \cdot \left(F_{BB}^{M}\right)^{-1} P_{BA}^{M} \qquad E_{BB}^{D} = \left(F_{BB}^{M}\right)^{-1}$$

$$E_{AB}^{D} = Q_{AB}^{M} \cdot \left(F_{BB}^{M}\right)^{-1} \qquad E_{BA}^{D} = -\left(F_{BB}^{M}\right)^{-1} P_{BA}^{M}$$

$$E_{A}^{D} = E_{A}^{M} - Q_{AB}^{M} \cdot \left(F_{BB}^{M}\right)^{-1} F_{B}^{M} \qquad E_{B}^{D} = -\left(F_{BB}^{M}\right)^{-1} F_{B}^{M}\,.$$

Theoretical relation of inverse flexibilities to mixed elasticities.

$$F_{AA}^{I} = \left(E_{AA}^{M}\right)^{-1} \qquad F_{BB}^{I} = F_{BB}^{M} - P_{BA}^{M}\left(E_{AA}^{M}\right)^{-1} Q_{AB}^{M}$$

$$F_{AB}^{I} = -\left(E_{AA}^{M}\right)^{-1} Q_{AB}^{M} \qquad F_{BA}^{I} = P_{BA}^{M}\left(E_{AA}^{M}\right)^{-1}$$

$$F_{A}^{I} = RowSum\left[\left(E_{AA}^{M}\right)^{-1} \vdots -\left(E_{AA}^{M}\right)^{-1} Q_{AB}^{M}\right] \qquad F_{B}^{I} = RowSum\left[P_{BA}^{M}\left(E_{AA}^{M}\right)^{-1} \vdots F_{BB}^{M} - P_{BA}^{M}\left(E_{AA}^{M}\right)^{-1} Q_{AB}^{M}\right].$$

Theoretical relation of mixed elasticities to direct elasticities

$$E_{AA}^{M} = E_{AA}^{D} - E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1} E_{BA}^{D} \qquad F_{BB}^{M} = \left(E_{BB}^{D}\right)^{-1},$$

$$Q_{AB}^{M} = E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1}, \qquad P_{BA}^{M} = -\left(E_{BB}^{D}\right)^{-1} E_{BA}^{D},$$

$$E_{A}^{M} = E_{A}^{D} - E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1} E_{B}^{D}, \qquad F_{B}^{M} = -\left(E_{BB}^{D}\right)^{-1} E_{B}^{D}\,.$$

Theoretical relations of mixed elasticities to inverse flexibilities

$$E_{AA}^{M} = \left(F_{AA}^{I}\right)^{-1} \qquad F_{BB}^{M} = F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}$$

$$P_{BA}^{M} = F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} \qquad Q_{AB}^{M} = -\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}$$

$$E_{A}^{M} = -RowSum\left[\left(F_{AA}^{I}\right)^{-1}\right] \qquad F_{B}^{M} = I - RowSum\left[F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}\right].$$

Theoretical relation of direct elasticities to inverse flexibilities

$$E_{AA}^{D} = \left(F_{AA}^{I}\right)^{-1} + \left(F_{AA}^{I}\right)^{-1} F_{AB}^{I} \cdot \left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}\right]^{-1} F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} \qquad E_{BB}^{D} = \left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}\right]^{-1}$$

$$E_{AB}^{D} = -\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I} \cdot \left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}\right]^{-1} \qquad E_{BA}^{D} = -\left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}\right]^{-1} F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$$

$$E_{A}^{D} = -RowSum\left[\left(F_{AA}^{I}\right)^{-1} + \left(F_{AA}^{I}\right)^{-1} F_{AB}^{I} \cdot \left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}\right]^{-1} F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} \vdots -\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I} \cdot \left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}\right]^{-1}\right]$$

$$E_{B}^{D} = -RowSum\left[-\left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}\right]^{-1} F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} \vdots \left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1} F_{AB}^{I}\right]^{-1}\right].$$

Theoretical relation of inverse flexibilities to direct elasticities

$$F_{AA}^{I} = \left[E_{AA}^{D} - E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1} E_{BA}^{D}\right]^{-1} \qquad F_{BB}^{I} = \left(E_{BB}^{D}\right)^{-1} + \left(E_{BB}^{D}\right)^{-1} E_{BA}^{D}\left[E_{AA}^{D} - E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1} E_{BA}^{D}\right]^{-1} E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1}$$

$$F_{AB}^{I} = -\left[E_{AA}^{D} - E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1} E_{BA}^{D}\right]^{-1} E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1} \qquad F_{BA}^{I} = -\left(E_{BB}^{D}\right)^{-1} E_{BA}^{D}\left[E_{AA}^{D} - E_{AB}^{D}\left(E_{BB}^{D}\right)^{-1} E_{BA}^{D}\right]^{-1}$$

$$F_A^I = RowSum\left[\left[E_{AA}^D - E_{AB}^D\left(E_{BB}^D\right)^{-1}E_{BA}^D\right]^{-1} \vdots -\left[E_{AA}^D - E_{AB}^D\left(E_{BB}^D\right)^{-1}E_{BA}^D\right]^{-1}E_{AB}^D\left(E_{BB}^D\right)^{-1}\right]$$

$$F_B^I = RowSum\left[-\left(E_{BB}^D\right)^{-1}E_{BA}^D\left[E_{AA}^D - E_{AB}^D\left(E_{BB}^D\right)^{-1}E_{BA}^D\right]^{-1} \vdots \left(E_{BB}^D\right)^{-1} + \left(E_{BB}^D\right)^{-1}E_{BA}^D\left[E_{AA}^D - E_{AB}^D\left(E_{BB}^D\right)^{-1}E_{BA}^D\right]^{-1}E_{AB}^D\left(E_{BB}^D\right)^{-1}\right].$$

Heretofore, the full modeling spectrums of monotone set of direct or inverse demand functions as well as mixed demand functions are explained and their relationships are derived based on the mixed demand system. Although the mixed demand system provides a plausible way to sidestep the estimation of both demand and supply functions in a full simultaneous equation framework, the choice among three specifications for demand system remains open issue. When the choice among them only relies on a subjective reasoning of product property or market characteristics of a specific commodity rather than empirical evidence, the coexistence of alternative specifications can even result in ambiguities. For example as Thurman (1986) mentioned, both direct (Wohlgenant and Hahn, 1982) and inverse (Shonkwiler and Taylor, 1984) demand functions are used for poultry market data.

Given that theory does not provide enough information for this choice and the full simultaneous equations approach has some ambiguities in choosing appropriate instrumental variables, the graphical causal models discussed in previous chapter provide an alternative approach for the choice of empirical modeling among direct, inverse, and mixed demand systems. The specification choice is closely related with the identification issue of the local causal structure between price and quantity for a specific commodity. When we choose either quantity-dependent or price-dependent specification, we implicitly assume a local causal structure, since the direct (or inverse) demand function is implied by the causal structure that price (or quantity) variable causes quantity (or price) variable. The empirically derived causal structures through the proposed methods of DAG can be used to decide dependent and explanatory variable for a specific commodity demand function within the demand system. Stockton, Capps, and Bessler (2005) use this approach for meat demand study and named this approach as a Causally-Identified Demand System (CIDS). The (probabilistic) stability condition of the graphical causal model, however, can be violated in using a high dimensional data as discussed in chapter II, given the observation that many variables in retail scanner data move very closely. The compositional stability condition is proposed to address this issue in using the graphical causal model, since the compositional stability condition makes it possible to capture disaggregated micro-relations by the aggregated macro-relations as the legitimate representatives.

Aggregation in Study of Consumer Behavior

The legitimate condition of classification and the appropriate way of aggregation, which are related with the (probabilistic) stability condition of the graphical causal model, have also been major issues in the context of the more general econometric considerations in empirical studies especially in using a high dimensional data set. The availability of scanner data makes it possible to define finer variables based on thousands of individual products at the store level on daily frequencies. However, econometric considerations such as the degrees-of-freedom and multicollinearity require classification and aggregation procedures for economy of parameters in empirical study. While classification and aggregation issues are involved with multi-dimensions such as commodity-wise, agent-wise or spatial, and temporal or time dimensions, the main focus in empirical studies has been on the commodity-wise dimension. Even though the level of classification and aggregation and the choice of a specific category have been often based on convenience for addressing specific research objectives rather than on the empirical evidence (Shumway and Davis, 2001 and reference in there), it has been argued that small departures from valid classification and/or aggregation can result in large mistakes in elasticity/flexibility and welfare estimates (Lewbel, 1996). For example, the decision on classification and aggregation can substantially affect the conclusions about elasticity estimates in multi-stage budgeting approach, because cross-price elasticity or cross-quantity flexibility between products in different groups is likely to be small by construction itself (Rubinfeld, 2000).

The classification and aggregation issues have been addressed by using homothetic or weak separability condition or generalized composite commodity condition in the context of quantity-dependent specification of demand function. However, there are some difficulties or ambiguities in using their conditions in empirical studies as discussed in chapter II. We propose to use the generalized form of the compositional stability condition derived from the Theil's aggregation theory to address classification and aggregation issue in more general context of all possible direct, inverse, and mixed demand functions. The Tornqvist-Theil index, based on the discussion of the index number theory in chapter II, is mainly used for the actual aggregation or the decision of weighting schemes for aggregating disaggregated micro-variables within each of the identified homogenous groups into representative macro-variables. The compositional stability condition of $Cov\left(d_n, X\right)=0$ in $x_n = X H_n + d_n$ are empirically tested by using a Hausman type misspecification test of $H_0 : \gamma_n = 0$ in $x_n = X H_n + IV \cdot \gamma_n + \varepsilon_n^{IV}$, where x_n are disaggregated micro-variables of either price or quantity of a specific group and X are corresponding aggregated macro-variables of either price or quantity of a specific group. The IV are Instrumental Variable such that IV is closely correlated with regressor X (relevance condition of IV) and independent of error d_n (validity condition of IV). In this study, we use the total expenditure variable, which is calculated by aggregating the price and quantity macro-variables within the demand system, as the instrumental variable based on the following reasoning. Given that the total expenditure is closely related with the aggregated price and quantity variables as in estimated aggregated demand systems, the relevance condition can holds. The validity condition of the total expenditure variable as an instrumental variable can also hold. Such possibility exists, since either each of the idiosyncratic variations of disaggregated price or quantity variable can cancel each other in calculating the total expenditure variable or the idiosyncratic variation of individual price or quantity variable, which is not captured by the common variation of representative macro-variables of a specific group, does not have dependencies on the total expenditure variable, which captures the common variation of an entire group of commodities within the demand system through group-representative price and quantity macro-variables.

The problem of forming suitable partitions before conducting any empirical test to justify those classifications has relied on researchers' intuition rather than empirical data patterns. The intuitive partitions based on the subjective reasoning are only a small set of possible partitions among an extremely large number of possible partitions. Thus when classification is empirically rejected, it might be simply because of researchers' unsuccessful identification of the partition itself, not because of non-existence of legitimate classification itself. Given the empirical implausibility of attempting all possible partitions, it can be helpful to pursue inductive partitions related with legitimate aggregation conditions based on the data pattern itself. The approximate form of the compositional stability condition is used to search for a specific homogeneous group. The homogeneous grouping or partitioning of related commodities is identified by the block-diagonal pattern of static and dynamic correlation matrix of price and quantity variables, based on the modified k-nearest neighbor algorithm.

The compositional stability condition as the consistent aggregation condition is closely related with the (probabilistic) stability condition as the fundamental condition for the graphical causal models. When generalized forms of the compositional stability condition can be identified in data set through grouping micro-variables based on their correlation or covariance matrix, there exist not only the possibility of obtaining interpretable aggregate indexes or macro-variables as the representative aggregate of homogeneous disaggregate micro-variables, but also the possibility of obtaining interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for subsequent analysis. This implies that when the micro-variables can be legitimately grouped and represented by macro-variables, it is possible to use aggregation methods to capture (causal) relationships among disaggregated variables through (causal) relationships among aggregated variables as the legitimate representatives.

Functional Form in Study of Consumer Behavior

While it is possible to define aggregated variables based on the consistent aggregation condition and to choose among direct, inverse, and mixed demand systems based on the graphical causal models, there remains another issue of deciding functional form to relate the dependent variable with explanatory variables in an empirical model. This issue has been a frequently discussed topic in empirical demand literatures. Many useful functional forms have been proposed and used for the direct and inverse demand functions. Several functional forms of direct demand system have been converted for use in inverse demand systems and vice versa, based on the polar relations between both specifications. However, when we want to compare direct, inverse, and mixed demand systems in the similar functional form specifications, the possible use of mixed demand system impose some limitations for considering possible range of functional forms. It is because the mixed demand system requires consistent and simultaneous specifications for both direct and indirect utility functions and the commonly used flexible functional forms, such as those underlying the translog and almost ideal systems, do not have a closed form dual representation for both direct and indirect utility functions. As Moschini and Vissa (1993) emphasize, an appropriate approach for a flexible demand system of mixed demand functions is to approximate each demand function directly by a differential Rotterdam demand system and to impose the theoretical restrictions.

The Rotterdam demand system has been a commonly used functional form for both direct and inverse demand systems, since it is regarded as flexible in that it provides a first-order approximation to an arbitrary demand system in either parameter or variable space. Another commonly used functional form is the Almost Ideal Demand Systems (AIDS) or the Linear Approximate Almost Ideal Demand Systems (LA/AIDS). While these two demand systems are common in demand system estimation in agricultural economics, especially for using scanner data, the assumptions used to parameterize these two systems have different implications. While the Rotterdam parameterization assumes that both the income (or scale) coefficient and the compensated price (or quantity) coefficient in the direct (or inverse) demand system are constant parameters, the LA/AIDS parameterization assumes that both the income (or scale) coefficient and the Slutsky (or Antonelli) coefficient in the direct (or inverse) demand system are variational parameters dependent on the budget shares. Two more logically possible combinations of constant/variational parameterization for the income (or scale) coefficient and the Slutsky (or Antonelli) coefficient are also used for both direct and inverse systems. While Keller and van Driel (1985) of Dutch Central Bureau of Statistics (CBS) introduce variational income (or scale) coefficient with constant Slutsky (or Antonelli) coefficient by reparameterizing the Rotterdam specification, Neves (1987) of Netherlands National Bureau of Research (NBR) introduce income (or scale) coefficient with variational Slutsky (or Antonelli) coefficient by reparameterizing the differential AIDS specification. Given that economic theory does not provide sufficient information for this issue, the use of intuitive reasoning rather than empirical evidence can result in coexistence of alternative specifications and thus generate ambiguities, since elasticities (or flexibilities) are sensitive to the functional form specification. Even though this general finding that elasticities (or flexibilities) are sensitive to the functional form specification makes this issue of functional form specification non trivial, empirical comparisons among alternative specification have been rarely done. The main difficulties are the alternative specifications are non-nested relative to each other and the non-nested hypotheses testing approach oftentimes does not provide a conclusive answer for this problem in general situations. An alternative method for this problem is using the principle of artificial nesting. In this respective, it has been demonstrated that the Rotterdam, the differential AIDS, and two hybrid demand specifications of CBS and NBR can be nested within a synthetic direct (Barten, 1993) and inverse (Brown, Lee, and Seal, 1995) demand system. It has been argued that these two synthetic direct and inverse demand systems can be considered as demand systems in their own right, beyond an artificial composite of known models. For example, Matsuda (2005) shows that one of the nesting coefficients in the inverse synthetic model of Brown, Lee, and Seal (1995) implies the transformation parameter of the Box-Cox scale curves. Using a similar idea based on the Box-Cox scale curves, Matsuda (2004) proposes a mixed demand specification, nesting Rotterdam and CBS specifications.

When we want to compare direct, inverse, and mixed demand systems, we need parameterize three demand systems in the similar degrees of flexibility in functional form specifications, when the flexibility means the capability of the empirical model to allow the possible combinations of constant/variational parameterization for the income (or scale) coefficient and the Slutsky (or Antonelli) coefficient. Given that the synthetic differential demand model exists for the direct and inverse demand system, the synthetic differential demand model is proposed for the mixed demand system based on the similar logic to derive synthetic demand model in direct and inverse demand systems. Furthermore Eales, Durham, and Wessells (1997) show that synthetic direct and inverse demand systems can be reparameterized to have common differential AIDS dependent variables, which makes it possible to compare direct and inverse demand functions. By extending the common logic of these approaches, a similar synthetic functional form for a mixed demand system can be specified in the common differential AIDS dependent variables, which makes it possible to compare direct, inverse, and mixed demand systems in the model selection frameworks. Rotterdam type and AIDS type dependent variable synthetic models can be directly derived from Rotterdam specification as explained below, which make it possible to derive synthetic mixed demand functions. Derivations of direct, inverse, mixed demand functions are explained in Appendix C. The original functional form and the Rotterdam type and AIDS type dependent variable synthetic model specifications can be summarized as follows.

The differential family of four direct demand systems can be summarized and nested in either Rotterdam or AIDS dependent variable forms of synthetic direct demand systems. If the expenditure coefficient is defined as $a_n \equiv [w_n \varepsilon_n]$ or $c_n \equiv [w_n \varepsilon_n - w_n]$ and the Slutsky coefficient is defined as $a_{n,n'} \equiv \lfloor w_n \varepsilon^c_{n,n'} \rfloor$ or $c_{n,n'} \equiv \lfloor w_n \varepsilon^c_{n,n'} - w_n (w_{n'} - \delta_{n,n'}) \rfloor$, then both are nested by synthetic parameters of $C_n \equiv [w_n \varepsilon_n - \theta^o_1 w_n]$ and $C_{n,n'} \equiv \lfloor w_n \varepsilon^c_{n,n'} - \theta^o_2 w_n (w_{n'} - \delta_{n,n'}) \rfloor$ respectively.

Rotterdam : $$w_n d\ln q_n = [w_n \varepsilon_n] d\ln Q + \sum_{n'=1}^{N} [w_n \varepsilon^c_{n,n'}] d\ln p_{n'} \text{ or}$$

$$w_n d\ln q_n = [a_n] d\ln Q + \sum_{n'=1}^{N} [a_{n,n'}] d\ln p_{n'} \text{ or}$$

$$dw_n = [a_n - w_n] d\ln Q + \sum_{n'=1}^{N} [a_{n,n'} - w_n (w_{n'} - \delta_{n,n'})] d\ln p_{n'}\ .$$

Differential AIDS: $$dw_n = [w_n \varepsilon_n - w_n] d\ln Q + \sum_{n'=1}^{N} [w_n \varepsilon^c_{n,n'} - w_n (w_{n'} - \delta_{n,n'})] d\ln p_{n'} \text{ or}$$

$$w_n d\ln q_n = [c_n + w_n] d\ln Q + \sum_{n'=1}^{N} [c_{n,n'} + w_n (w_{n'} - \delta_{n,n'})] d\ln p_{n'} \text{ or}$$

$$dw_n = [c_n] d\ln Q + \sum_{n'=1}^{N} [c_{n,n'}] d\ln p_{n'}\ .$$

CBS: $$w_n d\ln\left(\frac{q_n}{Q}\right) = [c_n] d\ln Q + \sum_{n'=1}^{N} a_{n,n'} \ln p_{n'} \text{ or}$$

$$w_n d\ln q_n = [c_n + w_n] d\ln Q + \sum_{n'=1}^{N} [a_{n,n'}] d\ln p_{n'} \text{ or}$$

$$dw_n = [c_n] d\ln Q + \sum_{n'=1}^{N} [a_{n,n'} - w_n(w_{n'} - \delta_{n,n'})] d\ln p_{n'} .$$

NBR: $$(dw_n + w_n d\ln Q) = [a_n] d\ln Q + \sum_{n'=1}^{N} [c_{n,n'}] d\ln p_{n'} \text{ or}$$

$$w_n d\ln q_n = [a_n] d\ln Q + \sum_{n'=1}^{N} [c_{n,n'} + w_n(w_{n'} - \delta_{n,n'})] d\ln p_{n'} \text{ or}$$

$$dw_n = [a_n - w_n] d\ln Q + \sum_{n'=1}^{N} [c_{n,n'}] d\ln p_{n'} .$$

Synthetic: $$w_n d\ln q_n = [C_n + \theta_1^O w_n] d\ln Q + \sum_{n'=1}^{N} [C_{n,n'} + \theta_2^O w_n(w_{n'} - \delta_{n,n'})] d\ln p_{n'} \text{ or}$$

$$dw_n = [C_n - (1-\theta_1^O) w_n] d\ln Q + \sum_{n'=1}^{N} [C_{n,n'} - (1-\theta_2^O) w_n(w_{n'} - \delta_{n,n'})] d\ln p_{n'} .$$

The Rotterdam type dependent variable synthetic forms can be derived as follows.

$$w_n d\ln q_n = [w_n\varepsilon_n - \theta_1^O w_n + \theta_1^O w_n] d\ln Q + \sum_{n'=1}^{N} [w_n\varepsilon_{n,n'}^c - \theta_2^O w_n(w_{n'} - \delta_{n,n'}) + \theta_2^O w_n(w_{n'} - \delta_{n,n'})] d\ln p_{n'} ,$$

which can be transformed into AIDS type dependent variable synthetic forms as follows.

$$\begin{aligned} dw_n &\equiv \langle w_n d\ln q_n \rangle + w_n d\ln p_n - w_n [d\ln Q + d\ln P] \\ &= \langle w_n d\ln q_n \rangle + w_n \left[\sum_{n'=1}^{N} \delta_{n,n'} d\ln p_{n'}\right] - w_n d\ln Q - w_n \left[\sum_{n'=1}^{N} w_{n'} d\ln p_{n'}\right] \\ &= \langle w_n d\ln q_n \rangle - w_n d\ln Q - w_n \left[\sum_{n'=1}^{N} (w_{n'} - \delta_{n,n'}) d\ln p_{n'}\right] \end{aligned}$$

$$dw_n = [w_n\varepsilon_n - \theta_1^O w_n - (1-\theta_1^O) w_n] d\ln Q + \sum_{n'=1}^{N} [w_n\varepsilon_{n,n'}^c - \theta_2^O w_n(w_{n'} - \delta_{n,n'}) - (1-\theta_2^O) w_n(w_{n'} - \delta_{n,n'})] d\ln p_{n'} .$$

Theoretical restrictions can be imposed by using following relations

(a) Homogeneity: $\sum_{n'=1}^{N} C_{n,n'} = 0$,

(b) Symmetry: $C_{n,n'} = C_{n',n}$,

(c) Adding-up: $\sum_{n=1}^{N} C_n = 1 - \theta_1^O$.

Because: (a) $\sum_{n'=1}^{N} C_{n,n'} \equiv \sum_{n'=1}^{N} [w_n\varepsilon_{n,n'}^c - \theta_2^O w_n(w_{n'} - \delta_{n,n'})] = \sum_{n'=1}^{N} w_n\varepsilon_{n,n'}^c - \theta_2^O w_n \sum_{n'=1}^{N} [w_{n'} - \delta_{n,n'}]$, which, by $\sum_{n'=1}^{N} w_n\varepsilon_{n,n'}^c = 0$, is $\sum_{n'=1}^{N} C_{n,n'} = -\theta_2^O w_n \left[\sum_{n'=1}^{N} w_{n'} - \sum_{n'=1}^{N} \delta_{n,n'}\right] = -\theta_2^O w_n [1-1] = 0$. (b) Using $w_n\varepsilon_{n,n'}^c = w_{n'}\varepsilon_{n',n}^c$, we can compare $C_{n,n'} = w_n\varepsilon_{n,n'}^c - \theta_2^O w_n(w_{n'} - \delta_{n,n'})$ with $C_{n',n} = w_{n'}\varepsilon_{n',n}^c - \theta_2^O w_{n'}(w_n - \delta_{n',n})$ as $w_n w_{n'} - w_n\delta_{n,n'}$

with $w_{n'}w_n - w_{n'}\delta_{n',n}$, which is equal because $w_n\delta_{n,n'} = w_{n'}\delta_{n',n}$. (c) Using $\sum_{n=1}^{N} w_n\varepsilon_n = 1$ and $\sum_{n=1}^{N} w_n = 1$, $\sum_{n=1}^{N} C_n \equiv \sum_{n=1}^{N}\left[w_n\varepsilon_n - \theta_1^O w_n\right] = \left(\sum_{n=1}^{N} w_n\varepsilon_n\right) - \theta_1^O\left(\sum_{n=1}^{N} w_n\right)$ can be written as $\sum_{n=1}^{N} C_n = 1 - \theta_1^O$.

The elasticities are calculated as follows

(a) Expenditure elasticity: $\varepsilon_n = \dfrac{C_n}{w_n} + \theta_1^O$,

(b) Compensated elasticity: $\varepsilon_{n,n'}^{c} = \dfrac{C_{n,n'}}{w_n} + \theta_2^O\left(w_{n'} - \delta_{n,n'}\right)$, and

(c) Uncompensated elasticity: $\varepsilon_{n,n'} = \left[\dfrac{C_{n,n'}}{w_n} + \theta_2^O\left(w_{n'} - \delta_{n,n'}\right)\right] - \left[C_n\left(\dfrac{w_{n'}}{w_n}\right) + \theta_1^O w_{n'}\right]$.

Because: (a) $C_n \equiv w_n\varepsilon_n - \theta_1^O w_n$, (b) $C_{n,n'} \equiv w_n\varepsilon_{n,n'}^{c} - \theta_2^O w_n\left(w_{n'} - \delta_{n,n'}\right)$, and (c) $\varepsilon_{n,n'} = \varepsilon_{n,n'}^{C} - \varepsilon_n w_{n'}$`

The differential family of four inverse demand systems can be summarized and nested in either Rotterdam or AIDS dependent variable forms of synthetic inverse demand systems. If the scale coefficient is defined as $b_n \equiv \left[w_n f_n\right]$ or $d_n \equiv \left[w_n f_n + w_n\right]$ and the Antonelli coefficient is defined as $b_{n,n'} \equiv \left[w_n f_{n,n'}^{c}\right]$ or $d_{n,n'} \equiv \left[w_n f_{n,n'}^{c} - w_n\left(w_{n'} - \delta_{n,n'}\right)\right]$, then both of them are nested by synthetic parameters of $D_n \equiv \left[w_n f_n + \theta_1^I w_n\right]$ and $D_{n,n'} \equiv \left[w_n f_{n,n'}^{c} - \theta_2^I w_n\left(w_{n'} - \delta_{n,n'}\right)\right]$ respectively.

Rotterdam:

$$w_n d\ln\pi_n = \left[w_n f_n\right] d\ln Q + \sum_{n'=1}^{N}\left[w_n f_{n,n'}^{c}\right] d\ln q_{n'} \text{ or}$$

$$w_n d\ln\pi_n = \left[b_n\right] d\ln Q + \sum_{n'=1}^{N}\left[b_{n,n'}\right] d\ln q_{n'} \text{ or}$$

$$dw_n = \left[b_n + w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[b_{n,n'} - w_n\left(w_{n'} - \delta_{n,n'}\right)\right] d\ln q_{n'}.$$

Differential AIDS:

$$dw_n = \left[w_n f_n + w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[w_n f_{n,n'}^{c} - w_n\left(w_{n'} - \delta_{n,n'}\right)\right] d\ln q_{n'} \text{ or}$$

$$w_n d\ln\pi_n = \left[d_n - w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[d_{n,n'} + w_n\left(w_{n'} - \delta_{n,n'}\right)\right] d\ln q_{n'} \text{ or}$$

$$dw_n = \left[d_n\right] d\ln Q + \sum_{n'=1}^{N}\left[d_{n,n'}\right] d\ln q_{n'}.$$

CBS:

$$w_n d\ln\left(\frac{p_n}{P}\right) = \left[d_n\right] d\ln Q + \sum_{n'=1}^{N}\left[b_{n,n'}\right] \ln q_{n'} \text{ or}$$

$$w_n d\ln\pi_n = \left[d_n - w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[b_{n,n'}\right] d\ln q_{n'} \text{ or}$$

$$dw_n = \left[d_n\right] d\ln Q + \sum_{n'=1}^{N}\left[d_{n,n'}\right] d\ln q_{n'}.$$

NBR: $\left(dw_n + w_n d\ln Q\right) = \left[b_n\right] d\ln Q + \sum_{n'=1}^{N}\left[d_{n,n'}\right] d\ln p_{n'}$ or

$$w_n d\ln\pi_n = \left[b_n\right] d\ln Q + \sum_{n'=1}^{N}\left[d_{n,n'} + w_n\left(w_{n'} - \delta_{n,n'}\right)\right] d\ln q_{n'} \text{ or}$$

$$dw_n = \left[b_n + w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[c_{n,n'}\right] d\ln p_{n'} .$$

Synthetic: $w_n d\ln\pi_n = \left[D_n - \theta_1^I w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[D_{n,n'} + \theta_2^I w_n\left(w_{n'} - \delta_{n,n'}\right)\right] d\ln q_{n'}$ or

$$dw_n = \left[D_n + \left(1-\theta_1^I\right) w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[D_{n,n'} - \left(1-\theta_2^I\right) w_n\left(w_{n'} - \delta_{n,n'}\right)\right] d\ln q_{n'} .$$

The Rotterdam type dependent variables synthetic forms can be derived as follows.

$$w_n d\ln\pi_n = \left[w_n f_n + \theta_1^I w_n - \theta_1^I w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[w_n f_{n,n'}^c - \theta_2^I w_n\left(w_{n'} - \delta_{n,n'}\right) + \theta_2^I w_n\left(w_{n'} - \delta_{n,n'}\right)\right] d\ln q_{n'} ,$$

which can be transformed into AIDS type dependent variables synthetic forms as follows.

$$\begin{aligned}
dw_n &\equiv w_n d\ln p_n + w_n d\ln q_n - w_n d\ln y \\
&= w_n\left[d\ln p_n - d\ln y\right] + w_n d\ln q_n \\
&= \left\langle w_n d\ln\pi_n\right\rangle + w_n d\ln q_n + \left(w_n d\ln Q - w_n d\ln Q\right) \\
&= \left\langle w_n d\ln\pi_n\right\rangle + w_n\left[\sum_{n'=1}^{N}\delta_{n,n'} d\ln q_{n'}\right] + w_n d\ln Q - w_n\left[\sum_{n'=1}^{N} w_{n'} d\ln q_{n'}\right] \\
&= \left\langle w_n d\ln\pi_n\right\rangle + w_n d\ln Q - w_n\left[\sum_{n'=1}^{N}\left(w_{n'} - \delta_{n,n'}\right) d\ln q_{n'}\right]
\end{aligned}$$

$$dw_n = \left[w_n f_n + \theta_1^I w_n + \left(1-\theta_1^I\right) w_n\right] d\ln Q + \sum_{n'=1}^{N}\left[w_n f_{n,n'}^c - \theta_2^I w_n\left(w_{n'} - \delta_{n,n'}\right) - \left(1-\theta_2^I\right) w_n\left(w_{n'} - \delta_{n,n'}\right)\right] d\ln q_{n'}$$

Theoretical restrictions can be imposed by using following relations

(a) Homogeneity: $\sum_{n'=1}^{N} D_{n,n'} = 0$,

(b) Symmetry: $D_{n,n'} = D_{n',n}$,

(c) Adding-up: $\sum_{n=1}^{N} D_n = -1 + \theta_1^I$.

Because: (a) $\sum_{n'=1}^{N} D_{n,n'} \equiv \sum_{n'=1}^{N}\left[w_n f_{n,n'}^c - \theta_2^I w_n\left(w_{n'} - \delta_{n,n'}\right)\right] = \sum_{n'=1}^{N} w_n f_{n,n'}^c - \theta_2^I w_n \sum_{n'=1}^{N}\left[w_{n'} - \delta_{n,n'}\right]$, which, by $\sum_{n'=1}^{N} w_n f_{n,n'}^c = 0$, is $\sum_{n'=1}^{N} D_{n,n'} = -\theta_2^I w_n\left[\sum_{n'=1}^{N} w_{n'} - \sum_{n'=1}^{N}\delta_{n,n'}\right] = -\theta_2^I w_n\left[1-1\right] = 0$. (b) Using $w_n f_{n,n'}^c = w_{n'} f_{n',n}^c$, we can compare $D_{n,n'} \equiv w_n f_{n,n'}^c - \theta_2^I w_n\left(w_{n'} - \delta_{n,n'}\right)$ with $D_{n',n} \equiv w_{n'} f_{n',n}^c - \theta_2^I w_{n'}\left(w_n - \delta_{n',n}\right)$ as $w_n w_{n'} - w_n \delta_{n,n'}$ with $w_{n'} w_n - w_{n'}\delta_{n',n}$, which is equal because $w_n \delta_{n,n'} = w_{n'}\delta_{n',n}$. (c) Using by $\sum_{n=1}^{N} w_n f_n = -1$ and $\sum_{n=1}^{N} w_n = 1$`, $\sum_{n=1}^{N} D_n \equiv \sum_{n=1}^{N}\left[w_n f_n + \theta_1^I w_n\right] = \left(\sum_{n=1}^{N} w_n f_n\right) + \theta_1^I\left(\sum_{n=1}^{N} w_n\right)$ can be written as $\sum_{n=1}^{N} D_n = -1 + \theta_1^I$

The elasticities are calculated as follows

(a) Scale flexibility: $f_n = \dfrac{D_n}{w_n} - \theta_1^I$,

(b) Compensated flexibility: $f_{n,n'}^c = \dfrac{D_{n,n'}}{w_n} + \theta_2^I\left(w_{n'} - \delta_{n,n'}\right)$, and

(c) Uncompensated flexibility: $f_{n,n'} = \left[\dfrac{D_{n,n'}}{w_n} + \theta_2^I\left(w_{n'} - \delta_{n,n'}\right)\right] + \left[D_n\left(\dfrac{w_{n'}}{w_n}\right) - \theta_1^I w_{n'}\right]$.

because (a) $D_n = w_n f_n + \theta_1^I w_n$, (b) $D_{n,n'} = w_n f_{n,n'}^c - \theta_2^I w_n\left(w_{n'} - \delta_{n,n'}\right)$, and (c) $f_{n,n'} = f_{n,n'}^C + f_n w_{n'}$

The differential family of mixed demand systems can be derived and nested in either Rotterdam or AIDS dependent variable forms of analogous synthetic mixed demand systems. The expenditure coefficients of group A and B are defined as $\alpha_i \equiv \left[w_i\varepsilon_i - \theta_1^M w_i\right]$ and $\beta_k \equiv \left[w_k f_k - \theta_1^M w_k\right]$ and the Slutsky coefficients are defined as $\alpha_{i,j} \equiv \left[w_i\varepsilon_{i,j}^c - \theta_2^M w_i\left(w_j - \delta_{i,j}\right)\right]$, $\beta_{k,s} \equiv \left[w_k f_{k,s}^c - \theta_2^M w_k\left(w_s - \delta_{k,s}\right)\right]$, $-\gamma_{k,j} \equiv \left[w_k p_{k,j}^c + \theta_2^M w_k w_j\right]$, and $g_{i,s} \equiv \left[w_i q_{i,s}^c - \theta_2^M w_i w_s\right]$.

Rotterdam:

$$\begin{aligned} w_i d\ln q_i = {} & \left[w_i\varepsilon_i\right]\cdot d\ln\bar{y} \\ & + \sum_{j=1}^{m}\left[w_i\varepsilon_{i,j}^c - w_i\varepsilon_i\cdot\left(\sum_{k=m+1}^{N} w_k\cdot p_{k,j}^c\right)\right]\cdot d\ln p_j \\ & + \sum_{s=m+1}^{N}\left[w_i q_{i,s}^c - w_i\varepsilon_i\cdot\left(\sum_{r=m+1}^{N} w_r f_{r,s}^c\right)\right]\cdot d\ln q_s \end{aligned}$$

$$\begin{aligned} w_k d\ln p_k = {} & \left[w_k f_k\right]\cdot d\ln\bar{y} \\ & + \sum_{j=1}^{m}\left[w_k p_{k,j}^c - w_k f_k\cdot\left(\sum_{r=m+1}^{N} w_r p_{r,j}^c\right)\right]\cdot d\ln p_j \\ & + \sum_{s=m+1}^{N}\left[w_k f_{k,s}^c - w_k f_k\cdot\left(\sum_{r=m+1}^{N} w_r f_{r,s}^c\right)\right]\cdot d\ln q_s \end{aligned}$$

Synthetic:

$$\begin{aligned} w_i d\ln q_i = {} & \left[\alpha_i + \theta_1^M w_i\right]\cdot d\ln\bar{y} \\ & + \sum_{j=1}^{m}\left[\alpha_{i,j} + \theta_2^M w_i\left(w_j - \delta_{i,j}\right) - \left(\alpha_i + \theta_1^M w_i\right)\cdot\left(\sum_{r=m+1}^{N} -\gamma_{r,j} - \theta_2^M w_r w_j\right)\right]\cdot d\ln p_j \\ & + \sum_{s=m+1}^{N}\left[g_{i,s} + \theta_2^M w_i w_s - \left(\alpha_i + \theta_1^M w_i\right)\cdot\left(\sum_{r=m+1}^{N}\beta_{r,s} + \theta_2^M w_r\left(w_s - \delta_{r,s}\right)\right)\right]\cdot d\ln q_s \end{aligned}$$

$$\begin{aligned} w_k d\ln p_k = {} & \left[\beta_k + \theta_1^M w_k\right]\cdot d\ln\bar{y} \\ & + \sum_{j=1}^{m}\left[-\gamma_{k,j} - \theta_2^M w_k w_j - \left(\beta_k + \theta_1^M w_k\right)\cdot\left(\sum_{r=m+1}^{N} -\gamma_{r,j} - \theta_2^M w_r w_j\right)\right]\cdot d\ln p_j \\ & + \sum_{s=m+1}^{N}\left[\beta_{k,s} + \theta_2^M w_k\left(w_s - \delta_{k,s}\right) - \left(\beta_k + \theta_1^M w_k\right)\cdot\left(\sum_{r=m+1}^{N}\beta_{r,s} + \theta_2^M w_r\left(w_s - \delta_{r,s}\right)\right)\right]\cdot d\ln q_s \end{aligned}$$

or

$$
\begin{aligned}
dw_i &= \left[\alpha_i + \left(\theta_1^M - 1\right) w_i \right] \cdot d \ln \bar{y} \\
&+ \sum_{j=1}^{m} \left[\alpha_{i,j} + \left(\theta_2^M - 1\right) w_i \left(w_j - \delta_{i,j}\right) - \left(\alpha_i + \theta_1^M w_i \right) \cdot \left(\sum_{r=m+1}^{N} - \gamma_{r,j} - \theta_2^M w_r w_j \right) \right] \cdot d \ln p_j \\
&+ \sum_{s=m+1}^{N} \left[g_{i,s} + \theta_2^M w_i w_s - \left(\alpha_i + \theta_1^M w_i \right) \cdot \left(\sum_{r=m+1}^{N} \beta_{r,s} + \theta_2^M w_r \left(w_s - \delta_{r,s}\right) \right) \right] \cdot d \ln q_s
\end{aligned}
$$

$$
\begin{aligned}
dw_k &= \left[\beta_k + \left(\theta_1^M - 1\right) w_k \right] \cdot d \ln \bar{y} \\
&+ \sum_{j=1}^{m} \left[- \gamma_{k,j} - \left(\theta_2^M + 1\right) w_k w_j - \left(\beta_k + \theta_1^M w_k\right) \cdot \left(\sum_{r=m+1}^{N} - \gamma_{r,j} - \theta_2^M w_r w_j \right) \right] \cdot d \ln p_j \\
&+ \sum_{s=m+1}^{N} \left[\beta_{k,s} + \theta_2^M w_k w_s + \left(1 - \theta_2^M\right) w_k \delta_{k,s} - \left(\beta_k + \theta_1^M w_k\right) \cdot \left(\sum_{r=m+1}^{N} \beta_{r,s} + \theta_2^M w_r \left(w_s - \delta_{r,s}\right) \right) \right] \cdot d \ln q_s
\end{aligned}
$$

where two synthetic forms of mixed demand function can be derived as follows by applying similar logics used to derive two synthetic forms of direct and inverse demand functions.

$$
\begin{aligned}
w_i d \ln q_i &= \left[w_i \varepsilon_i - \theta_1^M w_i + \theta_1^M w_i \right] \cdot d \ln \bar{y} \\
&+ \sum_{j=1}^{m} \left[w_i \varepsilon_{i,j}^c - \theta_2^M w_i \left(w_j - \delta_{i,j}\right) + \theta_2^M w_i \left(w_j - \delta_{i,j}\right) \right] \cdot d \ln p_j \\
&+ \sum_{j=1}^{m} \left[- \left(w_i \varepsilon_i - \theta_1^M w_i + \theta_1^M w_i\right) \cdot \left(\sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c + \theta_2^M w_r w_j - \theta_2^M w_r w_j \right) \right] \cdot d \ln p_j \\
&+ \sum_{s=m+1}^{N} \left[w_i q_{i,s}^c - \theta_2^M w_i w_s + \theta_2^M w_i w_s \right] \cdot d \ln q_s \\
&+ \sum_{s=m+1}^{N} \left[- \left(w_i \varepsilon_i - \theta_1^M w_i + \theta_1^M w_i\right) \cdot \left(\sum_{r=m+1}^{N} w_r f_{r,s}^c - \theta_2^M w_r \left(w_s - \delta_{r,s}\right) + \theta_2^M w_r \left(w_s - \delta_{r,s}\right) \right) \right] \cdot d \ln q_s
\end{aligned}
$$

$$
\begin{aligned}
w_i d \ln q_i &= \left[\alpha_i + \theta_1^M w_i \right] \cdot d \ln \bar{y} \\
&+ \sum_{j=1}^{m} \left[\alpha_{i,j} + \theta_2^M w_i \left(w_j - \delta_{i,j}\right) - \left(\alpha_i + \theta_1^M w_i \right) \cdot \left(\sum_{r=m+1}^{N} - \gamma_{r,j} - \theta_2^M w_r w_j \right) \right] \cdot d \ln p_j \\
&+ \sum_{s=m+1}^{N} \left[g_{i,s} + \theta_2^M w_i w_s - \left(\alpha_i + \theta_1^M w_i \right) \cdot \left(\sum_{r=m+1}^{N} \beta_{r,s} + \theta_2^M w_r \left(w_s - \delta_{r,s}\right) \right) \right] \cdot d \ln q_s
\end{aligned}
$$

$$
\begin{aligned}
dw_i &\equiv w_i d \ln q_i + w_i d \ln p_i - w_i \left[d \ln y \right] \\
&= w_i d \ln q_i + w_i d \ln p_i - w_i \left[d \ln y - d \ln P_A + d \ln P_A \right] \\
&= w_i d \ln q_i + w_i d \ln p_i - w_i \left[d \ln \bar{y} + d \ln P_A \right] \\
&= \left\langle w_i d \ln q_i \right\rangle + w_i d \ln p_i - w_i d \ln \bar{y} - w_i d \ln P_A \\
&= \left\langle w_i d \ln q_i \right\rangle + w_i \left[\sum_{j=1}^{m} \delta_{i,j} d \ln p_j \right] - w_i d \ln \bar{y} - w_i \left[\sum_{j=1}^{m} w_j d \ln p_j \right] \\
&= \left\langle w_i d \ln q_i \right\rangle - w_i d \ln \bar{y} - \sum_{j=1}^{m} w_i \left(w_j - \delta_{i,j}\right) d \ln p_j
\end{aligned}
$$

$$
\begin{aligned}
dw_i &= \left[\alpha_i + \left(\theta_1^M - 1\right) w_i \right] \cdot d \ln \bar{y} \\
&+ \sum_{j=1}^{m} \left[\alpha_{i,j} + \left(\theta_2^M - 1\right) w_i \left(w_j - \delta_{i,j}\right) - \left(\alpha_i + \theta_1^M w_i \right) \cdot \left(\sum_{r=m+1}^{N} - \gamma_{r,j} - \theta_2^M w_r w_j \right) \right] \cdot d \ln p_j \\
&+ \sum_{s=m+1}^{N} \left[g_{i,s} + \theta_2^M w_i w_s - \left(\alpha_i + \theta_1^M w_i \right) \cdot \left(\sum_{r=m+1}^{N} \beta_{r,s} + \theta_2^M w_r \left(w_s - \delta_{r,s}\right) \right) \right] \cdot d \ln q_s
\end{aligned}
$$

and

$$
\begin{aligned}
w_k d\ln p_k &= \left[w_k f_k - \theta_1^M w_k + \theta_1^M w_k \right] \cdot d\ln \bar{y} \\
&\quad + \sum_{j=1}^{m} \left[w_k p_{k,j}^c + \theta_2^M w_k w_j - \theta_2^M w_k w_j \right] \cdot d\ln p_j \\
&\quad + \sum_{j=1}^{m} \left[-\left(w_k f_k - \theta_1^M w_k + \theta_1^M w_k \right) \cdot \left(\sum_{r=m+1}^{N} w_r p_{r,j}^c + \theta_2^M w_r w_j - \theta_2^M w_r w_j \right) \right] \cdot d\ln p_j \\
&\quad + \sum_{s=m+1}^{N} \left[w_k f_{k,s}^c - \theta_2^M w_k \left(w_s - \delta_{k,s} \right) + \theta_2^M w_k \left(w_s - \delta_{k,s} \right) \right] \cdot d\ln q_s \\
&\quad + \sum_{s=m+1}^{N} \left[-\left(w_k f_k - \theta_1^M w_k + \theta_1^M w_k \right) \cdot \left(\sum_{r=m+1}^{N} w_r f_{r,s}^c - \theta_2^M w_r \left(w_s - \delta_{r,s} \right) + \theta_2^M w_r \left(w_s - \delta_{r,s} \right) \right) \right] \cdot d\ln q_s
\end{aligned}
$$

$$
\begin{aligned}
w_k d\ln p_k &= \left[\beta_k + \theta_1^M w_k \right] \cdot d\ln \bar{y} \\
&\quad + \sum_{j=1}^{m} \left[-\gamma_{k,j} - \theta_2^M w_k w_j - \left(\beta_k + \theta_1^M w_k \right) \cdot \left(\sum_{r=m+1}^{N} -\gamma_{r,j} - \theta_2^M w_r w_j \right) \right] \cdot d\ln p_j \\
&\quad + \sum_{s=m+1}^{N} \left[\beta_{k,s} + \theta_2^M w_k \left(w_s - \delta_{k,s} \right) - \left(\beta_k + \theta_1^M w_k \right) \cdot \left(\sum_{r=m+1}^{N} \beta_{r,s} + \theta_2^M w_r \left(w_s - \delta_{r,s} \right) \right) \right] \cdot d\ln q_s
\end{aligned}
$$

$$
\begin{aligned}
dw_k &\equiv w_k d\ln p_k + w_k d\ln q_k - w_k \left[d\ln y \right] \\
&= w_k d\ln p_k + w_k d\ln q_k - w_k \left[d\ln y - d\ln P_A + d\ln P_A \right] \\
&= w_k d\ln p_k + w_k d\ln q_k - w_k \left[d\ln \bar{y} + d\ln P_A \right] \\
&= \left\langle w_k d\ln p_k \right\rangle + w_k d\ln q_k - w_k d\ln \bar{y} - w_k d\ln P_A \\
&= \left\langle w_k d\ln p_k \right\rangle + w_k \left[\sum_{s=1}^{m} \delta_{k,s} d\ln q_s \right] - w_k d\ln \bar{y} - w_k \left[\sum_{j=1}^{m} w_j d\ln p_j \right] \\
&= \left\langle w_k d\ln p_k \right\rangle - w_k d\ln \bar{y} - \sum_{j=1}^{m} w_k w_j d\ln p_j + \sum_{s=1}^{m} w_k \delta_{k,s} d\ln q_s
\end{aligned}
$$

$$
\begin{aligned}
dw_k &= \left[\beta_k + \left(\theta_1^M - 1 \right) w_k \right] \cdot d\ln \bar{y} \\
&\quad + \sum_{j=1}^{m} \left[-\gamma_{k,j} - \left(\theta_2^M + 1 \right) w_k w_j - \left(\beta_k + \theta_1^M w_k \right) \cdot \left(\sum_{r=m+1}^{N} -\gamma_{r,j} - \theta_2^M w_r w_j \right) \right] \cdot d\ln p_j \\
&\quad + \sum_{s=m+1}^{N} \left[\beta_{k,s} + \theta_2^M w_k w_s + \left(1 - \theta_2^M \right) w_k \delta_{k,s} - \left(\beta_k + \theta_1^M w_k \right) \cdot \left(\sum_{r=m+1}^{N} \beta_{r,s} + \theta_2^M w_r \left(w_s - \delta_{r,s} \right) \right) \right] \cdot d\ln q_s
\end{aligned}
$$

Theoretical restrictions can be imposed by using following relations

(a) Homogeneity: $\sum_{j=1}^{m} \alpha_{i,j} = \theta_2^M w_i \left(1 - \sum_{j=1}^{m} w_j \right) = \theta_2^M w_i \left(\sum_{s=m+1}^{N} w_s \right)$ and

$$\sum_{j=1}^{m} \gamma_{r,j} = -w_r \left[1 + \theta_2^M \sum_{j=1}^{m} w_j \right] = -w_r \left[1 + \theta_2^M \left(1 - \sum_{s=m+1}^{N} w_s \right) \right]$$

(b) Symmetry: $\alpha_{i,j} = \alpha_{j,i}$, $\beta_{r,s} = \beta_{s,r}$, and $\gamma_{r,j} = g_{j,r}$,

(c) Adding-up: $\sum_{i=1}^{m} \alpha_i + \sum_{k=m+1}^{N} \beta_k = 1 - \theta_1^M$,

$$\sum_{i=1}^{m} \alpha_{i,j} = \theta_2^M w_i \left(1 - \sum_{j=1}^{m} w_j \right) = \theta_2^M w_i \left(\sum_{k=m+1}^{N} w_k \right), \text{ and}$$

$$\sum_{i=1}^{m} g_{i,r} = -w_r \left[1 + \theta_2^M \sum_{j=1}^{m} w_j \right] = -w_r \left[1 + \theta_2^M \left(1 - \sum_{s=m+1}^{N} w_s \right) \right]$$

Because: (a) $\sum_{j=1}^{m}\alpha_{i,j} \equiv \sum_{j=1}^{m}\left[w_i\varepsilon_{i,j}^{c} - \theta_2^M w_i\left(w_j - \delta_{i,j}\right)\right] = w_i\sum_{j=1}^{m}\varepsilon_{i,j}^{c} - \theta_2^M w_i\left[\sum_{j=1}^{m}w_j - \sum_{j=1}^{m}\delta_{i,j}\right]$, which, by $\sum_{j=1}^{m}w_i \cdot \varepsilon_{i,j}^{c} = 0$ and $\sum_{j=1}^{m}\delta_{i,j} = 1$, is $\sum_{j=1}^{m}\alpha_{i,j} = -\theta_2^M w_i\left[\sum_{j=1}^{m}w_j - 1\right] = \theta_2^M w_i\left[1 - \sum_{j=1}^{m}w_j\right] = \theta_2^M w_i\left[\sum_{s=m+1}^{N}w_s\right]$ and $\sum_{j=1}^{m}\gamma_{r,j} \equiv \sum_{j=1}^{m}\left[-w_r p_{r,j}^{c} - \theta_2^M w_r w_j\right] = -\sum_{j=1}^{m}w_r p_{r,j}^{c} - \theta_2^M w_r\sum_{j=1}^{m}w_j$, which, by $\sum_{j=1}^{m}w_r \cdot p_{r,j}^{c} = w_r$, is equal to $\sum_{j=1}^{m}\gamma_{r,j} = -w_r - \theta_2^M w_r\sum_{j=1}^{m}w_j = -w_r\left[1 + \theta_2^M\sum_{j=1}^{m}w_j\right] = -w_r\left[1 + \theta_2^M\left(1 - \sum_{s=m+1}^{N}w_s\right)\right]$. (b)Using $w_i \cdot \varepsilon_{i,j}^{c} = w_j \cdot \varepsilon_{j,i}^{c}$, we can compare $\alpha_{i,j} \equiv w_i\varepsilon_{i,j}^{c} - \theta_2^M w_i\left(w_j - \delta_{i,j}\right)$ with $\alpha_{j,i} \equiv w_j\varepsilon_{j,i}^{c} - \theta_2^M w_j\left(w_i - \delta_{j,i}\right)$ as $w_i w_j - w_i\delta_{i,j}$ with. $w_j w_i - w_j\delta_{j,i}$, which is equal because $w_i\delta_{i,j} = w_j\delta_{j,i}$. Using $w_k \cdot f_{k,s}^{c} = w_s \cdot f_{s,k}^{c}$, we can compare $\beta_{r,s} \equiv w_r f_{r,s}^{c} - \theta_2^M w_r\left(w_s - \delta_{r,s}\right)$ with $\beta_{s,r} \equiv w_s f_{s,r}^{c} - \theta_2^M w_s\left(w_r - \delta_{s,r}\right)$ as $w_r w_s - w_r\delta_{r,s}$ with $w_s w_r - w_s\delta_{s,r}$, which is equal because $w_r\delta_{r,s} = w_s\delta_{s,r}$. Using $-w_r \cdot p_{r,j}^{c} = w_j \cdot q_{j,r}^{c}$, we can compare $\gamma_{r,j} \equiv -\left[w_r p_{r,j}^{c} + \theta_2^M w_r w_j\right]$ with $g_{j,r} \equiv w_j q_{j,r}^{c} - \theta_2^M w_j w_r$ as $-w_r w_j$ with. $-w_j w_r$, which is equal. (c) $\sum_{i=1}^{m}\alpha_i + \sum_{k=m+1}^{N}\beta_k \equiv \left[\sum_{i=1}^{m}w_i\varepsilon_i - \theta_1^M w_i\right] + \left[\sum_{k=m+1}^{N}w_k f_k - \theta_1^M w_k\right] = \left[\sum_{i=1}^{m}w_i\varepsilon_i + \sum_{k=m+1}^{N}w_k f_k\right] - \theta_1^M\left[\sum_{i=1}^{m}w_i + \sum_{k=m+1}^{N}w_k\right]$, which, by $\sum_{i=1}^{m}w_i\varepsilon_i + \sum_{k=m+1}^{N}w_k f_k = 1$, can be written as $\sum_{i=1}^{m}\alpha_i + \sum_{k=m+1}^{N}\beta_k = 1 - \theta_1^M$. Other two restrictions can be also derived by using symmetry relationships of $\alpha_{i,j} = \alpha_{j,i}$ and $g_{i,r} = \gamma_{r,i}$.

The elasticities are calculated as follows

(a) Expenditure elasticities: $\varepsilon_i \equiv \dfrac{\alpha_i}{w_i} + \theta_1^M$ and $f_k \equiv \dfrac{\beta_k}{w_k} + \theta_1^M$,

(b) Compensated elasticities: $\varepsilon_{i,j}^{c} \equiv \dfrac{\alpha_{i,j}}{w_i} + \theta_2^M\left(w_j - \delta_{i,j}\right)$, $f_{k,s}^{c} \equiv \dfrac{\beta_{k,s}}{w_k} + \theta_2^M\left(w_s - \delta_{k,s}\right)$,

$p_{k,j}^{c} \equiv -\dfrac{\gamma_{k,j}}{w_k} - \theta_2^M w_j$, and $q_{i,s}^{c} \equiv \dfrac{g_{i,s}}{w_i} + \theta_2^M w_s$,

(c) Uncompensated elasticities:

$$\varepsilon_{i,j} = \left[\frac{\alpha_{i,j}}{w_i} + \theta_2^M\left(w_j - \delta_{i,j}\right)\right] - \left[\frac{\alpha_i}{w_i} + \theta_1^M\right]\cdot\left[w_j + \sum_{k=m+1}^{N}w_k\cdot\left(-\frac{\gamma_{k,j}}{w_k} - \theta_2^M w_j\right)\right]$$

$$f_{k,s} = \left[\frac{\beta_{k,s}}{w_k} + \theta_2^M\left(w_s - \delta_{k,s}\right)\right] - \left[\frac{\beta_k}{w_k} + \theta_1^M\right]\cdot\left[\sum_{r=m+1}^{N}w_r\cdot\left(\frac{\beta_{r,s}}{w_r} + \theta_2^M\left(w_s - \delta_{r,s}\right)\right)\right]$$

$$p_{k,j} = \left[-\frac{\gamma_{k,j}}{w_k} - \theta_2^M w_j \right] - \left[\frac{\beta_k}{w_k} + \theta_1^M \right] \cdot \left[w_j + \sum_{r=m+1}^{N} w_r \cdot \left(-\frac{\gamma_{r,j}}{w_r} - \theta_2^M w_j \right) \right]$$

$$q_{i,s} = \left[\frac{g_{i,s}}{w_i} + \theta_2^M w_s \right] - \left[\frac{\alpha_i}{w_i} + \theta_1^M \right] \cdot \left[\sum_{r=m+1}^{N} w_r \cdot \left(\frac{\beta_{r,s}}{w_r} + \theta_2^M \left(w_s - \delta_{r,s} \right) \right) \right].$$

Because: (a) $\alpha_i \equiv \left[w_i \varepsilon_i - \theta_1^M w_i \right]$ and $\beta_k \equiv \left[w_k f_k - \theta_1^M w_k \right]$, (b) $\alpha_{i,j} \equiv \left\lfloor w_i \varepsilon_{i,j}^c - \theta_2^M w_i \left(w_j - \delta_{i,j} \right) \right\rfloor$, $\beta_{k,s} \equiv \left\lfloor w_k f_{k,s}^c - \theta_2^M w_k \left(w_s - \delta_{k,s} \right) \right\rfloor$, $-\gamma_{k,j} \equiv \left\lfloor w_k p_{k,j}^c + \theta_2^M w_k w_j \right\rfloor$, and $g_{i,s} \equiv \left\lfloor w_i q_{i,s}^c - \theta_2^M w_i w_s \right\rfloor$, and (c) $\varepsilon_{i,j} = \varepsilon_{i,j}^c - \varepsilon_i \left\lfloor w_j + \sum_{k=m+1}^{N} w_k \cdot p_{k,j}^c \right\rfloor$, $f_{k,s} = f_{k,s}^c - f_k \left\lfloor \sum_{r=m+1}^{N} w_r \cdot f_{r,s}^c \right\rfloor$, $p_{k,j} = p_{k,j}^c - f_k \left\lfloor w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right\rfloor$, and $q_{i,s} = q_{i,s}^c - \varepsilon_i \left\lfloor \sum_{r=m+1}^{N} w_r \cdot f_{r,s}^c \right\rfloor$.

The synthetic parameters for direct, inverse, and mixed demand functions can be summarized as in Table 3.1. The value of 0 and 1 for θ_1 captures constant and variational expenditure or scale coefficients and the value of 0 and 1 for θ_2 captures constant and variational Slutsky and Antonelli coefficients respectively, where the variations rely on the budget share values. Even though it is difficult to directly compare each of four types of specifications, it is possible to indirectly compare each of them to a synthetic model, because the synthetic model nests all four specifications. The joint tests for combinations of possible values of θ_1 and θ_2 can be used to compare among the synthetic model itself and four nesting types of models within each of direct, inverse, and mixed demand systems respectively.

Table 3.1. Synthetic Parameters for Three Specifications

Model	Direct		Inverse		Mixed	
	θ_1^D	θ_2^D	θ_1^I	θ_2^I	θ_1^M	θ_2^M
Rotterdam	0	0	0	0	0	0
LA/AIDS	1	1	1	1	1	1
NBR	0	1	0	1	0	1
CBS	1	0	1	0	1	0

* Restrictions of synthetic parameters to nest popular functional forms for three specifications.
** Refer to synthetic demand equation. For example, synthetic parameters in the direct demand system corresponds to parameters in
$dw_n = \left[C_n - \left(1 - \theta_1^D\right) w_n \right] d\ln Q + \sum_{n'=1}^{N} \left[C_{n,n'} - \left(1 - \theta_2^D\right) w_n \left(w_{n'} - \delta_{n,n'} \right) \right] d\ln p_{n'}$

Model Comparison Method

While the issue of an appropriate functional form within each of direct, inverse, and mixed demand systems respectively can be addressed through synthetic approaches, it is not easy to nest all three specifications of direct, inverse, and mixed demand systems in composite model. The main difficulties to compare different specifications in terms of price-dependent and/or quantity-dependent modeling across direct, inverse, and mixed demand systems are again the alternative specifications are non-nested relative to each other and non-nested hypotheses testing approach oftentimes does not provide definite answer for this problem. The Likelihood Dominance Criterion, introduced by Pollak and Wales (1991), provides alternative method to rank competing models as long as the competing specifications have the common dependent variables. Unlike the non-nesting test procedures and artificial nesting approach, the model selection criterion does not require actually estimating the composite model. Saha, Shumway, and Talpaz (1994) demonstrated that the likelihood dominance criterion outperformed some widely used non-nested testing procedures such as Davidson-MacKinnon J test and Cox test in selecting the true model, using Monte Carlo evidence. Let H_1 and H_2 denote two non-nesting hypotheses, which are nested in composite hypothesis H_C and n_1 , n_2 , n_C and L_1 , L_2 , L_C are corresponding number of independent parameters and log-likelihood values with assumption of $n_1 \le n_2$. Let $C(v,\tau)$ denote the critical values of the chi-square distribution with v degrees-of-freedom at some fixed significant level τ . The model selection approach can be summarized as follow based on the Pollak and Wales (1991).

When the standard likelihood ratio test procedure is used to compare two hypotheses with the composite, the hypothesis H_i will not be rejected iff $2L_C - 2L_i < C(n_C - n_i, \tau)$ or $L_C < L_i + (1/2)\cdot C(n_C - n_i, \tau)$ and H_i will be rejected iff $C(n_C - n_i, \tau) < 2L_C - 2L_i$ or $L_i + (1/2)\cdot C(n_C - n_i, \tau) < L_C$. This test procedure can be understood based on the intuitive reasoning that the additional parameters in composite model can be accepted, only when they increase likelihood function values. Testing separately the restrictions on the composite corresponding to the two non-nesting hypotheses can result in one of four possible outcomes:

(a) reject H_1 and accept H_2 ,

iff $C(n_C - n_1, \tau) < 2L_C - 2L_1$ and $2L_C - 2L_2 < C(n_C - n_2, \tau)$

or $L_1 + (1/2)\cdot C(n_C - n_1, \tau) < L_C < L_2 + (1/2)\cdot C(n_C - n_2, \tau)$.

(b) reject H_2 and accept H_1 ,

iff $C(n_C - n_2, \tau) < 2L_C - 2L_2$ and $2L_C - 2L_1 < C(n_C - n_1, \tau)$

or $L_2 + (1/2)\cdot C(n_C - n_2, \tau) < L_C < L_1 + (1/2)\cdot C(n_C - n_1, \tau)$.

(c) reject both H_1 and H_2,

iff $C(n_C - n_1, \tau) < 2L_C - 2L_1$ and $C(n_C - n_2, \tau) < 2L_C - 2L_2$

or both $L_1 + (1/2) \cdot C(n_C - n_1, \tau)$ and $L_1 + (1/2) \cdot C(n_C - n_1, \tau)$ are less than L_C.

(d) accept both H_1 and H_2,

iff $2L_C - 2L_1 < C(n_C - n_1, \tau)$ and $2L_C - 2L_2 < C(n_C - n_2, \tau)$

or both $L_1 + (1/2) \cdot C(n_C - n_1, \tau)$ and $L_1 + (1/2) \cdot C(n_C - n_1, \tau)$ are greater than L_C.

According to the dominance ordering, unlike non-nesting testing procedure which may result in accepting or rejecting both hypotheses, when the likelihood ratio test accepts one hypothesis and reject the other, the decision of accepting one hypothesis and rejecting the other can be determined without actually estimating or even specifying a particular composite, although the determination require specifying the number of independent parameters of the composite or the composite parameteric size n_C. Ordering dominance among competing non-nesting hypotheses can result in one of three possible outcomes:

(a) H_2 dominates H_1,

iff $L_1 + (1/2) \cdot C(n_C - n_1, \tau) < L_2 + (1/2) \cdot C(n_C - n_2, \tau)$ and $L_1 < L_2$ when $n_1 = n_2$,

because reject H_1 and accept H_2, iff $L_1 + (1/2) \cdot C(n_C - n_1, \tau) < L_C < L_2 + (1/2) \cdot C(n_C - n_2, \tau)$

(b) H_1 dominates H_2,

iff $L_2 + (1/2) \cdot C(n_C - n_2, \tau) < L_1 + (1/2) \cdot C(n_C - n_1, \tau)$ and $L_2 < L_1$ when $n_1 = n_2$,

because reject H_2 and accept H_1, iff $L_2 + (1/2) \cdot C(n_C - n_2, \tau) < L_C < L_1 + (1/2) \cdot C(n_C - n_1, \tau)$.

(c) H_1 is indifferent to H_2,

iff $L_2 + (1/2) \cdot C(n_C - n_2, \tau) = L_1 + (1/2) \cdot C(n_C - n_1, \tau)$ and $L_1 = L_2$ when $n_1 = n_2$,

based on the above two possible outcomes.

Note that if $L_2 + (1/2) \cdot C(n_C - n_2, \tau)$ is very close to $L_1 + (1/2) \cdot C(n_C - n_1, \tau)$, then the likelihood ratio test will probably either accept or reject both hypotheses. In this respective, the significance level to determine the critical value should not be interpreted as the significance level of the dominance ordering per se, but as the significance level of the fictive likelihood ratio test.

The dependence of the dominance ordering on the composite parametric size is disturbing in the general case of $n_1 \neq n_2$, since specifying it may be fairy arbitrary and different composite sizes may imply a different ordering. This difficulty can be mitigated by using the likelihood dominance criterion, based on the proposition that $(1/2)\cdot[C(n_C - n_1, \tau) - C(n_C - n_2, \tau)]$ is a monotonically decreasing function of n_C, if the significance level τ is less than 0.40 and a range of composite parametric sizes is such that from one parameter more than the larger hypotheses to one parameter more than the sum of the number of parameters in the two hypotheses ($n_2 + 1 < n_C < n_1 + n_2 + 1$). This proposition implies that for $n_2 + 1 < n_C < n_1 + n_2 + 1$, $[C(n_2 + 1, \tau) - C(n_2 + 1, \tau)] < [C(n_C - n_1, \tau) - C(n_C - n_2, \tau)] < [C(n_1 + n_2 + 1, \tau) - C(1, \tau)]$. The use of Likelihood Dominance Criterion among competing non-nesting hypotheses can result in one of three possible outcomes:

(a) H_2 is preferred to H_1,

iff $(1/2)\cdot[C(n_2 - n_1 + 1, \tau) - C(1, \tau)] < L_2 - L_1$ or $L_1 < L_2$ for $n_1 = n_2$

because $(1/2)\cdot[C(n_C - n_1, \tau) - C(n_C - n_2, \tau)] < (1/2)\cdot[C(n_2 - n_1 + 1, \tau) - C(1, \tau)] < L_2 - L_1$

(b) H_1 is preferred to H_2,

iff $L_2 - L_1 < (1/2)\cdot[C(n_2 + 1, \tau) - C(n_1 + 1, \tau)]$ or $L_2 < L_1$ for $n_1 = n_2$

because $L_2 - L_1 < (1/2)\cdot[C(n_2 + 1, \tau) - C(n_1 + 1, \tau)] < (1/2)\cdot[C(n_C - n_1, \tau) - C(n_C - n_2, \tau)]$

(c) Indecisive between H_1 and H_2,

iff $(1/2)\cdot[C(n_2 + 1, \tau) - C(n_2 + 1, \tau)] < L_2 - L_1 < (1/2)\cdot[C(n_2 - n_1 + 1, \tau) - C(1, \tau)]$

or $L_1 = L_2$ for $n_1 = n_2$

because based on the above two possible outcomes and the relationship of $[C(n_2 + 1, \tau) - C(n_2 + 1, \tau)] < [C(n_C - n_1, \tau) - C(n_C - n_2, \tau)] < [C(n_1 - n_2 + 1, \tau) - C(1, \tau)]$.

Note that to narrow this indecisive range, the significant level τ be adjustably selected and/or the composite parametric size n_C can be determined directly from the significance tables for the chi-square distribution for given n_1, n_2 and L_1, L_2.

It can be seen that the likelihood dominance criterion has similar implication with the two common model selection criteria of Akaike Information criterion (Akaike, 1973) and Schwarz information criterion (Schwarz, 1978). The Akaike and Schwarz model selection rules of choosing the largest value of $L_i - n_i$ and $L_i - (\log T/2)\cdot n_i$ can be understood as pair-wise comparison rules for $L_2 - L_1$ in terms of relative penalty functions $(n_2 - n_1)$ and $(\log T/2)\cdot(n_2 - n_1)$ respectively. These two relative penalty functions have similar implications as the likelihood dominance criterion, since as Pollak and Wales (1991) argued that

$(1/2)\cdot[C(n_C - n_1, \tau) - C(n_C - n_2, \tau)]$ converges to $(1/2)\cdot(n_2 - n_1)$ as $n_C \to \infty$ based on the asymptotic normality property as a function of degrees-of-freedom of the chi-squared distribution. Based on this argument, it can be argued that the use of three model selection rules can result in one of three possible outcomes:

(a) H_2 is preferred to H_1,

iff $(1/2)\cdot(n_2 - n_1) < L_2 - L_1$ for likelihood dominance criterion of $n_C \to \infty$

or $(\log T/2)\cdot(n_2 - n_1) < L_2 - L_1$ for Schwarz model selection rule

or $(n_2 - n_1) < L_2 - L_1$ for Akaike model selection rule.

(b) H_1 is preferred to H_2,

iff $L_2 - L_1 < (1/2)\cdot(n_2 - n_1)$ for likelihood dominance criterion of $n_C \to \infty$

or $L_2 - L_1 < (\log T/2)\cdot(n_2 - n_1)$ for Schwarz model selection rule

or $L_2 - L_1 < (n_2 - n_1)$ for Akaike model selection rule.

(c) Indecisive between H_1 and H_2,

iff $L_2 - L_1 = (1/2)\cdot(n_2 - n_1)$ for likelihood dominance criterion of $n_C \to \infty$

or $L_2 - L_1 = (\log T/2)\cdot(n_2 - n_1)$ for Schwarz model selection rule

or $L_2 - L_1 = (n_2 - n_1)$ for Akaike model selection rule.

Note that non-nesting hypotheses and composite hypothesis should involve the same dependent variables for the above discussions. While Rotterdam-type synthetic models have different dependent variables across direct, inverse, and mixed demand systems, AIDS-type synthetic models have the common dependent variables across direct, inverse, and mixed demand systems. If the hypotheses involve different dependent variables but are functionally related, then the likelihood function must be adjusted by including the appropriate Jacobian bias term. To avoid difficulties involved this adjustment, the model selection approach is used for synthetic models of AIDS-type dependent variables for the comparison across direct, inverse, and mixed demand systems.

Summary and Proposed Method

There are significant advances in the study of demand from both theoretical and empirical perspective. In the theoretical perspective, the full modeling spectrums of monotone set of direct or inverse demand functions as well as mixed demand functions are developed.

While these theoretical advances bring modeling flexibility to the study of consumer behavior, they also bring forth the local identification issue of choosing one empirical model among three possible specifications of the direct, inverse, and mixed demand systems. Given that there is an empirical difficulty in studying all possible combinations for the mixed demand system as well as the direct and inverse demand systems, graphical causal models provide an inductive way to infer local causal structure among price and quantity variables for a particular commodity. After the local identification issue is guided by the graphical causal models, the model selection approaches, such as the likelihood dominance criterion, provide empirical method of comparing empirical demand model specifications. The AIDS type dependent variable synthetic functional forms for the direct, inverse, and mixed demand systems provide a flexible and comparable functional form to connect the graphical causal model and the model selection approaches, thus minimizing the effects of the chosen functional forms for three specifications. Note that the direct and inverse demand systems can be always used for comparison purposes, regardless of the identified causal structures from the graphical causal model. On the other hand, the identified causal structures from the graphical causal model provide inductive information for the possible combination of price and quantity dependent specifications for the mixed demand system. This inductive information based on the graphical causal models provides empirical guidance for the local identification issue, given that the researchers' intuition for this issue does not always provide objective specifications.

Recent advances in data processing capabilities have brought the possibility of analyzing larger number of detailed variables. The retail checkout scanner data have brought forth research potentials for significant advances in the micro-economic analysis of consumer behavior. Given the observation that many variables in this high dimensional data move very closely, the compositional stability condition, as a consistent aggregation condition, provides an inductive way to pursue the possibility of obtaining not only (a) interpretable aggregate indexes or macro-variables as the representative aggregate of homogeneous disaggregate micro-variables but also (b) interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for the subsequence analysis. This implies that when the micro-variables can be legitimately grouped and represented by macro-variables, it is possible to use aggregation methods (a) to incorporate broad range of information into the empirical demand models, while minimizing econometric issues such as the multicollinearity and degrees of freedom and (b) to capture (causal) relationships among disaggregated variables through (causal) relationships among aggregated variables as the legitimate representatives. This compositional stability condition is used (a) to provide an inductive way of forming suitable partitions before conducting any empirical test to justify those classifications based on the empirical data patterns rather than on researchers' intuition and (b) to address the possible violation of the (probabilistic) stability condition to use the graphical causal models for the high dimensional data. Note that it is conceivable and oftentimes observed that the (probabilistic) stability condition for the graphical causal models is violated for using high dimensional data in empirical study, given the observation that there exist close co-movements and thus near deterministic relations among variables in high dimensional data.

More specific procedures we propose are as follows: (a) Both standard static correlation matrix and dynamic correlation matrix over identified frequency bands are used to measure co-movement among original variables. Based on these similarity measures of disaggregate micro-variables, the modified k-nearest neighbor algorithm is used to sort the variables such that the highly correlated variables are near each other along the main diagonal in reordered correlation matrix. The block-diagonal pattern of reordered or sorted static and dynamic correlation matrixes are used to identify homogeneous groups of variables, based the approximate form of the compositional stability condition. (b) Based on identified classifications of original variables, index number theory is used for actual aggregation procedure to decide weighting schemes or aggregating disaggregated micro-variables into representative macro-variables within each identified group. (c) The identified classification and aggregation of micro-variables into macro-variables can be tested, as long as appropriate instrumental variables can be identified. The Hausman type misspecification test of $H_0 : \gamma_n = 0$ in the equation $x_n = X H_n + IV \cdot \gamma_n + \varepsilon_n^{IV}$, where x_n and X are micro- and macro-variables respectively and IV are Instrumental Variables such that IV is closely correlated with X and independent of d_n, provides statistical test framework for the generalized form of the compositional stability condition of independence between d_n and X in the set of equations $x_n = X H_n + d_n$. (d) Based on the implication that identified compositional stability condition in the data makes it possible to infer causal structures among micro-variables through relationships among representative aggregated macro-variables. PC algorithm or GES algorithm are used to infer causal structures among macro-variables as the legitimate representative causal relationships among micro-variables are used for the subsequent analysis. (e) Based on the local causal structure between price and quantity variables for a particular commodity, the AIDS type dependent variable synthetic functional forms for the direct, inverse, and mixed demand systems are estimated. (f) The Rotterdam, AIDS, NBR, and CBS type constant and/or variational parameterizations and synthetic model are statistically compared and the parameterizations for expenditure (scale) elasticities (flexibilities) and Slutsky (Antonelli) coefficients are chosen within each of direct, inverse, and mixed specifications. Based on the chosen parameterization, the direct, inverse, and mixed demand system are compared based on the model selection approaches, such as the Akaike Information, Schwarz information criterion, and the likelihood dominance criterion. Note that inductive properties are emphasized in every sequence of the proposed method, since some types of deductive properties can bring subjectivities or ambiguities into the empirical results. The remaining subjectivities in our proposed method are left as further research issues, with the hope that the remaining subjectivities bring fewer ambiguities relative to the previously used methods. The proposed method is illustrated with the applications for retail checkout scanner data as an example of the high dimensional data.

Empirical Analysis and Results

The proposed methodological procedure is illustrated with the soft drink products with size of 6/12 oz sold at Dominick's Finer Foods (DFF). Given that some types of deductive properties can bring subjectivities or ambiguities into the empirical results, inductive properties are emphasized. First, the data used for this study are described. Second, the common frequency bands for the estimated spectrum of variables are identified and the static and dynamic correlations among variables are measured and sorted for the classification. Third, based on the block diagonal pattern of the sorted correlation matrixes, the variables are classified and classified groups are interpreted, where variables within each of groups closely co-moves. Fourth, based on the classified groups, the index number theory is used to represent disaggregate variables by aggregate variables. And the compositional stability condition is empirically tested and the test results are compared with Lewbel's composite commodity conditions. Fifth, the local causal structure among price and quantity variables for each of aggregate commodities is inferred by the graphical causal model. Sixth, based on the local causal structure used for identification, the direct, inverse, and mixed demand systems are estimated based on the synthetic demand system approach. The estimated results of three specifications of demand system are related and compared. The empirical results are summarized and additional issues to be studied are discussed.

Data Description

The data set consists of weekly observations on 23 soft drink products with size of 6/12 oz sold at Dominick's Finer Foods (DFF) from 09:14:1989 through 09:22:1993 with the sample size 210. All the data are from the Dominick's database, which is publicly available from the University of Chicago Graduate School of Business (http://www.chicagogsb.edu/). The Dominick's Finer Foods (DFF) is the second largest supermarket chain in the Chicago metropolitan area with about 25% market share. Each soft drink used for this study is a specific soft drink of 6/12 oz size such as Coca-cola classic, Pepsi-cola cans, Seven-up diet can. The brand-level categories include Coke, Pepsi, Seven-up, Mountain Dew, Sprite, Rite-Cola, Dr. Pepper, A&W, Canada Dry, Sunkist, and Lipton Brisk. The size of 6/12 oz is chosen due to the data availability and identified homogeneity within this size of soft drinks in the preliminary study.

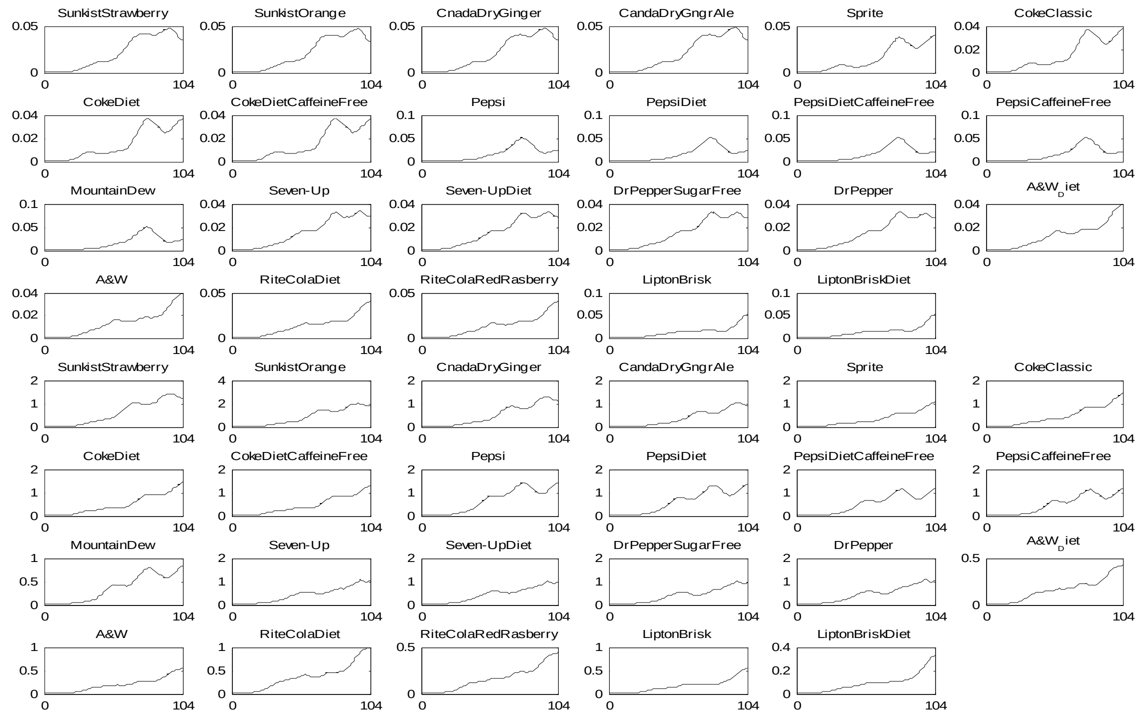
Although the original data set is the store level weekly retail scanner data for the specific items represented by UPC code, the aggregated chain level data is used for this study. In order to characterize the chain level characteristics, the store level data are aggregated across stores by using the simple sum and unit value for quantity and price variables, where unit value is total sale revenue divided by the total quantity sold. The reasons for this is the commodity-wise aggregation is the main issue to be addressed in this study and the aggregation across consumers or regions brings forth more difficult issues, which can be addressed only with additional information such as demographical and economical information. Another practical reason for this is that there are many missing observations in the original data set due to different data collection period or other reasons. Aggregation based on the unit value approach is one way to deal with this missing observation problem, whereas the use of other index formulas brings forth the difficult issue of how to handle a zero price or quantity in the data set.

For the purpose of estimating differential demand systems, the differential terms for price and quantity variables are approximated by the finite first differences ($d\ln p_n \approx \ln p_{n,t} - \ln p_{n,t-1}$ and $d\ln q_n \approx \ln q_{n,t} - \ln q_{n,t-1}$) and the market share terms are replaced by their moving average ($w_n \approx \left(w_{n,t} + w_{n,t-1}\right)/2$). The market share changes dw are approximated by using the log differential property ($dw = w \cdot d \ln w \approx (1/2) \cdot (w_{n,t} + w_{n,t-1}) \cdot (\ln w_{n,t} - \ln w_{n,t-1})$), since dw has a limited range of $(-1, 1)$, whereas $dw = w \cdot d \ln w$ has a range of $(-\infty, \infty)$ (Barten, 1993). The list of variables and detailed descriptions are given in Appendix D.

Classification and Aggregation

One of objectives of this study is to propose an inductive procedure for the construction of appropriate groupings of variables. An inductive property is emphasized due to the empirical implausibility of attempting all possible partitions before conducting any empirical test to justify those classifications. In this respect, it can be better to pursue inductive classifications related with legitimate aggregation conditions, which is based on the empirical data pattern itself rather than researchers' subjective intuition. Based on the compositional stability conditions, our inductive procedure is based on the idea that homogeneity or similarity of a group of variables can be identified through their dynamic movements. When the original disaggregate variables within a group have similar dynamic movements so that they co-move with each other very closely, their high co-movements suggest their underlying similarity.

Both the standard static correlation matrix and the dynamic correlation matrix over identified frequency bands are used to measure co-movement among the original variables. For the dynamic correlation over frequency band, several different frequency bands are chosen as the non-overlapping bands or regions approximately centered at peak λ_k so that $\{ \Lambda = [\lambda_i, \lambda_j) \cup [-\lambda_j, -\lambda_i) : 0 \le \lambda_i < \lambda_k < \lambda_j \le \pi \}$, where the frequency λ_k is specified as $\{ \lambda_k = 2\pi \cdot k/T : k = 1, \cdots, (T/2) \}$ and T is the sample size (Rodrigues, 1999). Note that if the frequency of a cycle is λ, the period of the cycle is $2\pi/\lambda$. Thus, a frequency of $\lambda_k = 2\pi \cdot k/T$ corresponds to a period of $2\pi/\lambda_k = T/k$. We choose common frequency bands to measure co-movement among variables with possible leads and lags, based on the estimated spectrums of variables, which capture dynamics of variables in terms of their cyclic properties with long or short run trends (Hamilton, 1994). The estimated spectrums of price and quantity variables are presented in Figure 3.1. The x-axis is the frequency in terms of k and the y-axis is the estimated spectrum.



* The full description of variables is provided in the Appendix D.
* The top 23 variables are the price variables and the bottom 23 variables are the quantity variables.
* The x- axis is the frequency in terms of k and the y-axis is the estimated spectrum.

Figure 3.1. Estimated Spectrums of Price and Quantity Variables

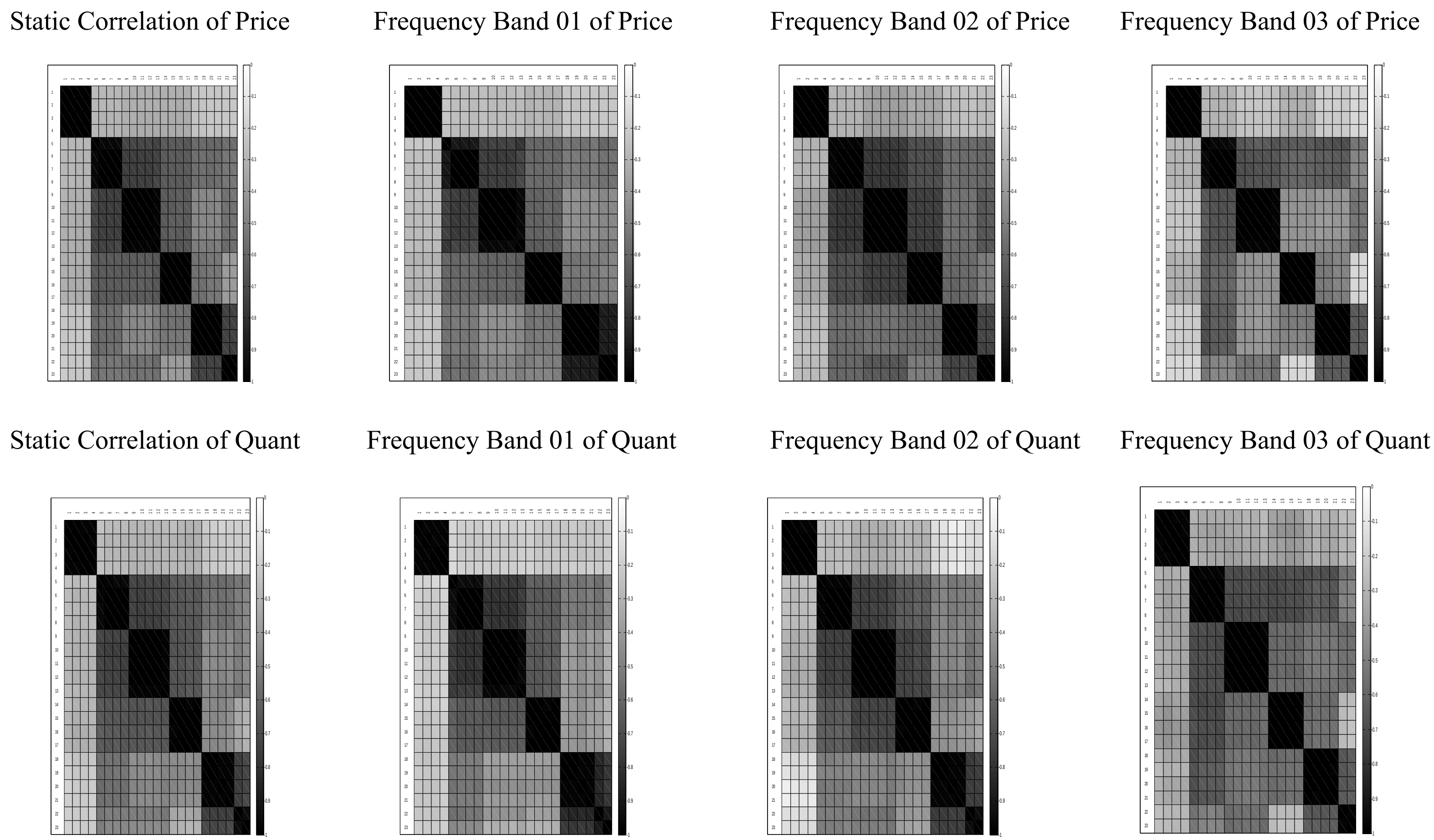
The top 23 variables are the price variables and the bottom 23 variables are the quantity variables. The full description of variables is provided in the Appendix D. Although there are some degrees of differences, the common frequency bands can be identified across price and quantity variables and thus among 23 commodities. We use three frequency bands: 0-62, 63-90, and 90-104.5 in terms of k . These correspond to a period more than 3.37 weeks (frequency Band 01), a period of 3.32 to 2.32 weeks (frequency Band 02), a period of less than 2.30 weeks (frequency Band 03) respectively. These ranges approximately correspond to 1 month, a half month, and less that a half month period ranges.

Based on these homogeneity or similarity measure of disaggregate micro-variables, the modified k-nearest neighbor algorithm is used to sort or reordered the variables such that the highly correlated variables are near each other along the main diagonal in the reordered correlation matrix. The block-diagonal pattern of sorted static and dynamic correlation matrixes are used to identify homogeneous group of variables, based on the approximate form of the compositional stability condition. The final results of the sorted static correlation matrix and dynamic correlation matrixes for different frequency bands are presented in Figure 3.2. The black/white color scheme is used to represent the absolute value of measured correlations, where the darkest black represents the correlation of 1 and the brightest white represents the correlation of 0. The full description of variables is provided in the Appendix D, where the variables are in the same order. More detailed information of measured correlation for the standard static correlation coefficient for the price variables (lower triangular matrix) and quantity variables (upper triangular matrix) is presented in Table 3.2.

The homogeneity within the group is identified based on the high co-movements of price and/or quantity variables in terms of measured pair-wise static and dynamic correlations among variables. For example in the static correlation of price and quantity variables, the correlations among pair of products within the identified group are larger than 0.954 and 0.948 respectively. Although the correlations of pair-wise variables across different groups show somewhat different degrees of correlation over the different frequency bands, the common groups of variables are identified over all the different frequency bands. It is also noticed that both price and quantity variables show similar correlation patterns. Based on the sorted static and dynamic correlation matrixes of price and quantity variables over the different frequency bands, the following six groups of soft drink products are identified as homogeneous groups.

Group 6: The Sunkist and Canada Dry product group (Product of 1 to 4)
Group 1: The Coca-Cola and Sprite product group (Products of 5 to 8)
Group 2: The Pepsi-Cola and Mountain Dew product group (Product of 9 to 13)
Group 3: The Seven-Up and Dr Pepper product group (Products of 14 to 17)
Group 5: The A&W and Rite-Cola product group (Products of 18 to 21)
Group 4: The Lipton Brisk product group (Products of 22 to 23)

The group of 1 and 2 are discriminated by their relatively different relationship with group 5, given that the variables in group 1 have higher correlation with the variables in group 5. The group of 2 and 3 are discriminated by their relatively different relationship with group 4, given that the variables in group 2 have higher correlation with the variables in group 4. The group of 5 and 4 are discriminated by their relatively different relationship with group 2, given that the variables in group 4 have higher correlation with the variables in group 2.



* The black/white color scheme is used to represent the absolute value of measured correlation, where the darkest black represents the correlation of 1 and the brightest white represents the correlation of 0.
* See Appendix D for the description of variables, where variables are in the same order.

Figure 3.2. Sorted Static and Dynamic Correlation Matrix

Table 3.2. Sorted Static Correlation Matrix

Var. #	Variable Names	dln(01)	dln(02)	dln(03)	dln(04)	dln(05)	dln(06)	dln(07)	dln(08)	dln(09)	dln(10)	dln(11)	dln(12)	dln(13)	dln(14)	dln(15)	dln(16)	dln(17)	dln(18)	dln(19)	dln(20)	dln(21)	dln(22)	dln(23)
01	SunkistStrawberry	1.000	0.988	0.983	0.975	0.248	0.272	0.269	0.274	0.282	0.287	0.281	0.289	0.269	0.264	0.282	0.300	0.297	0.212	0.191	0.187	0.189	0.196	0.187
02	SunkistOrange	0.998	1.000	0.988	0.982	0.270	0.297	0.294	0.302	0.308	0.313	0.311	0.316	0.294	0.298	0.317	0.332	0.327	0.237	0.222	0.215	0.210	0.207	0.202
03	CnadaDryGinger	0.998	0.999	1.000	0.994	0.264	0.291	0.287	0.291	0.304	0.310	0.306	0.311	0.293	0.288	0.302	0.317	0.314	0.239	0.223	0.220	0.214	0.218	0.216
04	CandaDryGngrAle	0.998	0.999	1.000	1.000	0.248	0.277	0.275	0.280	0.303	0.310	0.308	0.312	0.294	0.282	0.295	0.311	0.307	0.222	0.206	0.205	0.197	0.202	0.205
05	Sprite	0.279	0.282	0.287	0.291	1.000	0.971	0.968	0.967	0.734	0.740	0.730	0.734	0.728	0.654	0.653	0.637	0.645	0.570	0.568	0.587	0.575	0.537	0.507
06	CokeClassic	0.292	0.295	0.300	0.304	0.955	1.000	0.998	0.995	0.750	0.757	0.746	0.749	0.724	0.671	0.671	0.656	0.662	0.552	0.548	0.569	0.557	0.513	0.483
07	CokeDiet	0.291	0.295	0.300	0.304	0.954	0.999	1.000	0.995	0.748	0.756	0.745	0.748	0.722	0.661	0.661	0.647	0.652	0.550	0.544	0.568	0.555	0.506	0.480
08	CokeDietCaffeineFree	0.293	0.296	0.301	0.305	0.954	0.998	0.999	1.000	0.750	0.758	0.750	0.753	0.724	0.662	0.668	0.657	0.659	0.548	0.543	0.565	0.549	0.509	0.477
09	Pepsi	0.312	0.314	0.319	0.321	0.734	0.756	0.754	0.751	1.000	0.994	0.991	0.989	0.975	0.677	0.683	0.660	0.666	0.453	0.465	0.491	0.461	0.505	0.462
10	PepsiDiet	0.319	0.322	0.326	0.328	0.734	0.755	0.754	0.753	0.998	1.000	0.997	0.996	0.982	0.676	0.679	0.660	0.668	0.458	0.466	0.492	0.467	0.513	0.467
11	PepsiDietCaffeineFree	0.322	0.324	0.329	0.331	0.732	0.753	0.753	0.753	0.995	0.999	1.000	0.997	0.981	0.670	0.674	0.655	0.661	0.449	0.459	0.481	0.454	0.504	0.460
12	PepsiCaffeineFree	0.324	0.326	0.330	0.333	0.732	0.752	0.752	0.752	0.995	0.999	0.999	1.000	0.981	0.675	0.679	0.664	0.671	0.458	0.465	0.484	0.461	0.509	0.458
13	MountainDew	0.325	0.328	0.332	0.334	0.746	0.735	0.735	0.733	0.978	0.981	0.981	0.982	1.000	0.676	0.675	0.656	0.667	0.479	0.492	0.510	0.489	0.536	0.488
14	Seven-Up	0.319	0.321	0.325	0.329	0.652	0.648	0.644	0.642	0.646	0.644	0.641	0.641	0.662	1.000	0.993	0.982	0.988	0.467	0.477	0.490	0.464	0.359	0.316
15	Seven-UpDiet	0.321	0.325	0.329	0.332	0.648	0.645	0.642	0.641	0.643	0.642	0.640	0.641	0.661	0.998	1.000	0.990	0.991	0.458	0.469	0.481	0.449	0.353	0.304
16	DrPepperSugarFree	0.326	0.329	0.333	0.337	0.652	0.644	0.642	0.641	0.638	0.640	0.639	0.641	0.663	0.995	0.996	1.000	0.995	0.453	0.458	0.468	0.439	0.354	0.296
17	DrPepper	0.325	0.328	0.332	0.336	0.659	0.650	0.648	0.646	0.643	0.645	0.644	0.646	0.669	0.995	0.996	0.999	1.000	0.463	0.469	0.476	0.453	0.365	0.304
18	A&W_Diet	0.233	0.238	0.241	0.243	0.592	0.572	0.570	0.568	0.471	0.475	0.473	0.475	0.511	0.558	0.555	0.562	0.564	1.000	0.990	0.977	0.984	0.754	0.712
19	A&W	0.234	0.240	0.242	0.245	0.593	0.574	0.572	0.569	0.472	0.476	0.473	0.476	0.512	0.561	0.557	0.564	0.567	1.000	1.000	0.979	0.979	0.755	0.721
20	RiteColaDiet	0.222	0.228	0.230	0.233	0.601	0.588	0.586	0.584	0.482	0.486	0.483	0.484	0.516	0.541	0.537	0.539	0.544	0.990	0.989	1.000	0.979	0.754	0.722
21	RiteColaRedRasberry	0.224	0.230	0.232	0.235	0.598	0.579	0.578	0.576	0.476	0.479	0.477	0.479	0.515	0.538	0.534	0.540	0.546	0.994	0.994	0.996	1.000	0.750	0.717
22	LiptonBrisk	0.216	0.220	0.224	0.224	0.573	0.546	0.544	0.543	0.556	0.559	0.557	0.560	0.583	0.399	0.395	0.402	0.406	0.747	0.747	0.747	0.750	1.000	0.948
23	LiptonBriskDiet	0.218	0.223	0.226	0.227	0.568	0.541	0.539	0.538	0.547	0.552	0.550	0.553	0.577	0.394	0.391	0.398	0.402	0.748	0.748	0.747	0.751	0.999	1.000

* The lower triangular is for the static correlation coefficients of price variables and the upper triangular is for the static correlation coefficients of quantity variables
* The shaded areas represent the identified groups.
* See Appendix D for the description of variables, where variables are in the same order.

Note that the ordering of variables and groups, which is listed in Appendix D, correspond to the ordering in correlation matrix. The numbering for each of the groups follows the different ordering for the consistency in notations for the subsequent analyses.

The above classification results can be interpreted as follows: (a) The products of group 1 and 2 correspond to the products of Coca-Cola company (Coca-Cola and Sprite) and Pepsi company (Pepsi-Cola and Mountain Dew) respectively. (b) The products of group 3 and 5 correspond to the products of competing companies (Seven-Up and Dr Pepper) and following companies (A&W and Rite-Cola) respectively, given that the Coca-Cola and Pepsi companies can be interpreted as the market leaders. (c) The products of group 6 and 4 correspond to the products of different substitutive groups for the carbonate soft drink products. The Sunkist and Canada Dry brands are identified as a homogenous group, although they represent two different types of substitute for the carbonate soft drink products. The Lipton Brisk product group shows different relationships across other groups and thus it is identified distinct group, although this group is closely related with group 5.

The resulting classification based on the inductive procedure can be compared with other standard classifications, which rely on the researchers' intuitive choices, for the soft drink products in the literature. For example, one standard classifications scheme for the soft drink products is based on intuitive choices among possible combinations of assumed multi-stage budgeting structures as follows: (a) All soft drinks are classified as the branded, private label, and all-other products and (b) The branded soft drinks are classified as Cola and Clear sub-segments. (c) The Cola sub-segment consists of Coke, Pepsi, RC Cola and Dr Pepper. On the other hand, the Clear sub-segment consists of Sprite, 7-Up and Mt. Dew (Dhar, Chavas, and Gould, 2003). Comparing with this and other deductive classification, the inductive classification of this study has following distinctive features: (a) The Cola and Clear sub-segments are not identified. (i) Sprite and Mountain Dew brands belong in their companies' brands, Coca-Cola and Pepsi-Cola respectively. (ii) The Seven-Up brand forms a distinct group with the Dr Pepper brand. (iii) The Rite-Cola brand forms a distinct group with the A&W brand. (b) The substitutive products for the carbonate soft drink products are classified as two distinctive groups, where one group consists of Sunkist and Canada Dry brands and the other group consists of Lipton Brisk product. (c) Diet or caffeine free products do not form distinctive groups. Note that Dhar, Chavas, and Gould (2003) find that classifications based on the Cola and Clear sub-segments are empirically rejected. In this respect, it can be argued that the classification inductively identified from the data itself in this study provides another plausible classification scheme for soft drink products.

The consistent aggregation condition can be empirically tested, where the classification is based on the sorted correlation matrices and the aggregation is based on the index number theory. Note that different index number formulas are used for actual aggregation procedure to decide weighting schemes for aggregating disaggregated original variables into representative aggregate variables within each identified group. It is for the robustness check of test results, given that the test is actually a joint test for both classification and aggregation. The following different index number formulas are used: Tornqvist-Theil (dd), Fisher (ff), Paasche (pp), Laspeyres (ll), Fisher with chain (fc), Paasche with chain (pc), Laspeyres with chain (lc), Unit value (uv), Quantity share weighted index (qw), and Expenditure share weighted index (ew). The Tornqvist-Theil index is primary used in this study. The preference toward the Tornqvist-Theil index, especially rather than the Fisher index, is due to facts that unlike the Fisher index, the Tornqvist-Theil index does not invoke the problematic assumption of a homothetic or linear homogeneous utility function as discussed in chapter II. Two types of consistent aggregation conditions are empirically tested and compared. Note that both tests are conducted for both price and quantity variables due to our interest in the alternative specification among direct, inverse, and mixed demand system.

First, the compositional stability condition of $Cov(d_n, X)=0$ in $x_n = X H_n + d_n$ is empirically tested by using Hausman type misspecification test of $H_0 : \gamma_n = 0$ in $x_n = X H_n + IV \cdot \gamma_n + \varepsilon_n^{IV}$, where x_n are disaggregated micro-variables of a specific group and X are corresponding aggregated macro-variables of a specific group. The IV is an Instrumental Variable such that IV is closely correlated with regressor X (relevance condition of IV) and independent of error d_n (validity condition of IV). In this study, we choose to use the total expenditure variable, which is calculated by aggregating the price and quantity macro-variables, as the instrumental variable based on the following reasoning: (a) Given that the total expenditure is closely related with the aggregated price and quantity variables as in estimated aggregated demand systems, the relevance condition holds. (b) The validity condition of total expenditure variable as instrumental variable can also hold, either when each of the idiosyncratic variations of disaggregated price or quantity variable are canceled each other in calculating total expenditure, or when the idiosyncratic variation of individual price or quantity variable, which is not captured by the common variation of representative macro-variables of a specific group, does not have dependencies on the total expenditure variable, which captures the common variation of an entire group of commodities within the demand system through group-representative price and quantity macro-variables.

The empirical results of the compositional stability condition are presented in Table 3.3. The empirical test results of the compositional stability condition can be summarized as follows, given that a high p-value across almost all test implies a high probability of $H_0 : \gamma_n = 0$ in $x_n = X H_n + IV \cdot \gamma_n + \varepsilon_n^{IV}$, which in turn implies that $Cov\left(d_n, X\right)=0$ in $x_n = X H_n + d_n$: (a) The possible bias due to classification and aggregation for price variable can be ignored and thus the use of aggregate price variable for representing each group can be justified, when price variables are used as explanatory variables. (b) The possible bias due to classification and aggregation for quantity variable can be ignored and thus the use of aggregate quantity variable for representing each group can be justified, when quantity variables are used as explanatory variables. (c) The classification itself, which is inductively identified, can be empirically justified in terms of both price and quantity variables, given that the results are robust with respect to different index number formulas for aggregation.

For the comparison with the empirical finding for the Clear soft drink group in Dhar, Chavas, and Gould (2003), the Sprite, Mt. Dew, 7-up, and 7-up diet are tested as a one homogeneous group based on the compositional stability condition. The p-values for $H_0 : \gamma_n = 0$ are 0.0018 (Sprite), 0.0001 (Mt. Dew), 0.00027 (7-up), and 0.0029 (7-up diet) in terms of the price variables and 0.000 for all the products in terms of quantity variables, when the Tornqvist-Theil index is used for price and quantity aggregates. This result is consistent with the empirical rejection of homogeneity of Sprite, Mt Dew, and 7-up products in Dhar, Chavas, and Gould (2003) and thus provides additional evidence for the non-existence of the Clear sub-group.

Second, Lewbel's generalized compositional commodity condition for differential demand system is tested based on the correlation test of $H_0 : Corr\left(d_n^{Lewbel}, X\right)=0$, where $d_n^{Lewbel} \equiv x_n - X$. The empirical results of the unit root test (UR-test) for micro- and macro-variables imply stationarity of transformed variables in differential demand system, where unit root test results for disaggregate variables are in the column vector and those for aggregate variables are in the row vector under the heads of UR-Test for each group (Table 3.5). These results of unit root test are robust with respect to other specifications in unit root test. These results are consistent with the observation in the demand literature that the differential demand system has been considered as appropriate specification to deal with the possible non-stationarity problems.

The empirical results of the generalized compositional commodity condition are presented in Table 3.4. The empirical test results for Lewbel's generalized compositional commodity condition can be summarized as follows, given that high p-value implies high probability of $H_0 : Corr\left(d_n^{Lewbel}, X\right) = 0$ where $d_n^{Lewbel} \equiv x_n - X$: (a) The possible bias due to classification and aggregation for price variable can be ignored and thus the use of aggregate price variable for representing each group can be justified, when price variables are used as explanatory variables. (b) The possible bias due to classification and aggregation for quantity variable can not be ignored and thus the use of aggregate quantity variable for representing each group can not be justified, when quantity variables are used as explanatory variables. (c) The test results are ambiguous for classification itself. The classification itself can be empirically justified in terms of price variables but it can not be justified in terms of quantity variables.

The different implications from the two test approaches of the compositional stability condition and Lewbel's generalized compositional commodity condition for quantity variables can be explained based on the interpretation of the Lewbel's condition in the context of Theil's aggregation theory. As discussed, the ambiguity exists in the arbitrary choice on the proportionality factors $c = 1$ in relationship between micro-variables and macro-variable for each group $x_n = X c_n + \varepsilon_n$. The choice of $c = 1$ is restrictive in the context of Theil's aggregation theory, because it implies that the true macro-parameters should be a simple sum of micro-parameters. However, there is no a prior reason that the true macro-parameters can not be a simple average of micro-parameters ($c = 1/N$), for example. On the other hand, the compositional stability condition considers these proportional factors to be the empirically estimated, without imposing any numerical restrictions except their stability. When a high probability of the proportionality factor $c = 1$ in $x_n = X c_n + \varepsilon_n$ is empirically found, the same test results for the consistent aggregation condition are expected from the two test approaches. On the other hand, the low p-value of $H_0 : c_n = 1$ can explain the different results from the two test approaches. The empirical test results of $H_0 : c_n = 1$ in $x_n = X c_n + \varepsilon_n$ are presented in Table 3.5. In general, high p-values are found for price variables, which can explain the same implications of two test approaches. On the other hand, low p-values are found for quantity

Table 3.3. Test for Compositional Stability Condition

		Price variables										Quantity variables									
Var. #	Variable Names	dd	ff	pp	ll	fc	pc	lc	uv	qw	ew	dd	ff	pp	ll	fc	pc	lc	uv	qw	ew
01	SunkistStrawberry	0.146	0.070	0.153	0.070	0.149	0.205	0.178	0.152	0.064	0.048	0.014	0.012	0.013	0.012	0.012	0.012	0.012	0.015	0.012	0.031
02	SunkistOrange	0.077	0.207	0.174	0.595	0.076	0.063	0.142	0.172	0.761	0.778	0.689	0.692	0.688	0.704	0.688	0.686	0.691	0.688	0.696	0.730
03	CnadaDryGinger	0.050	0.113	0.113	0.052	0.048	0.081	0.057	0.111	0.022	0.020	0.700	0.695	0.695	0.699	0.698	0.704	0.695	0.699	0.709	0.898
04	CandaDryGngrAle	0.296	0.427	0.375	0.805	0.289	0.254	0.314	0.378	0.659	0.638	0.549	0.537	0.549	0.540	0.536	0.543	0.533	0.545	0.538	0.379
05	Sprite	0.468	0.542	0.990	0.143	0.535	0.597	0.156	0.993	0.145	0.190	0.256	0.241	0.131	0.414	0.296	0.156	0.443	0.133	0.139	0.665
06	CokeClassic	0.577	0.645	0.552	0.137	0.673	0.695	0.587	0.585	0.111	0.155	0.927	0.935	0.877	0.805	0.951	0.894	0.909	0.878	0.893	0.560
07	CokeDiet	0.672	0.738	0.500	0.247	0.765	0.644	0.496	0.535	0.213	0.269	0.781	0.795	0.992	0.651	0.822	0.737	0.879	0.991	0.759	0.402
08	CokeDietCaffeineFree	0.978	0.977	0.382	0.898	0.959	0.513	0.323	0.418	0.990	0.961	0.913	0.912	0.821	0.946	0.911	0.961	0.764	0.818	0.945	0.893
09	Pepsi	0.218	0.264	0.937	0.119	0.267	0.815	0.194	0.933	0.127	0.165	0.082	0.080	0.100	0.076	0.092	0.096	0.080	0.099	0.077	0.020
10	PepsiDiet	0.628	0.606	0.627	0.132	0.673	0.827	0.892	0.652	0.175	0.181	0.206	0.219	0.250	0.175	0.222	0.252	0.171	0.245	0.292	0.041
11	PepsiDietCaffeineFree	0.713	0.786	0.356	0.825	0.715	0.511	0.352	0.362	0.752	0.832	0.735	0.716	0.718	0.766	0.730	0.713	0.791	0.709	0.663	0.653
12	PepsiCaffeineFree	0.275	0.333	0.164	0.186	0.289	0.275	0.067	0.164	0.160	0.198	0.148	0.153	0.165	0.132	0.156	0.169	0.124	0.177	0.183	0.066
13	MountainDew	0.051	0.113	0.190	0.020	0.066	0.187	0.019	0.216	0.012	0.017	0.624	0.594	0.487	0.745	0.599	0.467	0.758	0.484	0.552	0.680
14	Seven-Up	0.057	0.039	0.071	0.033	0.054	0.015	0.027	0.064	0.041	0.047	0.206	0.261	0.205	0.211	0.202	0.127	0.271	0.236	0.217	0.131
15	Seven-UpDiet	0.152	0.165	0.123	0.233	0.153	0.225	0.149	0.112	0.271	0.244	0.088	0.065	0.090	0.085	0.096	0.093	0.086	0.092	0.084	0.048
16	DrPepperSugarFree	0.147	0.169	0.132	0.235	0.140	0.069	0.058	0.128	0.235	0.261	0.594	0.641	0.587	0.600	0.588	0.550	0.630	0.605	0.603	0.392
17	DrPepper	0.069	0.085	0.066	0.154	0.065	0.031	0.026	0.059	0.156	0.168	0.986	0.984	0.997	0.971	0.986	0.972	0.998	0.997	0.977	0.661
18	A&W_Diet	0.029	0.035	0.042	0.040	0.027	0.011	0.046	0.042	0.035	0.061	0.019	0.017	0.018	0.017	0.008	0.026	0.020	0.018	0.017	0.014
19	A&W	0.019	0.022	0.028	0.025	0.017	0.005	0.026	0.028	0.023	0.056	0.066	0.049	0.060	0.064	0.062	0.075	0.053	0.064	0.058	0.039
20	RiteColaDiet	0.064	0.051	0.054	0.069	0.062	0.075	0.042	0.052	0.068	0.196	0.022	0.018	0.023	0.024	0.013	0.025	0.027	0.025	0.025	0.064
21	RiteColaRedRasberry	0.206	0.129	0.186	0.074	0.202	0.367	0.156	0.190	0.106	0.151	0.015	0.014	0.015	0.013	0.013	0.015	0.011	0.015	0.013	0.013
22	LiptonBrisk	0.795	0.717	0.897	0.583	0.795	0.681	0.763	0.898	0.562	0.555	0.039	0.033	0.052	0.034	0.034	0.033	0.035	0.033	0.033	0.046
23	LiptonBriskDiet	0.398	0.426	0.329	0.576	0.403	0.386	0.350	0.332	0.554	0.548	0.105	0.092	0.138	0.090	0.094	0.094	0.097	0.096	0.092	0.127

* Aggregate variables are calculated based on different index number formulas.
For example, dd represents aggregation based on Tornqvist-Theil Index number. See the discussion part in the text for detail pp.108.

* All the values are the p-values for $H_0 : \gamma_n = 0$ in $x_n = X H_n + IV \cdot \gamma_n + \varepsilon_n^{IV}$, where IV is the total expenditure variable as the instrumental variable.

Table 3.4. Test for Lewbel's Composite Commodity Condition

		Price variables										Quantity variables									
Var. #	Variable Names	dd	ff	pp	ll	fc	pc	lc	uv	qw	ew	dd	ff	pp	ll	fc	pc	lc	uv	qw	ew
01	SunkistStrawberry	0.458	0.559	0.550	0.572	0.457	0.441	0.478	0.550	0.494	0.495	0.197	0.202	0.203	0.202	0.196	0.194	0.199	0.203	0.203	0.019
02	SunkistOrange	0.126	0.087	0.077	0.098	0.126	0.128	0.126	0.077	0.100	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
03	CnadaDryGinger	0.070	0.264	0.269	0.305	0.071	0.094	0.071	0.269	0.200	0.200	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
04	CandaDryGngrAle	0.807	0.908	0.900	0.909	0.807	0.796	0.831	0.900	0.963	0.963	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
05	Sprite	0.748	0.670	0.209	0.595	0.659	0.483	0.774	0.212	0.614	0.610	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
06	CokeClassic	0.854	0.804	0.206	0.547	0.754	0.433	0.378	0.204	0.552	0.551	0.005	0.006	0.036	0.000	0.007	0.014	0.004	0.036	0.036	0.959
07	CokeDiet	0.740	0.699	0.177	0.797	0.654	0.382	0.305	0.176	0.802	0.802	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
08	CokeDietCaffeineFree	0.694	0.658	0.175	0.930	0.619	0.368	0.303	0.174	0.934	0.934	0.038	0.038	0.038	0.036	0.038	0.046	0.030	0.038	0.038	0.000
09	Pepsi	0.072	0.094	0.352	0.076	0.090	0.333	0.067	0.370	0.079	0.079	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	PepsiDiet	0.688	0.659	0.951	0.603	0.706	0.996	0.996	0.920	0.783	0.783	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
11	PepsiDietCaffeineFree	0.334	0.391	0.361	0.175	0.366	0.392	0.132	0.344	0.146	0.146	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	PepsiCaffeineFree	0.127	0.159	0.188	0.044	0.149	0.207	0.037	0.178	0.037	0.037	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	MountainDew	0.225	0.263	0.367	0.144	0.251	0.394	0.123	0.354	0.133	0.133	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
14	Seven-Up	0.112	0.113	0.085	0.150	0.112	0.108	0.122	0.088	0.152	0.152	0.732	0.726	0.739	0.712	0.732	0.733	0.730	0.737	0.737	0.888
15	Seven-UpDiet	0.976	0.966	0.990	0.947	0.976	0.978	0.979	0.998	0.935	0.934	0.727	0.720	0.734	0.706	0.727	0.729	0.725	0.732	0.732	0.578
16	DrPepperSugarFree	0.559	0.543	0.542	0.542	0.559	0.561	0.555	0.536	0.584	0.585	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
17	DrPepper	0.066	0.067	0.069	0.065	0.066	0.067	0.065	0.067	0.064	0.064	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010
18	A&W_Diet	0.972	0.967	0.931	0.968	0.974	0.825	0.904	0.931	0.888	0.889	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
19	A&W	0.678	0.660	0.633	0.662	0.680	0.559	0.788	0.633	0.613	0.614	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20	RiteColaDiet	0.725	0.856	0.864	0.888	0.724	0.869	0.632	0.864	0.822	0.823	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
21	RiteColaRedRasberry	0.800	0.862	0.988	0.753	0.799	0.944	0.709	0.988	0.743	0.743	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
22	LiptonBrisk	0.268	0.204	0.191	0.220	0.269	0.239	0.306	0.191	0.226	0.226	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
23	LiptonBriskDiet	0.196	0.243	0.273	0.218	0.196	0.217	0.182	0.273	0.224	0.224	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

* Aggregate variables are calculated based on different index number formulas.
For example, dd represents aggregation based on Tornqvist-Theil Index number. See the discussion part in the text for detail pp.108.

* All the values are the p-values for $H_0 : Corr\left(d_n^{Lewbel}, X\right) = 0$ where $d_n^{Lewbel} \equiv x_n - X$.

Table 3.5. Tests for the Unit Root and the Proportionality Factors

Price Variables		dd	ff	pp	ll	fc	pc	lc	uv	qw	ew	Quantity Variables		dd	ff	pp	ll	fc	pc	lc	uv	qw	ew
dlnP06	UR-Test	-11.55	-11.54	-11.55	-11.54	-11.55	-11.53	-11.57	-11.55	-11.54	-11.54	dlnQ06	UR-Test	-10.95	-10.95	-10.95	-10.95	-10.95	-10.95	-10.95	-10.95	-10.95	-10.93
dln(p_01)	-11.61	0.57	0.67	0.66	0.50	0.58	0.34	0.44	0.65	0.52	0.51	dln(q_01)	-11.13	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.00
dln(p_02)	-11.52	0.93	0.78	0.76	0.55	0.95	0.72	0.77	0.77	0.67	0.68	dln(q_02)	-10.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_03)	-11.54	0.27	0.35	0.68	0.46	0.26	0.40	0.26	0.68	0.78	0.79	dln(q_03)	-10.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_04)	-11.51	0.43	0.14	0.18	0.11	0.43	0.47	0.41	0.18	0.12	0.13	dln(q_04)	-10.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dlnP01	UR-Test	-11.10	-11.09	-10.98	-11.14	-11.09	-13.81	-10.72	-10.98	-11.14	-11.14	dlnQ01	UR-Test	-10.86	-10.85	-10.84	-10.87	-10.85	-10.76	-10.90	-10.84	-10.84	-10.88
dln(p_05)	-10.69	0.86	0.84	0.31	0.35	0.79	0.93	0.77	0.31	0.36	0.36	dln(q_05)	-10.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_06)	-11.15	0.53	0.54	0.08	0.85	0.47	0.26	0.16	0.08	0.82	0.82	dln(q_06)	-10.89	0.02	0.03	0.10	0.00	0.03	0.02	0.04	0.10	0.10	0.81
dln(p_07)	-11.16	0.67	0.66	0.10	0.37	0.59	0.30	0.20	0.10	0.37	0.37	dln(q_07)	-10.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_08)	-11.04	0.94	0.92	0.15	0.19	0.83	0.40	0.35	0.15	0.20	0.19	dln(q_08)	-10.90	0.31	0.32	0.30	0.31	0.32	0.43	0.18	0.29	0.29	0.00
dlnP02	UR-Test	-13.19	-13.19	-13.11	-11.45	-13.17	-13.10	-13.20	-13.11	-11.46	-11.46	dlnQ02	UR-Test	-10.38	-10.38	-10.37	-10.39	-10.38	-10.37	-10.38	-10.37	-10.37	-10.39
dln(p_09)	-11.59	0.34	0.47	0.94	0.29	0.42	0.94	0.12	0.92	0.30	0.31	dln(q_09)	-10.28	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_10)	-11.43	0.72	0.78	0.62	0.21	0.80	0.62	0.78	0.59	0.29	0.28	dln(q_10)	-10.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_11)	-13.10	0.53	0.55	0.30	0.72	0.52	0.31	0.25	0.29	0.64	0.64	dln(q_11)	-10.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_12)	-13.11	0.16	0.18	0.12	0.15	0.17	0.12	0.04	0.11	0.13	0.13	dln(q_12)	-10.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_13)	-12.51	0.28	0.41	0.51	0.20	0.33	0.55	0.16	0.48	0.18	0.17	dln(q_13)	-14.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dlnP03	UR-Test	-11.25	-11.26	-11.26	-11.26	-11.25	-11.27	-11.22	-11.26	-11.26	-11.26	dlnQ03	UR-Test	-13.53	-13.52	-13.53	-13.52	-13.53	-13.53	-13.52	-13.53	-13.53	-13.47
dln(p_14)	-11.27	0.25	0.28	0.27	0.30	0.26	0.33	0.13	0.28	0.29	0.27	dln(q_14)	-13.39	0.46	0.46	0.47	0.45	0.42	0.46	0.48	0.50	0.50	0.53
dln(p_15)	-11.25	0.85	0.75	0.82	0.73	0.85	0.88	0.92	0.82	0.75	0.75	dln(q_15)	-13.38	0.80	0.80	0.80	0.82	0.80	0.81	0.80	0.80	0.80	0.93
dln(p_16)	-11.17	0.84	0.88	0.86	0.91	0.84	0.97	0.65	0.85	0.98	0.98	dln(q_16)	-13.54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_17)	-11.26	0.06	0.06	0.06	0.07	0.06	0.09	0.03	0.06	0.07	0.06	dln(q_17)	-13.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
dlnP05	UR-Test	-11.92	-11.93	-11.93	-11.94	-11.92	-11.87	-11.94	-11.93	-11.93	-11.93	dlnQ05	UR-Test	-10.45	-10.45	-10.45	-10.45	-10.45	-10.44	-10.46	-10.45	-10.45	-10.47
dln(p_18)	-11.91	0.81	0.69	0.80	0.62	0.80	0.91	0.70	0.80	0.71	0.71	dln(q_18)	-10.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_19)	-11.99	0.64	0.72	0.63	0.78	0.64	0.47	0.82	0.63	0.74	0.74	dln(q_19)	-11.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_20)	-11.90	0.83	0.98	0.90	0.98	0.83	0.88	0.75	0.90	0.90	0.90	dln(q_20)	-10.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_21)	-9.86	0.74	0.69	0.56	0.85	0.75	0.53	0.85	0.56	0.88	0.88	dln(q_21)	-10.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dlnP04	UR-Test	-12.63	-12.63	-12.63	-12.64	-12.63	-12.64	-12.62	-12.63	-12.64	-12.64	dlnQ04	UR-Test	-11.69	-11.69	-11.69	-11.69	-11.69	-11.69	-11.69	-11.69	-11.69	-11.71
dln(p_22)	-12.63	0.04	0.03	0.03	0.02	0.04	0.05	0.06	0.03	0.01	0.07	dln(q_22)	-11.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
dln(p_23)	-12.64	0.04	0.05	0.07	0.02	0.04	0.06	0.04	0.07	0.01	0.07	dln(q_23)	-15.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

* Aggregate variables are calculated based on different index number formulas.
For example, dd represents aggregation based on Tornqvist-Theil Index number. See the discussion part in the text for detail pp.108.

* Unit Root test (UR-Test) is based on no constant and no trend with BIC lag length selection specification, where critical values are -2.58 (1%), -1.95 (5%), -1.62 (10%).
The column vector of UR-Test is for disaggregate variables and row vector of UR-Test is for aggregate variables.

* All other values are the p-values for $H_0 : c_n = 1$ in $x_n = X \cdot c_n + \varepsilon_n$.

variables, which can explain the different implications of two test approaches of the compositional stability condition and Lewbel's generalized compositional commodity condition.

Causality for Identification

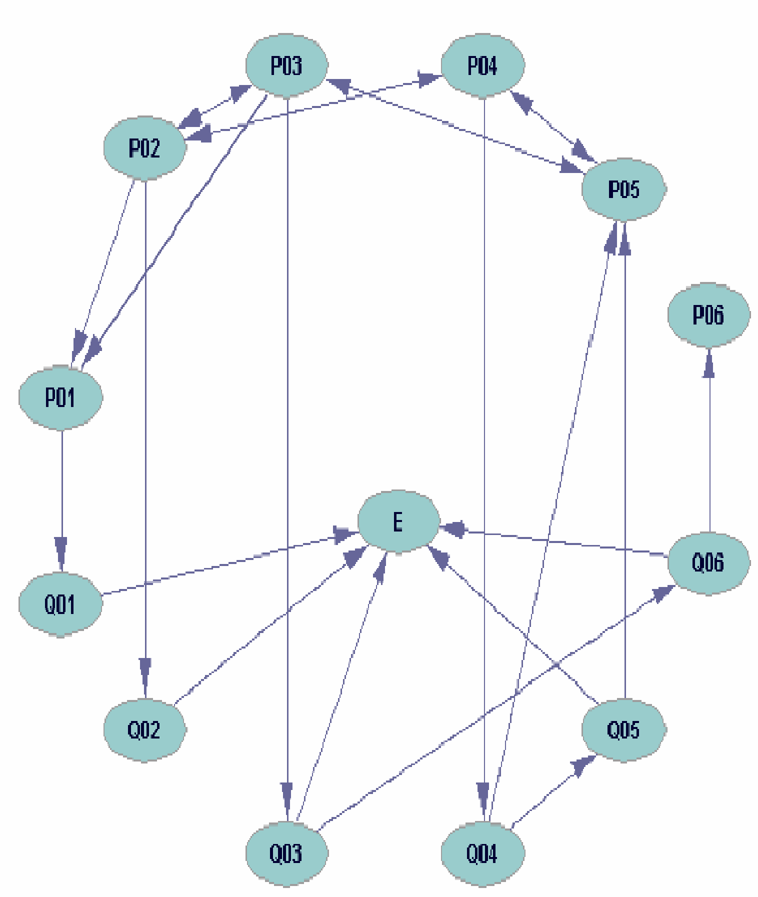
In consumer behavior study, the demand theory provides the full modeling spectrums of monotone set of direct or inverse demand functions and mixed demand functions with their properties such as homogeneity, symmetry, negativity, adding-up, and relation of compensated and uncompensated demand functions. The choice among direct, inverse, and mixed specifications has been usually based on the researchers' intuition about product properties or market characteristics of a specific commodity. However, given that (a) the choice of specifications is not trivial in empirical modeling to measure consumers' responsiveness. (b) some types of deductive properties can bring subjectivities or ambiguities into the empirical results, it is better to pursue an inductive procedure for this identification issue.

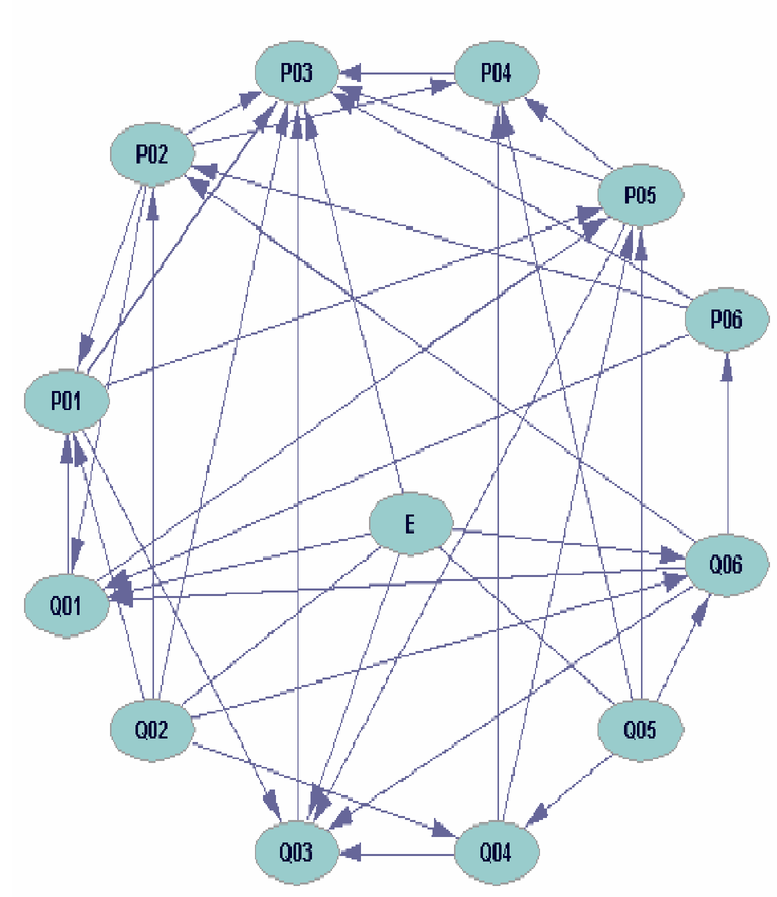
The specification choice is closely related with the identification issue of the local causal structure between price and quantity for a specific commodity. When we choose either quantity-dependent or price-dependent specification, we implicitly assume a local causal structure, since the direct (or inverse) demand function is implied by the causal structure that price (or quantity) variable causes quantity (or price) variable. Here we use graphical causal models to inductively derive this local causal structure. This empirically derived causal structure through the proposed methods of DAG can be used to decide dependent and explanatory variable for a specific commodity demand function within the demand system. Note that in the preliminary study for causal structures in the disaggregated original level data set, some causal relationships between price p_1 and quantity q_1 for the same commodity are statistically broken. It is because high correlation between p_1 and p_2 can induce $P(p_1 \mid p_2, q_1) = P(p_1 \mid p_2)$ through $P(p_1 \mid p_2) \approx 1$, when the other commodity's price p_2 has a high co-movement with p_1. Given the observation that many variables in soft drink products move very closely as empirically measured in aggregation section, the (probabilistic) stability condition of the graphical causal model is violated and thus DAG method can not be used for disaggregate level data set. Note that this problem is similar to the multicollinearity problem, which makes it difficult to obtain precise estimates of the separate effects of the variables in the regression method.

The compositional stability condition provides the possibility to address this issue in using the graphical causal model. The use of aggregate variables to infer causal relationships among observed disaggregate variables can be justified based on the compositional stability condition discussed in the aggregation theory. The identified block diagonal pattern of correlation matrixes and the empirically tested compositional stability condition discussed in aggregation section imply that the observed disaggregate variables meet the condition of compositional stability condition. This condition in turn implies that there exists not only the possibility of obtaining interpretable macro-variables as the representative aggregate of homogeneous disaggregate micro-variables, but also the possibility of getting interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for the subsequence analysis. This means that when the disaggregate variables can be legitimately grouped and represented by aggregate variables, it is possible to use aggregate variables to capture (causal) relationships among disaggregate variables through (causal) relationships among aggregate variables as the legitimate representatives as long as the compositional stability conditions hold among disaggregate variables. Based on the identified compositional stability condition discussed in aggregation section, (causal) relationships among disaggregate micro-variables through relationships among representative aggregated macro-variables are inferred. The PC algorithm or GES algorithm is used to infer local causal structures among macro-variables as the legitimate representative causal relationships among micro-variables. The empirical results are presented in Figure 3.3. and 3.4.

Before interpreting local causal information between price and quantity for each product for the full use of theoretical information from the demand theory, the reason to restrict causal information to local one need to be discussed. We do not pursue structural equation models approach based on the full causal structures identified from two resulting causal structures of PC and GES algorithms, since (a) One of main objectives of this study is to propose inductive methods to infer local causal structure between price and quantity for the full use of theoretical development in three possible specifications of direct, inverse, and mixed demand functions. And thus the issues to be addressed in this study are restricted to ones related with this objective. (b) There remain several undecided causal directions in both results and such directions can not be decided without additional causal information. The undirected edges in the result of the GES algorithms represent the limitations to identify causal directions based on the statistical observations only (observational equivalence). On the other hand, the bi-directed edges in the result of PC algorithm imply the existence of unobserved factors. The capability of identifying unobserved factors between two variables, based on the tetrad relationship among partial correlations, is one advantage of the PC algorithm relative to the GES algorithm. On the other hand, given the Markov condition (causal sufficiency and acyclic assumptions), the GES algorithm has following advantages relative to the PC algorithm (i) The GES algorithm does not require the choice of the significant level. This is advantage, given that the result of PC algorithm oftentimes is sensitive to the choice of the significant level. (ii) The GES algorithm oftentimes provides finer results than the PC algorithm. The difference is due to the fact that the GES algorithm is based on the numerical scores on the overall hypothetic causal structures, whereas the PC algorithm is based on the categorical decision on individual edges and such categorical decisions can be sensitive to the chosen significant level. In our results, the GES algorithm provides all the edges (skeleton) identified by the PC algorithm with additional edges. Sometimes these additional edges are important to decide the causal directions among variables. For example of the empirical results for soft drink data, the edge $P01-Q02$ is crucial to orient $Q01 \rightarrow P01$ in the GES algorithm, because this orientation is based on the unshielded collider pattern of $Q01 \rightarrow P01 \leftarrow Q02$. In the PC algorithm, the edge $P01-Q02$ is statistically removed and this categorical decision can be sensitive to the specified significant level. Similar patterns such as $P02-P06$ for $Q02 \rightarrow P02 \leftarrow P06$ and $Q02-P03$ for $Q02 \rightarrow P03 \leftarrow Q03$ can be used to explain the different implications for local causal structure between price and quantity between PC and GES algorithms. In this respect, the results of the PC algorithm need to be carefully used for the choice of the significant level. In fact, the local causal structure between price and quantity variables inferred by the PC algorithm is not robust to the change of the significant level. In this study, the final result of PC algorithm is based on the significant level of 0.1, which is recommended for sample size of 100-300 (Spirtes et al., 2000).

For the full use of theoretical information from the demand theory, all we need is the local causal structures between price and quantity variables for each commodity. This local information provides the possibility to inductively address the choice issue among three possible specifications of direct, inverse, and mixed demand functions. The local causal structures between price and quantity variables among six aggregated commodity groups identified by PC algorithm implies the mixed demand system, where quantity dependent specifications are suggested for aggregate commodities of groups of 01, 02, 03, and 04 and price dependent specifications are suggested for aggregate commodities of groups of 05 and 06. On the other hand, the local causal structure identified by GES algorithm implies the inverse demand system, where price dependent specifications are suggested for all the aggregate commodities.





* P and Q denotes representative price and quantity indices for each group defended as
Group 01: Coca-Cola and Sprite, Group 02: Pepsi-Cola and Mountain Dew, Group 03: Seven-Up and Dr Pepper,
Group 04: Lipton Brisk., Group 05: A&W and Rite-Cola, Group 06: Sunkist and Canada Dry, and E denote total expenditure variable.
* The result of PC algorithm is based on the significant level of 0.1, which is recommended for sample size of 100-300 (Spirtes et al., 2000).

Figure 3.3. Causal Structure Inferred by PC Algorithm

Figure 3.4. Causal Structure Inferred by GES Algorithm

Given that the direct demand system or quantity dependent specification is oftentimes used in empirical models, the possibility of the price dependent or mixed demand specification implied from the GES algorithm and the PC algorithm results need to be interpreted. One possible interpretation is that (a) The soft drinks are differentiated products, where the differentiated products are defined as the products differentiated by taste, packing and brand-base advertisement to influence consumers' perception of different brands, and (b) The retail prices for differentiated products can be determined by strategic pricing rules of firms incorporating supply and demand characteristics for these products (Dhar, Chavas, and Gould, 2003).

Note that Dhar, Chavas, and Gould (2003) use the reduced form specification for price and expenditure equations to deal with possible endogeneity problem in price and expenditure variables. Based on the Durbin, Wu, and Hausman test, they empirically found price and expenditure endogeneity problem. While price endogeneity problem can be addressed by the price dependent specification, the expenditure endogeneity problem is not fully addressed in this study. The reason for this is that (a) The instrumental variables in the expenditure equation need to be exogenous. To identify the exogeneity of those instrumental variables, we need additional causal information, which requires more information of additional variables. Or exogeneity of instrumental variables is assumed as like the exogeneity of expenditure variables is assumed. In addition, (b) Developing fully structural models, where price and expenditure equations are specified in the analytical and estimable forms with flexible demand specifications, results in econometric models, which is difficult to work with either analytically or empirically due to its highly non-linearity (Dhar, Chavas, and Gould, 2003). However, the main reason why we do not pursue instrumental variable approach is the same reason why we do not pursue structural equation models approach based on the full causal structures identified: one of main objectives of this study is to propose inductive methods for the full use of theoretical development in three possible specifications of direct, inverse, and mixed demand functions. And thus the issues to be addressed in this study are restricted to ones related with this objective.

Direct, Inverse, and Mixed Demand Systems

Heretofore, the consistent aggregation condition of the compositional stability condition is used to define variables and the empirically derived causal structure through DAG on the aggregated variables is used to decide dependent and explanatory variable for a specific commodity demand function within the demand system. There remains another issue of deciding functional form to relate dependent variable with explanatory variables for the empirical study of consumer behavior. Another objective in this study is to propose flexible and comparable functional forms for the direct, inverse, and mixed demand system.

When we want to compare direct, inverse, and mixed demand systems, we need parameterize direct, inverse, and mixed demand systems in the similar degrees of flexibility in functional form specifications, when the flexibility means the capability of empirical model to allow the possible combinations of constant/variational parameterization for income (or scale) coefficient and Slutsky (or Antonelli) coefficient. While the Rotterdam type parameterization assumes that both income (or scale) coefficient and compensated price (or quantity) coefficient in direct (or inverse) demand system are constant parameters, the LA/AIDS parameterization assumes that both income (or scale) coefficient and Slutsky (or Antonelli) coefficient in direct (or inverse) demand system are variational parameters dependent on the budget shares. For both direct and inverse systems, the synthetic approach in differential family provides the flexible way of parameterization to incorporate the logically possible combinations of constant and/or variational parameterization for income (or scale) coefficient and Slutsky (or Antonelli) coefficient. Based on the similar logic to derive synthetic demand model in direct and inverse demand systems, the synthetic differential demand model is proposed for the mixed demand system. When we want to compare direct, inverse, and mixed demand systems, the Likelihood Dominance Criterion, introduced by Pollak and Wales (1991), provides plausible method to rank competing models as long as the competing specifications have the common dependent variables. If the hypotheses involve different dependent variables but are functionally related, then the likelihood function must be adjusted by including the appropriate Jacobian bias term. To avoid difficulties involved with this adjustment, the synthetic direct and inverse demand systems are reparameterized to have common differential AIDS dependent variables, given that the Rotterdam type dependent variable of synthetic models have different dependent variables among direct, inverse, and mixed demand function. Rotterdam type and AIDS type dependent variable synthetic models can be directly derived from the Rotterdam specification without requiring consistent and simultaneous specifications for both direct and indirect utility functions. By extending the common logic of these approaches, a similar synthetic functional form for the mixed demand system is specified in the common differential AIDS dependent variables.

The synthetic models of direct, inverse, and mixed demand systems of the common differential AIDS type dependent variable are proposed for the flexible and comparable functional form for the direct, inverse, and mixed demand system, which makes it possible to compare direct, inverse, and mixed demand systems in model selection frameworks. The direct demand system is estimated for the comparison purpose with the inverse and mixed demand systems, which are chosen based on the local causal structure of the GES and PC algorithms respectively. The estimated parameters in all three direct, inverse, and mixed synthetic demand systems of the common differential AIDS type dependent variable are presented in Table 3.6. All three types of demand systems are estimated by the nonlinear seemingly unrelated regression estimation method with allowing autoregressive errors (SHAZAM). The first order autocorrelation is used with the restriction that the autocorrelation coefficients are constrained to be the same in all equations. The homogeneity, symmetry, and adding-up properties are used for the economy of parameters in empirical models. One equation is dropped in estimation step and recovered by using homogeneity, symmetry, and adding-up conditions for the direct and inverse demand. Since the adding-up condition in direct or inverse demand makes the demand system singular. On the other hand, for the mixed demand, the adding-up condition holds only at a point and thus does not induce the singularity in the resulting system. All the equations are used in estimation for the mixed demand. The number of independent parameters in all the demand system is 23, which include the two synthetic parameters and one autocorrelation correction term.

For the comparison of different parameterization assumptions of the constant and/or variation for the income (or scale) coefficient and Slutsky (or Antonelli) coefficient within each of direct, inverse, and mixed demand system, the Wald statistic, which is distributed chi-square with the same degrees of freedom as the number of restrictions, is used. For the comparison of competing models of three different specifications of the direct, inverse, and mixed demand system, three model selection rules, the Akaike Information, Schwarz information criterion, and the Pollak and Wales' likelihood dominance criterion, are used. The results of the model selection rules can be interpreted as the ranking among the competing models, rather than the rejection or accepting one of the competing models. Given that all three competing models have the same number of independent parameters, all three model selection rules are used based on the comparison of the estimated log-likelihood function values such as the higher log-likelihood value, the higher ranking among competing models. The empirical results of these comparison statistics are presented in Table 3.7.

Table 3.6. Parameter Estimates

Coefficient	Direct Model Estimate	Std. Error	t-Statistic	p-value
th1	1.3852	0.0338	41.0025	0.0000
th2	4.7255	0.1193	39.6028	0.0000
c01	-0.1119	0.0110	-10.2124	0.0000
c02	-0.0813	0.0114	-7.1276	0.0000
c03	-0.0771	0.0086	-8.9905	0.0000
c04	-0.0363	0.0021	-17.0796	0.0000
c05	-0.0700	0.0070	-9.9813	0.0000
c06*	-0.0086	0.0071	-1.2171	0.2236
c11	0.1552	0.0486	3.1933	0.0014
c12	0.0393	0.0319	1.2314	0.2182
c13	-0.0851	0.0289	-2.9473	0.0032
c14	-0.0083	0.0108	-0.7693	0.4417
c15	-0.0626	0.0241	-2.5933	0.0095
c16*	-0.0385	0.0209	-1.8430	0.0653
c22	0.0690	0.0466	1.4810	0.1386
c23	-0.0027	0.0276	-0.0965	0.9232
c24	-0.0375	0.0111	-3.3800	0.0007
c25	-0.0464	0.0256	-1.8114	0.0701
c26*	-0.0218	0.0208	-1.0484	0.2944
c33	0.1427	0.0435	3.2816	0.0010
c34	-0.0133	0.0095	-1.3949	0.1631
c35	-0.0224	0.0209	-1.0699	0.2847
c36*	-0.0192	0.0215	-0.8928	0.3720
c44	0.1097	0.0156	7.0345	0.0000
c45	-0.0322	0.0132	-2.4347	0.0149
c46*	-0.0183	0.0065	-2.8149	0.0049
c55	0.1622	0.0391	4.1501	0.0000
c56*	0.0014	0.0224	0.0644	0.9487
c66*	0.0964	0.0340	2.8321	0.0046
rho	-0.3569	0.0303	-11.7773	0.0000

Coefficient	Inverse Model Estimate	Std. Error	t-Statistic	p-value
th1	0.9609	0.0084	113.9911	0.0000
th2	0.1852	0.0068	27.0705	0.0000
d01	-0.0144	0.0027	-5.3288	0.0000
d02	-0.0102	0.0030	-3.4277	0.0006
d03	-0.0104	0.0023	-4.5423	0.0000
d04	-0.0072	0.0007	-10.2280	0.0000
d05	-0.0085	0.0019	-4.5498	0.0000
d06*	0.0116	0.0041	2.8590	0.0043
d11	-0.0046	0.0024	-1.9450	0.0518
d12	-0.0019	0.0013	-1.4661	0.1426
d13	0.0002	0.0013	0.1935	0.8465
d14	0.0002	0.0005	0.4582	0.6468
d15	-0.0003	0.0012	-0.2691	0.7879
d16*	0.0064	0.0012	5.5284	0.0000
d22	-0.0019	0.0024	-0.8231	0.4105
d23	-0.0034	0.0012	-2.8513	0.0044
d24	0.0006	0.0005	1.2309	0.2183
d25	-0.0005	0.0012	-0.3980	0.6906
d26*	0.0071	0.0013	5.4636	0.0000
d33	-0.0003	0.0022	-0.1390	0.8894
d34	0.0005	0.0004	1.0416	0.2976
d35	-0.0025	0.0011	-2.2266	0.0260
d36*	0.0055	0.0011	4.8862	0.0000
d44	-0.0052	0.0010	-4.9888	0.0000
d45	0.0017	0.0007	2.5608	0.0104
d46*	0.0022	0.0004	5.5577	0.0000
d55	0.0005	0.0018	0.2620	0.7933
d56*	0.0011	0.0011	0.9812	0.3265
d66*	-0.0224	0.0035	-6.4711	0.0000
rho	-0.3614	0.0296	-12.2266	0.0000

Coefficient	Mixed Model Estimate	Std. Error	t-Statistic	p-value
th1	0.1086	0.0502	2.1641	0.0305
th2	-0.1618	0.0464	-3.4893	0.0005
a01	0.2790	0.0183	15.2047	0.0000
a02	0.3470	0.0200	17.3620	0.0000
a03	0.2233	0.0165	13.5553	0.0000
a04	0.0280	0.0060	4.6852	0.0000
b05	-0.0010	0.0047	-0.2020	0.8399
b06*	0.0150	0.0031	4.9234	0.0000
a11	-1.1976	0.0683	-17.5455	0.0000
a12	0.6802	0.0566	12.0154	0.0000
a13	0.4324	0.0504	8.5805	0.0000
a14*	0.0759	0.0259	2.9288	0.0034
a22	-1.2570	0.0726	-17.3187	0.0000
a23	0.4667	0.0583	8.0043	0.0000
a24*	0.1007	0.0279	3.6095	0.0003
a33	-0.9733	0.0751	-12.9635	0.0000
a34*	0.0678	0.0232	2.9224	0.0035
a44*	-0.2708	0.0253	-10.7210	0.0000
b55	-0.0374	0.0075	-5.0124	0.0000
b56	0.0071	0.0021	3.4392	0.0006
b66	-0.0383	0.0066	-5.8044	0.0000
r51	-0.0079	0.0085	-0.9321	0.3513
r52	-0.0393	0.0103	-3.7990	0.0002
r53	-0.0331	0.0107	-3.0846	0.0020
r54*	-0.0156	0.0056	-2.7905	0.0053
r61	-0.0137	0.0077	-1.7791	0.0752
r62	-0.0215	0.0085	-2.5286	0.0115
r63	-0.0394	0.0092	-4.2923	0.0000
r64*	-0.0092	0.0035	-2.6017	0.0093
rho	-0.3660	0.0278	-13.1655	0.0000

+ Each number represent each group defended as Group01: Coca-Cola and Sprite, Group02: Pepsi-Cola and Mountain Dew, Group03: Seven-Up and Dr Pepper, Group04: Lipton Brisk., Group05: A&W and Rite-Cola, and Group06: Sunkist and Canada Dry. For example, c12 corresponds to parameter in quantity equation of group01 w.r.t. group02 price variable in

$$dw_n = \left[C_n - \left(1-\theta_1^o\right)w_n\right]d\ln Q + \sum_{n'=1}^{N}\left[C_{n,n'} - \left(1-\theta_2^o\right)w_n\left(w_{n'} - \delta_{n,n'}\right)\right]d\ln p_{n'}$$

* Coefficients with * mark are derived based on the adding-up and homogeneity conditions.

Table 3.7. Comparison Statistics for Three Specifications

Restrictions on Synthetic parameters	Direct		Inverse		Mixed	
	Wald statistic	p-value	Wald statistic	p-value	Wald statistic	p-value
th1 = 0	1681.2049	0.0000	12993.9780	0.0000	4.6833	0.0305
th2 = 0	1568.3829	0.0000	732.8099	0.0000	12.1754	0.0005
th1 = 1	129.9852	0.0000	21.5216	0.0000	315.5424	0.0000
th2 = 1	974.8223	0.0000	14180.2140	0.0000	628.0337	0.0000
th1 = 0 & th2 = 0	3032.4904	0.0000	13000.9610	0.0000	12.6597	0.0018
th1 = 1 & th2 = 1	1059.2406	0.0000	14640.0880	0.0000	3708.4420	0.0000
th1 = 0 & th2 = 1	2485.3570	0.0000	34603.8330	0.0000	1267.3297	0.0000
th1 = 1 & th2 = 0	1642.1024	0.0000	847.4041	0.0000	967.7887	0.0000
	Log-Likelihood	Paramter Number	Log-Likelihood	Paramter Number	Log-Likelihood	Paramter Number
Synthetic model	1332.2280	23	2698.7700	23	1269.1490	23

Within each of direct, inverse, and mixed demand system, all the nested Rotterdam, LA/AIDS, NBR, and CBS specifications, which assume the fixed restriction on the synthetic parameters, are strongly rejected. This test results imply that none of the four nested models is adequate and the synthetic model is a statistically better specification. In this respect, the same synthetic functional form of the common differential AIDS type dependent variable is used for the comparison across the direct, inverse, and mixed demand system. The estimated log-likelihood values suggest that the inverse demand specification strongly dominates both the direct and the mixed demand specifications and the direct demand specification statistically dominates the mixed demand specifications. Note that this ordering of the statistical dominance is interpreted as the ranking among the competing models rather than the rejection one of the competing models.

The compensated and uncompensated elasticities/flexibilities estimates with their standard errors and corresponding p-values for the direct, inverse, and mixed demand systems are presented in Table 3.8. In the results of direct demand system, the own elasticities are all negative and statistically significant. The expenditure elasticities are close to unity, as expected for the normal goods. The soft drinks are net and gross p-substitutes for each other, given that negative estimates $\varepsilon_{4,5}^{c,D}$, $\varepsilon_{5,4}^{c,D}$, $\varepsilon_{4,5}^{D}$, $\varepsilon_{5,4}^{D}$, and $\varepsilon_{6,4}^{D}$ are insignificant, where $\varepsilon_{n,n'}^{c,D}$ and $\varepsilon_{n,n'}^{D}$ denote the compensated and uncompensated elasticities in the direct demand system. In the results of inverse demand system, the own flexibilities are all negative and statistically significant. The scale flexibilities are close to unity in absolute values, as expected for the normal goods. The soft drinks are gross q-substitutes for each other. Note that the compensated flexibilities in inverse demand system are imperfect measures of the interaction of goods in their satisfaction of wants, since the dominating complementarity $f^{c}_{n,n'} > 0$ does not come from the preference structures but from the adding-up or homogeneity condition $\sum_{n'=1}^{N} f^{c}_{n,n'} = 0$ together with the negativity condition $f^{c}_{n,n'} < 0$ (Barten and Bettendorf, 1989). Note that the magnitudes of the compensated cross flexibilities are relatively small. In the results of mixed demand system, the own elasticities and/or flexibilities are all negative and statistically significant. The expenditure elasticities are close to unity, as expected for the normal goods. The soft drinks are net and gross substitutes each other, given that negative estimate $p^{M}_{5,1}$ is insignificant. The exceptions are $f^{c,M}_{5,6}$, $f^{c,M}_{6,5}$, $f^{M}_{5,6}$, and $f^{M}_{6,5}$, whose magnitudes are relatively small compared to other estimates. Note that the substitutability of the mixed compensated elasticities need not be equivalent to either p-substitutability in terms of the direct system, nor q-substitutability in terms of the inverse system, where the $\partial q_n / \partial p_{n'} > 0$ means p-substitutability in terms of the direct system and the $\partial p_n / \partial q_{n'} < 0$ q- substitutability in terms of the inverse system (Moschini and Vissa, 1993). Note also that the expenditure elasticities for quantity dependent group (group 01-04) measure percentage changes in consumption with respect to one percent increase in total expenditure as in the direct demand system, whereas the expenditure elasticities for price dependent group (group 05-06) measure percentage changes in willingness to pay with respect to one percent increase in total expenditure. On the other hand, the scale flexibilities measure percentage changes in normalized price with respect to one percent increase in the proportionate increase in consumption. For example, for group 05 (A&W and Rite Cola), the percentage increase in consumption with respect to one percent increase in total expenditure is 0.749 estimated in the direct demand system, the percentage increase in willingness to pay with respect to one percent increase in total expenditure is 0.100 estimated in the mixed demand system, and the percentage decrease in normalized price with respect to one percent increase in the proportionate increase in consumption is 1.038 estimated in the inverse demand system.

Table 3.8. Elasticities/Flexibilities Estimates

	Direct Compensated						
	P01	P02	P03	P04	P05	P06	note
Q01	-2.871	1.468	0.596	0.193	0.289	0.313	Group01
	0.149	0.112	0.104	0.040	0.087	0.075	Coke
	0.000	0.000	0.000	0.000	0.001	0.000	Sprite
Q02	1.424	-3.156	0.900	0.090	0.354	0.376	Group02
	0.109	0.134	0.094	0.039	0.091	0.072	Pepsi
	0.000	0.000	0.000	0.020	0.000	0.000	Mt. Dew
Q03	0.841	1.310	-3.075	0.155	0.403	0.354	Group03
	0.147	0.137	0.194	0.049	0.108	0.110	7-up
	0.000	0.000	0.000	0.002	0.000	0.001	Dr Pepper
Q04	1.108	0.533	0.629	-2.189	-0.160	0.068	Group04
	0.228	0.228	0.200	0.266	0.275	0.135	Lipton
	0.000	0.020	0.002	0.000	0.561	0.616	Brisk
Q05	0.714	0.902	0.705	-0.069	-2.731	0.467	Group05
	0.214	0.232	0.189	0.118	0.324	0.203	A&W
	0.001	0.000	0.000	0.561	0.000	0.022	Rite Cola
Q06	0.883	1.097	0.709	0.034	0.535	-3.269	Group06
	0.212	0.211	0.220	0.067	0.233	0.313	Sunkist
	0.000	0.000	0.001	0.616	0.022	0.000	Canada Dry

	Inverse Compensated						
	Q01	Q02	Q03	Q04	Q05	Q06	Group
P01	-0.152	0.045	0.037	0.010	0.019	0.041	Group01
	0.006	0.004	0.004	0.002	0.004	0.004	Coke
	0.000	0.000	0.000	0.000	0.000	0.000	Sprite
P02	0.043	-0.140	0.023	0.011	0.019	0.043	Group02
	0.004	0.006	0.004	0.002	0.004	0.005	Pepsi
	0.000	0.000	0.000	0.000	0.000	0.000	Mt. Dew
P03	0.052	0.034	-0.151	0.011	0.007	0.046	Group03
	0.006	0.006	0.009	0.002	0.006	0.006	7-up
	0.000	0.000	0.000	0.000	0.198	0.000	Dr Pepper
P04	0.055	0.065	0.045	-0.286	0.056	0.064	Group04
	0.011	0.010	0.009	0.018	0.014	0.008	Lipton
	0.000	0.000	0.000	0.000	0.000	0.000	Brisk
P05	0.047	0.048	0.013	0.024	-0.161	0.028	Group05
	0.010	0.010	0.010	0.006	0.014	0.010	A&W
	0.000	0.000	0.198	0.000	0.000	0.006	Rite Cola
P06	0.117	0.126	0.093	0.032	0.032	-0.400	Group06
	0.012	0.013	0.011	0.004	0.012	0.034	Sunkist
	0.000	0.000	0.000	0.000	0.006	0.000	Canada Dry

	Mixed Compensated						
	P01	P02	P03	P04	Q05	Q06	Group
Q01	-4.291	2.459	1.561	0.272	-0.047	-0.066	Group01
	0.257	0.210	0.186	0.095	0.031	0.028	Coke
	0.000	0.000	0.000	0.004	0.134	0.018	Sprite
Q02	2.385	-4.372	1.635	0.352	-0.158	-0.092	Group02
	0.204	0.264	0.208	0.100	0.037	0.030	Pepsi
	0.000	0.000	0.000	0.000	0.000	0.002	Mt. Dew
Q03	2.203	2.380	-4.928	0.345	-0.190	-0.220	Group03
	0.263	0.303	0.394	0.121	0.055	0.048	7-up
	0.000	0.000	0.000	0.004	0.001	0.000	Dr Pepper
Q04	1.555	2.078	1.399	-5.555	-0.347	-0.209	Group04
	0.546	0.589	0.490	0.526	0.118	0.074	Lipton
	0.004	0.000	0.004	0.000	0.003	0.005	Brisk
P05	0.028	0.312	0.270	0.135	-0.196	0.049	Group05
	0.079	0.097	0.100	0.051	0.035	0.017	A&W
	0.725	0.001	0.007	0.008	0.000	0.003	Rite Cola
P06	0.098	0.179	0.378	0.088	0.056	-0.252	Group06
	0.083	0.092	0.096	0.037	0.019	0.032	Sunkist
	0.236	0.052	0.000	0.018	0.003	0.000	Canada Dry

	Direct Uncompensated						
	P01	P02	P03	P04	P05	P06	Expenditure
Q01	-3.135	1.196	0.409	0.147	0.182	0.219	0.973
	0.150	0.111	0.104	0.040	0.087	0.075	0.020
	0.000	0.000	0.000	0.000	0.036	0.004	0.000
Q02	1.126	-3.462	0.689	0.038	0.234	0.271	1.095
	0.109	0.133	0.094	0.039	0.091	0.073	0.018
	0.000	0.000	0.000	0.322	0.011	0.000	0.000
Q03	0.574	1.034	-3.264	0.108	0.295	0.260	0.985
	0.146	0.137	0.193	0.050	0.109	0.110	0.028
	0.000	0.000	0.000	0.029	0.007	0.018	0.000
Q04	0.939	0.359	0.509	-2.219	-0.228	0.008	0.621
	0.228	0.227	0.200	0.266	0.275	0.136	0.035
	0.000	0.114	0.011	0.000	0.407	0.952	0.000
Q05	0.511	0.692	0.561	-0.104	-2.814	0.395	0.749
	0.213	0.231	0.189	0.119	0.325	0.204	0.055
	0.017	0.003	0.003	0.378	0.000	0.053	0.000
Q06	0.532	0.734	0.460	-0.028	0.392	-3.393	1.295
	0.211	0.210	0.220	0.067	0.233	0.314	0.069
	0.012	0.000	0.036	0.677	0.092	0.000	0.000

	Inverse Uncompensated						
	Q01	Q02	Q03	Q04	Q05	Q06	Scale
P01	-0.427	-0.239	-0.159	-0.038	-0.092	-0.056	-1.014
	0.006	0.004	0.004	0.002	0.004	0.004	0.006
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
P02	-0.227	-0.420	-0.168	-0.036	-0.091	-0.053	-0.997
	0.004	0.005	0.004	0.002	0.004	0.005	0.006
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
P03	-0.224	-0.250	-0.347	-0.037	-0.104	-0.051	-1.015
	0.006	0.005	0.009	0.002	0.006	0.006	0.008
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
P04	-0.247	-0.247	-0.169	-0.339	-0.066	-0.043	-1.113
	0.010	0.009	0.009	0.019	0.014	0.008	0.013
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
P05	-0.234	-0.243	-0.187	-0.025	-0.275	-0.072	-1.038
	0.011	0.009	0.010	0.006	0.014	0.010	0.015
	0.000	0.000	0.000	0.000	0.000	0.000	0.000
P06	-0.111	-0.109	-0.069	-0.008	-0.061	-0.481	-0.840
	0.015	0.016	0.013	0.004	0.012	0.035	0.041
	0.000	0.000	0.000	0.057	0.000	0.000	0.000

	Mixed Uncompensated						
	P01	P02	P03	P04	Q05	Q06	Expenditure
Q01	-4.614	2.083	1.267	0.191	-0.028	-0.044	1.136
	0.260	0.208	0.186	0.096	0.031	0.028	0.038
	0.000	0.000	0.000	0.046	0.363	0.110	0.000
Q02	2.002	-4.819	1.287	0.257	-0.136	-0.067	1.348
	0.203	0.263	0.209	0.097	0.036	0.030	0.042
	0.000	0.000	0.000	0.008	0.000	0.025	0.000
Q03	1.843	1.959	-5.256	0.255	-0.169	-0.196	1.269
	0.263	0.302	0.398	0.121	0.055	0.048	0.066
	0.000	0.000	0.000	0.035	0.002	0.000	0.000
Q04	1.357	1.846	1.218	-5.604	-0.336	-0.196	0.699
	0.548	0.578	0.487	0.533	0.118	0.074	0.113
	0.013	0.001	0.012	0.000	0.004	0.008	0.000
P05	-0.001	0.278	0.244	0.127	-0.194	0.051	0.100
	0.079	0.095	0.100	0.051	0.034	0.017	0.028
	0.993	0.004	0.015	0.012	0.000	0.002	0.000
P06	0.023	0.091	0.310	0.069	0.060	-0.247	0.265
	0.085	0.094	0.097	0.037	0.019	0.031	0.045
	0.787	0.331	0.001	0.058	0.002	0.000	0.000

* P and Q denotes representative price and quantity indices for each group defended as Group 01: Coca-Cola and Sprite, Group 02: Pepsi-Cola and Mountain Dew, Group 03: Seven-Up and Dr Pepper, Group 04: Lipton Brisk., Group 05: A&W and Rite-Cola, Group 06: Sunkist and Canada Dry, and E denote total expenditure variable
* In each cell, the first element is the estimates, the second is the standard error, and the third is the associated p-value.

The convenient and familiar forms of comparison are possible across the direct, inverse, and mixed demand systems in terms of one of three possible forms: the elasticities in the form of direct demand system, the flexibilities in the form of inverse demand system, and the elasticities in the form of mixed demand system. These results are retrieved based on the derived relationships among elasticities and/or flexibilities across the direct, inverse, and mixed demand systems. The relationships across the direct, inverse, and mixed demand system in terms of uncompensated elasticities/flexibilities retrieved from the direct, inverse, and mixed demand system are presented in Table 3.9. The tables in diagonal positions are replicated from the estimated ones and the own and expenditure/scale elasticities/flexibilities are summarized in the tables at the bottom positions. The own elasticities and/or flexibilities are all negative and the soft drinks are gross substitutes each other, given that the insignificance estimates imply the insignificant corresponding retrieved ones. For example, the insignificant estimate $\varepsilon_{5,4}^{D}$ in the direct demand system implies the corresponding insignificant retrieved one $p_{5,4}^{M}$ in the mixed demand form retrieved from the direct system estimates. In general, the expenditure elasticities and scale flexibilities are close to unity, as expected for the normal goods. Recall that the expenditure elasticities for the direct demand system and for the quantity dependent variables group in the mixed demand system, the expenditure elasticities for the price dependent variables group in the mixed demand system, and the scale flexibility for the inverse demand system measure different responses of consumers with respect to the changes in different variables as discussed.

The magnitudes of consumers' response measured in three different specifications are different in general and some differences are not trivial. For the group 05 (A&W and Rite Cola) as an example, (a) The percentage increase in consumption with respect to one percent increase in total expenditure measured in the direct, inverse, and mixed demand systems are 0.749, 0.785, and 0.847 represented in the direct demand form. (b) The percentage decrease in normalized price with respect to one percent increase in the proportionate increase in each consumption measured in the direct, inverse, and mixed demand systems are 1.056, 1.038, and 0.818 represented in the inverse demand form. (c) The percentage increase in willingness to pay with respect to one percent increase in total expenditure measured in the direct, inverse, and mixed demand systems are 0.325, 0.194, and 0.100 represented in the mixed demand form. (d) The percentage decrease in consumption with respect to one percent increase in its own price measured in the direct, inverse, and mixed demand systems are 2.814, 5.132, and 5.494

Table 3.9. Elasticities/Flexibilities Comparisons

Direct Form Estimated from Direct Model

	P01	P02	P03	P04	P05	P06	Expenditure
Q01	-3.135	1.196	0.409	0.147	0.182	0.219	0.973
Q02	1.126	-3.462	0.689	0.038	0.234	0.271	1.095
Q03	0.574	1.034	-3.264	0.108	0.295	0.260	0.985
Q04	0.939	0.359	0.509	-2.219	-0.228	0.008	0.621
Q05	0.511	0.692	0.561	-0.104	-2.814	0.395	0.749
Q06	0.532	0.734	0.460	-0.028	0.392	-3.393	1.295

Inverse Form Retrieved from Direct Model

	Q01	Q02	Q03	Q04	Q05	Q06	Scale
P01	-0.460	-0.230	-0.133	-0.037	-0.069	-0.066	-0.995
P02	-0.215	-0.438	-0.146	-0.025	-0.073	-0.069	-0.965
P03	-0.187	-0.220	-0.405	-0.031	-0.080	-0.070	-0.993
P04	-0.254	-0.198	-0.158	-0.477	-0.017	-0.047	-1.151
P05	-0.187	-0.212	-0.152	-0.003	-0.413	-0.089	-1.056
P06	-0.163	-0.184	-0.124	-0.012	-0.085	-0.339	-0.907

Mixed Form Retrieved from Direct Model

	P01	P02	P03	P04	Q05	Q06	Expenditure
Q01	-3.058	1.301	0.484	0.137	-0.075	-0.073	1.124
Q02	1.224	-3.329	0.785	0.026	-0.096	-0.091	1.285
Q03	0.682	1.181	-3.157	0.094	-0.117	-0.090	1.189
Q04	0.894	0.297	0.460	-2.210	0.082	0.007	0.550
P05	0.207	0.281	0.222	-0.039	-0.361	-0.042	0.325
P06	0.181	0.249	0.161	-0.013	-0.042	-0.300	0.419

Direct Form Retrieved from Inverse Model

	P01	P02	P03	P04	P05	P06	Expenditure
Q01	-3.841	1.327	0.731	0.173	0.502	0.135	0.972
Q02	1.261	-4.139	1.086	0.144	0.477	0.106	1.065
Q03	1.029	1.604	-4.684	0.159	0.849	0.061	0.981
Q04	1.080	0.967	0.709	-3.244	-0.174	0.010	0.652
Q05	1.292	1.292	1.523	-0.081	-5.132	0.322	0.785
Q06	0.274	0.224	0.048	-0.029	0.304	-2.185	1.363

Inverse Form Estimated from Inverse Model

	Q01	Q02	Q03	Q04	Q05	Q06	Scale
P01	-0.427	-0.239	-0.159	-0.038	-0.092	-0.056	-1.014
P02	-0.227	-0.420	-0.168	-0.036	-0.091	-0.053	-0.997
P03	-0.224	-0.250	-0.347	-0.037	-0.104	-0.051	-1.015
P04	-0.247	-0.247	-0.169	-0.339	-0.066	-0.043	-1.112
P05	-0.234	-0.243	-0.187	-0.025	-0.275	-0.072	-1.038
P06	-0.111	-0.109	-0.069	-0.008	-0.061	-0.481	-0.840

Mixed Form Retrieved from Inverse Model

	P01	P02	P03	P04	Q05	Q06	Expenditure
Q01	-3.687	1.476	0.891	0.163	-0.102	-0.077	1.157
Q02	1.403	-4.000	1.236	0.134	-0.097	-0.063	1.226
Q03	1.262	1.834	-4.424	0.144	-0.169	-0.053	1.185
Q04	1.036	0.923	0.657	-3.241	0.034	0.001	0.625
P05	0.262	0.260	0.301	-0.017	-0.197	-0.029	0.194
P06	0.162	0.139	0.064	-0.016	-0.027	-0.462	0.651

Direct Form Retrieved from Mixed Model

	P01	P02	P03	P04	P05	P06	Expenditure
Q01	-4.619	2.002	1.146	0.148	0.215	0.223	1.055
Q02	1.993	-5.092	0.946	0.119	0.837	0.442	1.147
Q03	1.820	1.533	-5.867	0.032	1.191	1.036	0.875
Q04	1.330	1.148	0.325	-5.958	2.108	1.224	0.164
Q05	0.022	1.632	1.688	0.778	-5.494	-1.123	0.847
Q06	0.098	0.763	1.662	0.468	-1.332	-4.314	1.277

Inverse Form Retrieved from Mixed Model

	Q01	Q02	Q03	Q04	Q05	Q06	Scale
P01	-0.405	-0.256	-0.168	-0.033	-0.086	-0.075	-1.022
P02	-0.246	-0.393	-0.163	-0.034	-0.100	-0.076	-1.012
P03	-0.245	-0.248	-0.318	-0.034	-0.106	-0.097	-1.048
P04	-0.232	-0.245	-0.164	-0.205	-0.137	-0.099	-1.082
P05	-0.157	-0.201	-0.144	-0.044	-0.265	-0.007	-0.818
P06	-0.124	-0.136	-0.129	-0.029	0.007	-0.293	-0.703

Mixed Form Estimated from Mixed Model

	P01	P02	P03	P04	Q05	Q06	Expenditure
Q01	-4.614	2.083	1.267	0.191	-0.028	-0.044	1.136
Q02	2.002	-4.819	1.287	0.257	-0.136	-0.067	1.348
Q03	1.843	1.959	-5.256	0.255	-0.169	-0.196	1.269
Q04	1.357	1.846	1.218	-5.604	-0.336	-0.196	0.699
P05	-0.001	0.278	0.244	0.127	-0.194	0.051	0.100
P06	0.023	0.091	0.310	0.069	0.060	-0.247	0.265

Comparison of Own/Expenditure Elasticities in Ordinary Form

Own	Direct	Inverse	Mixed	Direct	Inverse	Mixed	Expenditure
Q01	-3.135	-3.841	-4.619	0.973	0.972	1.055	Coke, Sprite
Q02	-3.462	-4.139	-5.092	1.095	1.065	1.147	Pepsi, Mt. Dew
Q03	-3.264	-4.684	-5.867	0.985	0.981	0.875	7-up, Dr Pepper
Q04	-2.219	-3.244	-5.958	0.621	0.652	0.164	Lipton Brisk
Q05	-2.814	-5.132	-5.494	0.749	0.785	0.847	A&W, Rite Cola
Q06	-3.393	-2.185	-4.314	1.295	1.363	1.277	Sunkist,Canada

Comparison of Own/Scale Flexibilities in Inverse Form

Own	Direct	Inverse	Mixed	Direct	Inverse	Mixed	Sclae
P01	-0.460	-0.427	-0.405	-0.995	-1.014	-1.022	Coke, Sprite
P02	-0.438	-0.420	-0.393	-0.965	-0.997	-1.012	Pepsi, Mt. Dew
P03	-0.405	-0.347	-0.318	-0.993	-1.015	-1.048	7-up, Dr Pepper
P04	-0.477	-0.339	-0.205	-1.151	-1.112	-1.082	Lipton Brisk
P05	-0.413	-0.275	-0.265	-1.056	-1.038	-0.818	A&W, Rite Cola
P06	-0.339	-0.481	-0.293	-0.907	-0.840	-0.703	Sunkist,Canada

Comparison of Own/Expenditure Elasticities in Mixed Form

Own	Direct	Inverse	Mixed	Direct	Inverse	Mixed	Expenditure
Q01	-3.058	-3.687	-4.614	1.124	1.157	1.136	Coke, Sprite
Q02	-3.329	-4.000	-4.819	1.285	1.226	1.348	Pepsi, Mt. Dew
Q03	-3.157	-4.424	-5.256	1.189	1.185	1.269	7-up, Dr Pepper
Q04	-2.210	-3.241	-5.604	0.550	0.625	0.699	Lipton Brisk
P05	-0.361	-0.197	-0.194	0.325	0.194	0.100	A&W, Rite Cola
P06	-0.300	-0.462	-0.247	0.419	0.651	0.265	Sunkist,Canada

* P and Q denotes representative price and quantity indices for each group defended as Group01: Coca-Cola and Sprite, Group02: Pepsi-Cola and Mountain Dew, Group03: Seven-Up and Dr Pepper, Group04: Lipton Brisk., Group05: A&W and Rite-Cola, Group06: Sunkist and Canada Dry.

represented in the direct demand form. (e) The percentage decrease in normalized price with respect to one percent increase in its own consumption measured in the direct, inverse, and mixed demand systems are 0.413, 0.275, and 0.265 represented in the inverse demand form. (f) The percentage decrease in willingness to pay with respect to one percent increase in its own consumption measured in the direct, inverse, and mixed demand systems are 0.361, 0.197, and 0.194 represented in the mixed demand form. Recall that these relationships are based on the partitioning quantity-dependent and price-dependent groups of commodities or the legitimate mixed demand system, which is identified by the PC algorithm.

Given the observation that the magnitudes of consumers' response measured in three different specifications are different in general, interpretation of the overall empirical results is not easy. However, one plausible comparison among three different demand systems of direct, inverse, and mixed demand systems is possible based on the model selection approach. Given that all three competing models have the same number of independent parameters (23), all three model selection rules, the Akaike Information, Schwarz information criterion, and the Pollak and Wales' likelihood dominance criterion, are used based on the comparison of the estimated log-likelihood function values, such as the higher log-likelihood value, the higher ranking among competing models. The estimated log-likelihood values suggest that the inverse demand specification (2698.77) strongly dominate both direct and mixed demand specifications and the direct demand specification (1332.23) statistically dominates the mixed demand specifications (1269.15). Additional empirical result that might lead one to prefer the inverse demand system is that the overall standard errors for the flexibility estimates of the inverse demand system are smaller than the overall standard errors for the elasticity estimates of the direct and mixed demand system. For example, the simple average of standard errors for the inverse, direct, and mixed uncompensated flexibility/elasticity estimates are 0.009, 0.159, and 0.164 respectively. These empirical evidences are consistent with the local causal structure inductively inferred by the GES algorithm. It can be also argued that the information inferred by the PC algorithm is also useful, given the observations that (i) The comparisons of three different specifications in three different forms are possible due to the reasonable partitioning of quantity-dependent and price-dependent groups of commodities or legitimate mixed demand system, which is identified by the PC algorithm. (ii) The magnitudes of consumers' response measured in three different specifications do not deviate too far with each other and thus provide plausible bounds in all the three different forms, although they are different in general and some differences are not trivial.

In this respect, another possible approach to interpret the overall empirical results is to pursue the model averaging method rather than model selection method taken in this study, given that the model selection ordering of the statistical dominance need to be interpreted as the ranking among the competing models, rather than the rejection one of the competing models and accepting the other. The overall results imply that the graphical causal model method can provide reliable and helpful guidelines for the local identification issue of the choice among the direct, inverse, and mixed demand systems.

Summary and Discussion

The proposed methodological procedure to address three methodological issues in the study of consumer behavior is illustrated by using retail checkout scanner data of soft drinks products. The three methodological issues are the aggregation, causal identification, and functional form issues. For the aggregation issue to incorporate broad information into empirical model, the compositional stability condition is used. The legitimate classification is inductively identified among soft drinks products and the empirical evidence with comparison of Lewbel's consistent aggregation condition is provided. The following six groups are used for subsequent analyses: Coca-Cola and Sprite product group, Pepsi-Cola and Mountain Dew product group, Seven-Up and Dr Pepper product group, Lipton Brisk product group, A&W and Rite-Cola product group, and Sunkist and Canada Dry product group. For the local (causal) identification issue between price and quantity variables or the model specification issue among three possible specifications of the direct, inverse, and mixed demand systems, the graphical causal model and model selection methods are used. To connect these two methods with minimizing the effect of parameterization assumptions, the AIDS type dependent variable form synthetic models are use for all the three demand systems of the direct, inverse, and mixed demand systems. The GES algorithm result implies the inverse demand specification, whereas the PC algorithm result suggests the mixed demand system. Based on these inductively inferred local causal structures between price and quantity variables of a particular product, the inverse and mixed demand systems are estimated as well as the direct demand system for comparison purpose. In all three demand systems, four nested parameterizations of Rotterdam, LA/AIDS, NBR, and CBS are statistically rejected and thus the synthetic differential functional forms are used for three demand systems. Based on the classification of the price dependent variable group (the A&W and Rite-Cola and the Sunkist and Canada Dry product groups) and the quantity dependent variable group (all other three groups) in the mixed demand system, which is identified by the PC algorithm, the estimated elasticities and flexibilities of three specifications are compared in the direct, inverse, and mixed demand system forms. Based on the model selection approach of the Akaike Information, Schwarz information criterion, and the Pollak and Wales' likelihood dominance criterion, the competing three demand systems are compared. Statistical evidences imply that the data support the inverse demand system, which is identified by the GES algorithm. Overall empirical evidences suggest that the graphical causal model provide helpful and reliable information for the identification issues in the study of consumer behavior.

As future research directions, several methodological issues to be studied can be suggested. A first issue is how to fully use the overall empirical findings. The model averaging approach, rather than model selection approach used in this study, can provide more precise understanding of consumer behavior. One possible approach for the model averaging method is to use the relative log-likelihood values of the direct, inverse, and mixed demand systems. The main issue is how to decide relative weights among competing models. A second issue is how to fully use the causal information inferred by the graphical causal models. Although only the local causal structure between the price and quantity variables are used in this study, other causal information can provide the possibility of a more full understanding of the interactions in the market, which in turn allow a more precise measurements of consumer behavior. The main issue is how to combine the full causal information into the theoretical properties of demand functions with maintaining flexibility and estimable functional form specification. A third issue is how to decide the boundary of the variables included in the empirical models. For example, there can be latent causal structures or interactions with other (size) commodities, although the size of 6/12 oz is used to decide which commodities are included in the study. The causal structure identified by the PC algorithm suggests that there may be latent causal variables among the price variables. The main issue is how to satisfy or how to relax the causal sufficiency conditions in the analysis, especially in the GES algorithm with discriminating the possible cyclic phenomenon. A fourth issue is how to incorporate the possible dynamic interactions and non-linearity in consumer behavior. Although the differential functional form approach provides useful framework to deal with the possible non-stationarity of variables, incorporating the possible lagged interaction and structural change in consumer behavior can provide more precise understanding of consumer behavior. The main issue is how to capture the possible dynamic interactions and non-linearity phenomena without sacrificing the theoretical properties of demand functions with maintaining flexibility and estimable functional form specification. A fifth issue is how to study consumer behavior at the original disaggregate level beyond the aggregated level used in this study, given that close co-movement among variables implies that the (probabilistic) stability condition is violated and multicollinearity problem is severe. One possible way is to use the mixed estimator. The main issue is how to combine aggregate level information into the mixed estimator to study disaggregate level. Although there remain other methodological issues to be addressed in empirical study, this study provides one plausible inductive procedure for the understanding of consumer behavior, while minimizing the deductive properties or ambiguities. The remaining subjectivities in our proposed method are left as further research topics, with the hope that the remaining subjectivities bring fewer ambiguities relative to the previously used methods.

CHAPTER IV
USE OF HIGH DIMENSIONAL PANEL DATA IN MACRO-ECONOMETRICS

Understanding how monetary policy affects overall economic activity has been the primary topic for theoretical and empirical studies in macro-economics for a long time. In this respect, the macro-econometrics has significantly advanced from methodological and empirical perspectives. In addition, recently available high dimensional macro-economic panel data has brought forth potential for significant advances in the macro-econometric study of monetary policy effect. However, there remain two methodological issues for the full realization of the research potential brought by these advances. One is the identification problem of how to infer the underlying causal structure from the data, given that the causal structure is generally underdetermined by the statistical properties of the data (induction problem) and theory does not provide sufficient causal information. While there have been many approaches to study the monetary policy transmission mechanism, the structural vector autoregressive (VAR) framework is widely used since Sims (1980) introduced the VAR approach as an alternative to structural equation model (SEM) approach. Although the structural VAR framework provides the possibility of inferring causal information from statistical properties of the data without pretending to have too much *a priori* theory and/or without demanding too much information from the data, how to inductively infer the causal structure to relate empirical regularities captured in reduced form model to theoretical properties represented by the structural form model remains an open methodological issue. The other methodological issue to be addressed is how to incorporate an available large information set into an empirical model, given that econometric considerations such as degrees-of-freedom and multicollinearity require the economy of parameters in empirical models. This information problem is important, since misspecification problems can exist due to the small information set incorporated in empirical macro-econometric model, given the observation that monetary authorities monitor a large number of economic variables and there can be many possible channels through which the monetary policy affects the economy.

In this chapter we propose inductive empirical methods to address these two methodological issues in the study of monetary policy effects based on the discussions on the causality and aggregation issues chapter II. A method to infer the causal structures for the study of the monetary policy transmission mechanism and a method to incorporate a broad range of information into the empirical macro-model are main issues to be addressed. More specifically, first, the SEM and VAR approaches are compared in terms of the identification problem. The relative advantage of the VAR approach beyond the recursive Wold causal chain system and the possibility of an inductive inference of the causal structures are discussed. Second, the possible misspecification problems due to the small information set incorporated in standard VAR approach is discussed in the context of the monetary transmission mechanism literature. The possibility both to incorporate high dimensional macro-economic panel data into a standard VAR model and to infer a structural interpretation for this large information set is discussed based on the factor augmented vector autoregressive (FAVAR) framework and the compositional stability conditions. Third, an identification issue in the FAVAR model is addressed by using inductively inferred causal information based on the graphical causal models. The proposed methods are illustrated with the applications for the study of the monetary policy effects using macro-economic panel data.

Theoretical Considerations

Causality in Study of Monetary Policy Effect

Empirical studies in economics have been developed along two distinctive interpretations of the relative roles of deduction and induction. One approach emphasizes deduction and interprets econometrics as an instrument of empirical application of economic theory. The other approach emphasizes induction and interprets statistical method as an instrument for the empirical discovery of economic relationships. While the first interpretation leads to empirical studies which aim to measure the strength of causal relationships deductively derived from *a priori* theory, the second interpretation leads to empirical studies which aim to inductively infer the causal structure itself with a minimum of *a priori* restrictions. The extreme arguments of these two approaches sometimes even bring the tension between economists who devoted to develop formalized theory without measurement and those devoted to develop measurement without theory. Macro-econometrics is an area where this kind of tension has been clearly observed. Given that identifying a system of equations means determining the causal structure, the different interpretations of the relative roles of deduction and induction in inferring the causal or structural information from the observationally equivalent statistical properties of data or the reduced form information is the main issue in the debate between the Cowles Commission and the National Bureau of Economic Research (Koopmans, 1949). “The development of methods for causal inference in macro-econometrics has been fragile with a tension between a deductive approach and an inductive approach. The first conceives of causes as something that economic theory must provide and that statistical method must measure. The second considers economic theory a not very reliable source of causal knowledge and opens the possibility of inferring causes form statistical properties of the data without pretending to have too much *a priori* theory. The first conception was advocated by some exponents of the Cowles Commission during 1950s and is fashionable among the calibration approach to econometrics. The second conception was formalized by Granger's (1969) test of causality and by Sims' (1980) vector autoregressive models, methods which are still very popular in nowadays econometrics (Moneta, 2007)."

In general, the first deductive approach is incorporated in the structural equation model (SEM) framework, whereas the second inductive approach is incorporated in the vector autoregressive (VAR) framework. Two distinctive econometric approaches can be summarized and compared in the context of the required causal information for identification. The structural equation model (SEM) for M endogenous variables Y and K predetermined variables X can be written as follows, where predetermined variables means exogenous, lagged exogenous, and lagged endogenous variables.

· The structural form SEM:

$Y_t^T \alpha + X_t^T \beta = \varepsilon_t^T$ or $Y_t^T = -X_t^T \beta \alpha^{-1} + \varepsilon_t^T \alpha^{-1}$, where $Cov(\varepsilon_t) = \Sigma$

· The reduced form SEM:

$Y_t^T = X_t^T \Pi + u_t^T$ where $\Pi = -\beta\alpha^{-1}$ and $u_t^T = \varepsilon_t^T \alpha^{-1}$, thus $Cov(u_t) = \hat{\Sigma} = (\alpha^{-1})^T \Sigma (\alpha^{-1})$

The observational equivalence or under-identification in SEM can be intuitively understood by simply counting parameters in the structure and reduced forms. Since the structural form has $M \times M$ parameters in coefficient matrix α, $K \times M$ parameters in coefficient matrix β, and $M \cdot (M+1)/2$ parameters in covariance matrix Σ and the reduced form has $K \times M$ parameters in coefficient matrix Π, and $M \cdot (M+1)/2$ parameters in covariance matrix $\hat{\Sigma}$, SEM has M^2 excessive number of parameters to be specified. When the normalization such that one endogenous variable in each equation has a coefficient of one are used (M restrictions), there remain $M(M-1)$ undetermined excessive parameters. When the additional assumption that $Cov(\varepsilon_t) = \Sigma$ is a diagonal matrix is also introduced ($M \cdot (M-1)/2$ restrictions), there remain $M \cdot (M-1)/2$ undetermined excessive parameters which should be resolved through non-sample information. This implies that many different structural models, which correspond to different $M \cdot (M-1)/2$ restrictions, can have the same reduced form.

The approach to this identification problem in SEM framework can be understood as follows. The main issue is how to specify the undetermined excessive parameters in α and β. Mathematically the reduced form SEM can be transformed into the structural form SEM for a single jth equation with following matrix partition. The M endogenous variable matrix Y^T is partitioned into the normalized jth endogenous variable y_j with a coefficient of one, M_j endogenous variables Y_j^T included in the jth equation, and M_j^* endogenous variables Y_j^{*T} excluded from the jth equation. The K exogenous variable matrix X^T is partitioned into K_j exogenous variables X_j^T included in the jth equation and K_j^* exogenous variables X_j^{*T} excluded from the jth equation.

· The general reduced form SEM for the jth equation with suitable matrix partition:

$$\begin{bmatrix} y_j & Y_j^T & Y_j^{*T} \end{bmatrix} = \begin{bmatrix} X_j^T & X_j^{*T} \end{bmatrix} \cdot \begin{bmatrix} \pi_j & \underline{\Pi}_j & \overline{\Pi}_j \\ \pi_j^* & \underline{\Pi}_j^* & \overline{\Pi}_j^* \end{bmatrix} + \begin{bmatrix} u_j & U_j & U_j^* \end{bmatrix}$$

· The corresponding structural form SEM for the jth equation with normalization:

$$\begin{bmatrix} y_j & Y_j^T & Y_j^{*T} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -\alpha_j \\ -\alpha_j^* \end{bmatrix} = \begin{bmatrix} X_j^T & X_j^{*T} \end{bmatrix} \cdot \begin{bmatrix} \pi_j & \underline{\Pi}_j & \overline{\Pi}_j \\ \pi_j^* & \underline{\Pi}_j^* & \overline{\Pi}_j^* \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -\alpha_j \\ -\alpha_j^* \end{bmatrix} + \begin{bmatrix} u_j & U_j & U_j^* \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -\alpha_j \\ -\alpha_j^* \end{bmatrix} \text{ or}$$

$$y_j - Y_j^T \alpha_j - Y_j^{*T} \alpha_j^* = \begin{bmatrix} X_j^T & X_j^{*T} \end{bmatrix} \cdot \begin{bmatrix} \pi_j - \underline{\Pi}_j \alpha_j - \overline{\Pi}_j \alpha_j^* \\ \pi_j^* - \underline{\Pi}_j^* \alpha_j - \overline{\Pi}_j^* \alpha_j^* \end{bmatrix} + v_j = \begin{bmatrix} X_j^T & X_j^{*T} \end{bmatrix} \cdot \begin{bmatrix} \beta_j \\ \beta_j^* \end{bmatrix} + v_j \ .$$

· The specific reduced form SEM with exclusion assumptions of $\alpha_j^* = 0$ and $\beta_j^* = 0$:

$$y_j = Y_j^T \alpha_j + X_j^T \beta_j + v_j = \begin{bmatrix} Y_j^T & X_j^T \end{bmatrix} \cdot \begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} + v_j = W_j^T \cdot \delta_j + v_j$$

The exclusion assumptions of $\alpha_j^* = 0$ and $\beta_j^* = 0$ transform the general reduced form into the specific reduced form, which can be used for system estimation by two-stage or three-stage least square methods. The required exclusion assumptions of $\alpha_j^* = 0$ and $\beta_j^* = 0$ implies that $\pi_j - \underline{\Pi}_j \alpha_j = \beta_j$ and $\pi_j^* - \underline{\Pi}_j^* \alpha_j = 0$. Since the system of equations $\pi_j^* = \underline{\Pi}_j^* \alpha_j$ is K_j^* equations in

M_j unknowns, the solution of α_j exists if there were at least as many equations as unknowns (order condition) and is unique if $rank\left\lfloor \pi_j^* \ \underline{\Pi_j^*} \right\rfloor = rank\left\lfloor \underline{\Pi_j^*} \right\rfloor = M_j$ (rank condition). Intuitively the order condition ($K_j^* \geq M_j$) can be understood as the condition that the number of exogenous variables excluded from a single j th equation must be at least as large as the number of endogenous variables included in a single j th equation. With the rank condition, the algebraic identification conditions through the exclusion assumptions in both $\alpha_j^* = 0$ and $\beta_j^* = 0$ can be understood as the condition that the simultaneous relationships of the j th endogenous variable and other endogenous variables included in the j th equation are discriminated by the exogenous variables, which are not in the j th equation but in other equations for endogenous variables included in the j th equation, as the specific shifters or additional causal determinants. For example, the demand (supply) shifters allow identifying supply (demand) equation. In this respect, the SEM approach to the identification problem can be understood as one that looks for additional causal determinants that discriminate among simultaneous relationships.

The vector autoregressive (VAR) approach can be understood as follows. Note that initially the VAR approach is proposed to pursue the absolutely inductive method without using any deductive structural information (at least in the estimation step) and aims to study how various shocks would affect the variables of the system, minimizing the structural concept itself. Such objective, however, faces a difficult issue that the residual terms in a reduced form VAR are not in general independent, so that a shock to one becomes a shock to others depending on the correlation structure among them. Orthogonalization takes into account this co-movement of the residual terms in the reduced form VAR and makes it possible to interpret the innovations in structural form VAR as fundamental structural shocks. Henceforth the statistical properties of economic time series are summarized by the reduced form VAR and the causal structures are imposed in the structural form VAR based on either the theoretical implications or institutional knowledge.

· The structural form VAR for $N \times 1$ vector of variables Z_t:

$A_0 Z_t - A_1 Z_{t-1} - \cdots - A_P Z_{t-P} = \varepsilon_t$ or $Z_t = A_0^{-1} A_1 Z_{t-1} + \cdots + A_0^{-1} A_P Z_{t-P} + A_0^{-1} \varepsilon_t$, where $Cov(\varepsilon_t) = \Omega$

· The reduced form VAR:

$Z_t = B_1 Z_{t-1} + \cdots + B_P Z_{t-P} + u_t$ where $B_p = A_0^{-1} A_p$ and $u_t = A_0^{-1} \varepsilon_t$, thus $Cov(u_t) = \hat{\Omega} = A_0^{-1} \Omega \left(A_0^{-1}\right)^T$

· Derivation of vector moving average (VMA) form or impulse response function (IRF) by either solving analytically or recursively backwards using lag operator L:

$$Z_t = B_1 \cdot LZ_t + \cdots + B_P \cdot L^P Z_t + u_t \qquad Z_t = B_1 Z_{t-1} + \cdots + B_P Z_{t-P} + u_t$$

$$\left(I - B_1 L - \cdots - B_P L^P\right) \cdot Z_t = u_t \quad \text{or} \quad = B_1\left(B_1 Z_{t-2} + \cdots + B_{P-1} Z_{t-P-1} + u_{t-1}\right) + \cdots + B_P Z_{t-P} + u_t$$

$$Z_t = \left(I - B_1 L - \cdots - B_P L^P\right)^{-1} u_t \qquad = \left(B_1 B_1 Z_{t-2} + \cdots + B_1 B_{P-1} Z_{t-P-1}\right) + \cdots + B_P Z_{t-P} + \left(u_t + B_1 u_{t-1}\right)$$

$$Z_t = u_t + C_1 u_{t-1} + \cdots + C_\infty u_{t-\infty} \qquad \text{or } Z_t = A_0^{-1}\varepsilon_t + C_1 A_0^{-1}\varepsilon_{t-1} + \cdots + C_\infty A_0^{-1}\varepsilon_{t-\infty}.$$

The observational equivalence or under-identification in VAR framework can be intuitively understood by simply counting parameters in the structure and reduced forms. Since the structural form has N^2 parameters in coefficient matrix A_0, $P \times N^2$ parameters in the sequence of coefficient matrix $\{A_1, \cdots, A_P\}$, and $N \cdot (N+1)/2$ parameters in covariance matrix $Cov(\varepsilon_t) = \Omega$ and the reduced form has $P \cdot N^2$ parameters in the sequence of coefficient matrix $\{B_1, \cdots, B_P\}$ and $N \cdot (N+1)/2$ parameters in covariance matrix $Cov(u_t) = \hat{\Omega}$. The VAR approach has N^2 excessive number of parameters. When the normalization such that one endogenous variable in each equation has a coefficient of one are used (N restrictions), there remain $N(N-1)$ undetermined excessive parameters. When the additional assumption that $Cov(\varepsilon_t) = \Omega$ is diagonal matrix is also introduced ($N \cdot (N-1)/2$ restrictions), there remain $N \cdot (N-1)/2$ undetermined excessive parameters which should be resolved through non-sample information. This implies that many different structural models, which correspond to different $N \cdot (N-1)/2$ restrictions, can have the same reduced form.

The approach to this identification problem in VAR framework can be understood based on the following simple two-variable VAR example. The main issue is how to specify A_0 coefficient matrix, which relates the structural and reduced form VAR specifications and controls how the endogenous variables are linked to each other contemporaneously.

· The structural form VAR with normalization on diagonal elements in A_0:

$$\begin{bmatrix} 1 & a_0^{12} \\ a_0^{21} & 1 \end{bmatrix} \cdot \begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} a_p^{11} & a_p^{12} \\ a_p^{21} & a_p^{22} \end{bmatrix} \cdot \begin{bmatrix} Z_{t-p}^1 \\ Z_{t-p}^2 \end{bmatrix} + \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix} \text{ where } Cov\left(\begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix}\right) = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{11}^2 & 0 \\ 0 & \sigma_{22}^2 \end{bmatrix} = D$$

· Derivation of structural form VAR with normalization on diagonal elements in $Cov(\varepsilon_t) = D$

$$\begin{bmatrix} \sigma_{11}^{-1} & 0 \\ 0 & \sigma_{22}^{-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & a_0^{12} \\ a_0^{21} & 1 \end{bmatrix} \cdot \begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} \sigma_{11}^{-1} & 0 \\ 0 & \sigma_{22}^{-1} \end{bmatrix} \cdot \begin{bmatrix} a_p^{11} & a_p^{12} \\ a_p^{21} & a_p^{22} \end{bmatrix} \cdot \begin{bmatrix} Z_{t-p}^1 \\ Z_{t-p}^2 \end{bmatrix} + \begin{bmatrix} \sigma_{11}^{-1} & 0 \\ 0 & \sigma_{22}^{-1} \end{bmatrix} \cdot \begin{bmatrix} e_t^1 \\ e_t^2 \end{bmatrix}$$

$$\begin{bmatrix} \sigma_{11}^{-1} & a_0^{12}\sigma_{11}^{-1} \\ a_0^{21}\sigma_{22}^{-1} & \sigma_{22}^{-1} \end{bmatrix} \cdot \begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} a_p^{11}\sigma_{11}^{-1} & a_p^{12}\sigma_{11}^{-1} \\ a_p^{21}\sigma_{22}^{-1} & a_p^{22}\sigma_{22}^{-1} \end{bmatrix} \cdot \begin{bmatrix} Z_{t-p}^1 \\ Z_{t-p}^2 \end{bmatrix} + \begin{bmatrix} e_t^1\sigma_{11}^{-1} \\ e_t^2\sigma_{22}^{-1} \end{bmatrix}$$

$$\begin{bmatrix} A_0^{11} & A_0^{12} \\ A_0^{21} & A_0^{22} \end{bmatrix} \cdot \begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} A_p^{11} & A_p^{12} \\ A_p^{21} & A_p^{22} \end{bmatrix} \cdot \begin{bmatrix} Z_{t-p}^1 \\ Z_{t-p}^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \text{ where } Cov\left(\begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

· Derivation of the reduced form VAR with normalization of $Cov(\varepsilon_t) = D = I$

$$\begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} A_0^{11} & A_0^{12} \\ A_0^{21} & A_0^{22} \end{bmatrix}^{-1} \begin{bmatrix} A_p^{11} & A_p^{12} \\ A_p^{21} & A_p^{22} \end{bmatrix} \cdot \begin{bmatrix} Z_{t-p}^1 \\ Z_{t-p}^2 \end{bmatrix} + \begin{bmatrix} A_0^{11} & A_0^{12} \\ A_0^{21} & A_0^{22} \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix}$$

$$\begin{bmatrix} Z_t^1 \\ Z_t^2 \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} B_p^{11} & B_p^{12} \\ B_p^{21} & B_p^{22} \end{bmatrix} \cdot \begin{bmatrix} Z_{t-p}^1 \\ Z_{t-p}^2 \end{bmatrix} + \begin{bmatrix} u_t^1 \\ u_t^2 \end{bmatrix},$$

$$\text{where } Cov\left(\begin{bmatrix} u_t^1 \\ u_t^2 \end{bmatrix} \right) \equiv \begin{bmatrix} \hat{\sigma}_{11}^2 & \hat{\sigma}_{12}^2 \\ \hat{\sigma}_{21}^2 & \hat{\sigma}_{22}^2 \end{bmatrix} = \hat{\Omega} = \begin{bmatrix} A_0^{11} & A_0^{12} \\ A_0^{21} & A_0^{22} \end{bmatrix}^{-1} \left(\begin{bmatrix} A_0^{11} & A_0^{12} \\ A_0^{21} & A_0^{22} \end{bmatrix}^{-1} \right)^T .$$

First, it can be assumed that $Cov(\varepsilon_t) = \Omega$ is diagonal matrix $Cov(\varepsilon_t) = \Omega = D$, which can be justified based on the argument that the innovations in structural form VAR are to be independent with each other, so that they can be interpreted as the fundamental structural shocks. Second, for recovering structural parameters from the estimated reduced form parameters, it is convenient to transform the normalization on diagonal elements in A_0 into the normalization on diagonal elements in $Cov(\varepsilon_t) = \Omega = D$ such that (a) $Cov(\varepsilon_t) = D = I$, (b) The absolute value of diagonal elements in A_0 are the inverse of the standard deviations of the structural shocks, and (c) The impulse responses with respect to structural innovation equal to its unity shock is equivalent to the impulse response with respect to structural innovation equal to its standard deviation shock. Third, since the reduced form VAR system is a system of seemingly unrelated regressions with usually the same regressor in each equation, applying the ordinary least squares method on each equation is equivalent with applying the generalized least square method or the maximum likelihood method with the assumption of normal distribution. The covariance of the estimated reduced form VAR $Cov(u_t) = \hat{\Omega}$ can be obtained with the appropriate choice of lag length, which allows assuming that structural and reduced form innovations are white noise. Fourth, based on the system of equations $\hat{\Omega} = A_0^{-1}(A_0^{-1})^T$, the unknown elements in A_0 coefficient matrix can be solved or recovered in terms of the estimated elements of $Cov(u_t) = \hat{\Omega}$ covariance matrix. Given that there exists the solution for the system of equations if there were at least as many equations as unknowns, $N \cdot (N-1)/2$ restrictions in A_0 need to be imposed for the existence of the solution of A_0 in $\hat{\Omega} = A_0^{-1}(A_0^{-1})^T$, since there are $N \cdot (N+1)/2$ equations in $Cov(u_t) = \hat{\Omega}$ and N^2 unknown parameters in A_0. Note that restrictions on the dynamic structure in the sequence of coefficient matrices $\{A_1, \cdots, A_P\}$ are not required for the identification.

There exists certain degree of corresponding relationship between SEM and VAR approaches in some special circumstance, although the VAR approach is proposed as an alternative to the SEM approach. When it is assumed (endogenous) variables have the special causal structure of the Wold causal chain or recursive system, where the first variable causes second variable and first and second variables cause third variable and so on, the assumed recursive causal structure among (endogenous) variables provides enough restrictions for the identification problem through the triangular restrictions on A_0 in VAR approach ($N \cdot (N-1)/2$ restrictions) and α in SEM approach ($M \cdot (M-1)/2$ restrictions) with the conformable diagonal covariance matrix. In this case, the SEM approach to the identification problem depends only on restrictions on α without requiring restrictions on β, as the VAR approach to the identification problem does not require restrictions on the sequence of coefficient matrix $\{A_1, \cdots, A_P\}$. In fact, when the fully recursive causal model is assumed among (endogenous) variables, the SEM and VAR approaches to identification problem become almost equivalent beside the required block recursive assumption for discriminating endogenous and predetermined variables. This relationship between SEM and VAR approaches can be understood as follows.

· Derivation of the SEM framework from the VAR model in structural form:

$A_0 Z_t - A_1 Z_{t-1} - \cdots - A_P Z_{t-P} = \varepsilon_t$, by assumption of $A_p = \begin{bmatrix} \Gamma_p & 0 \\ \beta_p & \alpha_p \end{bmatrix}$ for $Z_t = \begin{bmatrix} X_t' \\ Y_t' \end{bmatrix}$

$$\begin{bmatrix} \Gamma_0 & 0 \\ \beta_0 & \alpha_0 \end{bmatrix} \cdot \begin{bmatrix} X_t' \\ Y_t' \end{bmatrix} - \begin{bmatrix} \Gamma_1 & 0 \\ \beta_1 & \alpha_1 \end{bmatrix} \cdot \begin{bmatrix} X_{t-1}' \\ Y_{t-1}' \end{bmatrix} - \cdots - \begin{bmatrix} \Gamma_P & 0 \\ \beta_P & \alpha_P \end{bmatrix} \cdot \begin{bmatrix} X_{t-P}' \\ Y_{t-P}' \end{bmatrix} = \begin{bmatrix} \varepsilon_t^{X\prime} \\ \varepsilon_t^{Y\prime} \end{bmatrix}$$

$$\Gamma_0 X_t' - \Gamma_1 X_{t-1}' - \cdots - \Gamma_P X_{t-P}' = \varepsilon_t^{X\prime}$$
$$\alpha_0 Y_t' - \alpha_1 Y_{t-1}' - \cdots - \alpha_P Y_{t-P}' + \beta_0 X_t' - \beta_1 X_{t-1}' - \cdots - \beta_P X_{t-P}' = \varepsilon_t^{Y\prime}$$

The first equation become $X_t' = \varepsilon_t^{X\prime}$ by the assumptions of $\Gamma_0 = I$ and $\Gamma_1 = \cdots = \Gamma_L = 0$. On the other hand, the second equation can be written as the structural form of SEM, $Y_t^T \alpha + X_t^T \beta = \varepsilon_t^T$, where predetermined variables denote exogenous, lagged exogenous, and lagged endogenous variables. Note that the distinction among endogenous and predetermined variables is incorporated by block triangular restrictions on A_0 as well as the sequence of coefficient matrix $\{A_1, \cdots, A_p\}$ with the conformable block diagonal covariance matrix. Note also that although the SEM is usually formulated so that every parameter has an economic interpretation in the structural form of SEM ($Y_t^T \alpha + X_t^T \beta = \varepsilon_t^T$), based on the same logic of expressing VAR(2) $Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + \varepsilon_t$ as the canonical form of VAR(1) $\begin{bmatrix} z_t \\ z_{t-1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \cdot \begin{bmatrix} z_{t-1} \\ z_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$ or $Z_t^* = A_1^* Z_{t-1}^* + \varepsilon_t^*$, the dynamic form of SEM ($\alpha_0 Y_t' - \sum_{p=1}^{P} \alpha_p Y_{t-p}' + \beta_0 X_t' - \sum_{p=1}^{P} \beta_p X_{t-p}' = \varepsilon_t^{Y}{}'$) can be written as $Y_t^T \alpha + X_t^T \beta + Y_{t-1}^T \Phi = \varepsilon_t^T$, which can be transformed into the final form of a dynamic SEM as follows.

· The structural form of a dynamic SEM:

$Y_t^T \alpha + X_t^T \beta + Y_{t-1}^T \Phi = \varepsilon_t^T$ or $Y_t^T = -X_t^T \beta \alpha^{-1} - Y_{t-1}^T \Phi \alpha^{-1} + \varepsilon_t^T \alpha^{-1}$

· The reduced form of a dynamic SEM:

$Y_t^T = X_t^T \Pi + Y_{t-1}^T \Delta + u_t^T$, where $\Pi = -\beta \alpha^{-1}$, $\Delta = -\Phi \alpha^{-1}$, and $u_t^T = \varepsilon_t^T \alpha^{-1}$

· Derivation of the final form of a dynamic SEM by solving recursively backwards:

$$\begin{aligned} Y_t^T &= X_t^T \Pi + Y_{t-1}^T \Delta + u_t^T \\ &= X_t^T \Pi + \left(X_{t-1}^T \Pi + Y_{t-2}^T \Delta + u_{t-1}^T \right) \Delta + u_t^T \\ &= \left(X_t^T \Pi + X_{t-1}^T \Pi \Delta \right) + Y_{t-2}^T \Delta^2 + \left(u_{t-1}^T \Delta + u_t^T \right) \end{aligned}$$

$Y_t^T = \sum_{s=0}^{t-1} X_{t-s}^T \Pi \Delta^s + Y_0^T \Delta^t + \sum_{s=0}^{t-1} u_{t-s}^T \Delta^s$ or $Y_t^T = \sum_{s=0}^{\infty} X_{t-s}^T \Pi \Delta^s + \sum_{s=0}^{\infty} u_{t-s}^T \Delta^s$ by assumption of $\lim_{t \to \infty} \Delta^t = 0$

$\overline{Y}_t^T = \sum_{s=0}^{\infty} \overline{X}^T \Pi \Delta^s = \overline{X}^T \Pi \sum_{s=0}^{\infty} \Delta^s = \overline{X}^T \Pi (I - \Delta)^{-1}$ by using $\sum_{s=0}^{\infty} \Delta^s = (I - \Delta)^{-1}$,

where $\frac{\partial y_{t,m}}{\partial x_{t,k}} = [\Pi]_{km}$, $\frac{\partial y_{t,m}}{\partial x_{t-s,k}} = [\Pi \Delta^s]_{km}$, and $\frac{\partial \overline{y}_m}{\partial \overline{x}_k} = \left[\Pi (I - \Delta)^{-1} \right]_{km}$ are named as the impact multipliers, the dynamic multipliers, and the long-run or equilibrium multipliers respectively (Green, 2000). Note that the final form of a dynamic SEM can be interpreted as the analogue correspondence to the vector moving average (VMA) form or impulse response function (IRF) in VAR approach, except the conceptual difference between altering an entire time path of exogenous variable and giving a single shock to (exogenous) innovations. In this respect, it can be argued that there exists a correspondence relationship between SEM and VAR approaches, when the full recursive Wold causal chain structure is assumed in addition to the required block recursive assumption for discriminating endogenous and predetermined variables.

Even the restrictions of $\Gamma_0 = I$ and $\Gamma_1 = \cdots = \Gamma_P = 0$ can be relaxed to condition of lower triangular matrix of Γ_0 so that the entire coefficient matrix A_0 becomes lower triangular, since it is demonstrated that the recursive Wold causal chain structure in the VAR approach can provide partial identification for a certain specific purpose of study, i.e. the understanding of (monetary) policy effects. This argument can be understood as follows with the assumption that all variables Z_t can be partitioned into variables X_t and variables Y_t, where X_t come before the policy variable S_t and Y_t come after the policy variable S_t in causal order.

· The structural form VAR:

$A_0 Z_t - A_1 Z_{t-1} - \cdots - A_p Z_{t-P} = \varepsilon_t$, $\forall\ p = 1, 2, \cdots, P$,

by assumption of $A_0 = \begin{bmatrix} A_0^{11} & 0 & 0 \\ A_0^{21} & a_0^{22} & 0 \\ A_0^{31} & a_0^{32} & A_0^{33} \end{bmatrix}$ and $A_p = \begin{bmatrix} A_p^{11} & a_p^{12} & A_p^{13} \\ A_p^{21} & a_p^{22} & A_p^{23} \\ A_p^{31} & a_p^{32} & A_p^{33} \end{bmatrix}$ for $Z_t = \begin{bmatrix} X_t \\ S_t \\ Y_t \end{bmatrix}$

$$\begin{bmatrix} A_0^{11} & 0 & 0 \\ A_0^{21} & a_0^{22} & 0 \\ A_0^{31} & a_0^{32} & A_0^{33} \end{bmatrix} \cdot \begin{bmatrix} X_t \\ S_t \\ Y_t \end{bmatrix} - \begin{bmatrix} A_1^{11} & a_1^{12} & A_1^{13} \\ A_1^{21} & a_1^{22} & A_1^{23} \\ A_1^{31} & a_1^{32} & A_1^{33} \end{bmatrix} \cdot \begin{bmatrix} X_{t-1} \\ S_{t-1} \\ Y_{t-1} \end{bmatrix} - \cdots - \begin{bmatrix} A_L^{11} & a_L^{12} & A_L^{13} \\ A_L^{21} & a_L^{22} & A_L^{23} \\ A_L^{31} & a_L^{32} & A_L^{33} \end{bmatrix} \cdot \begin{bmatrix} X_{t-L} \\ S_{t-L} \\ Y_{t-L} \end{bmatrix} = \begin{bmatrix} \varepsilon_t^X \\ \varepsilon_t^S \\ \varepsilon_t^Y \end{bmatrix}$$

where a_0^{22} is a scalar, and A_0^{11} and A_0^{33} are lower triangular matrices, thus A_0 is lower triangular, whose diagonal elements are not necessarily equal to one by using a normalization of $Cov(\varepsilon_t) = \Omega = I$. Note that A_0^{11}, which is analogous to Γ_0, is not identity matrix I but lower triangular matrix.

· The corresponding policy reaction function:

$$\begin{bmatrix} A_0^{21} & a_0^{22} & 0 \end{bmatrix} \cdot \begin{bmatrix} X_t \\ S_t \\ Y_t \end{bmatrix} - \begin{bmatrix} A_1^{21} & a_1^{22} & A_1^{23} \end{bmatrix} \cdot \begin{bmatrix} X_{t-1} \\ S_{t-1} \\ Y_{t-1} \end{bmatrix} - \cdots - \begin{bmatrix} A_L^{21} & a_L^{22} & A_L^{23} \end{bmatrix} \cdot \begin{bmatrix} X_{t-L} \\ S_{t-L} \\ Y_{t-L} \end{bmatrix} = \varepsilon_t^S$$

$$S_t = -\left(a_0^{22}\right)^{-1} A_0^{21} X_t + \left(a_0^{22}\right)^{-1} \begin{bmatrix} A_1^{21} & \alpha_1^{22} & A_1^{23} \end{bmatrix} \cdot \begin{bmatrix} X_{t-1} \\ S_{t-1} \\ Y_{t-1} \end{bmatrix} + \cdots + \left(a_0^{22}\right)^{-1} \begin{bmatrix} A_L^{21} & \alpha_L^{22} & A_L^{23} \end{bmatrix} \cdot \begin{bmatrix} X_{t-L} \\ S_{t-L} \\ Y_{t-L} \end{bmatrix} + \left(a_0^{22}\right)^{-1} \varepsilon_t^S .$$

The policy in period t is determined by (a) a policy reaction rule which depends only on the contemporaneous X_t but not Y_t, (b) all the lagged variables in the VAR system, and (c) a policy shocks in ε_t^S.

· The reduced form VAR:

$Z_t = B_1 Z_{t-1} + \cdots + B_P Z_{t-P} + u_t$ where $B_p = A_0^{-1} A_p$ and $u_t = A_0^{-1} \varepsilon_t$, thus $Cov(u_t) = \hat{\Omega} = A_0^{-1} \Omega \left(A_0^{-1}\right)^T$.

When a normalization of $Cov(\varepsilon_t) = \Omega = I$ is used, A_0^{-1} can be obtained by applying the Cholesky decomposition rule of lower triangular matrix for the symmetric positive definite covariance matrix as $A_0^{-1} = chol\left(\hat{\Omega}\right)$, instead of solving the system of equations $\hat{\Omega} = A_0^{-1}\left(A_0^{-1}\right)^T$.

· The vector moving average (VMA) form or impulse response function (IRF):

$Z_t = u_t + C_1 u_{t-1} + \cdots + C_\infty u_{t-\infty}$ or $Z_t = A_0^{-1}\varepsilon_t + C_1 A_0^{-1}\varepsilon_{t-1} + \cdots + C_\infty A_0^{-1}\varepsilon_{t-\infty}$.

When the policy variable is the jth element in Z_t, the impulse response with respect to the policy shock is the jth columns of the sequence $\left\{A_0^{-1}, C_1 A_0^{-1}, \cdots, C_\infty A_0^{-1}\right\}$ with the assumption that the jth element in u_t unity and all other elements zero. Given that the inverse of lower triangular (A_0) is also lower triangular (A_0^{-1}), (a) the policy shock in period t (ε_t^S) has a contemporaneous effect only on Y_t but not X_t. Thus the partitioning of all variables Z_t into variables X_t and variables Y_t is important for impulse response function of entire variables with respect to the policy variable innovation shock ε_t^S. However, (b) for the study of (monetary) policy effects, the orderings within X_t and Y_t blocks do not matter for the impulse response function of any variable with respect to ε_t^S. Note that all other elements in u_t are assumed to be zero, except jth element in u_t (Christiano, Eichenbaum, and Evans, 1999). This implies that the identification problem to decide causal ordering among variables Z_t in the recursive assumption can be reduced into the partial identification problem to decide which variables come before and after the policy instrument variable in contemporaneous time, since the ordering within those blocks can be unimportant for specific object of study: understanding effects of (monetary) policy shocks.

Note that when the policy variable is assumed to be in either the first or the last causal order, the identification problem becomes trivial for specific object of study: understanding effects of (monetary) policy shocks. The typical identification assumption in much of Sims' earlier work (for example, Sims, 1980) is that the monetary policy variable is unaffected by contemporaneous innovations in other variables. In latter work by Sims and others, the monetary policy variable is assumed to be potentially affected by contemporaneous macro-economic variables instead. This ordering change of monetary policy variable from first to last can be understood by the change of variable choice to represent monetary policy variable from the money aggregate to the federal fund rate. This change to represent monetary policy instrument is based on following arguments among others. (a) A policy variable should be able to predict macro-economic variables and it is found that the federal funds rate produces better forecasts of output, employment and consumption than monetary aggregates such as M1 and M2. (b) While the expansionary monetary policy shock is expected to increase output and decrease money stock and interest rate, the positive shock to M1 leads to decrease output and increase federal funds rate in typical VAR of the U.S. economy. The estimates of policy reaction based on federal funds rate functions produce reasonable responses to inflation and unemployment shocks. (c) It is observed that the federal funds rate was raised at all cyclical peaks (NBER) and at most of the Romer dates (see Bernake and Blinder, 1992 and Eichenbaum, 1992 for examples).

Even for the general purposes, the entire causal ordering among variables of Z_t in the full recursive system can be unimportant in a certain circumstance. When the covariance matrix of the estimated residuals is almost an identity or diagonal matrix and the assumption of the full recursive system is used, the relationship of $Cov(u_t) = \hat{\Omega} \approx I = A_0^{-1}\left(A_0^{-1}\right)^T$ implies that (a) There is only one lower triangular matrix returned by a Choleski decomposition. (b) A_0^{-1} is not only lower triangular but also diagonal, which in turn suggest that neither the ordering in full recursive assumption nor the identification itself is important. When A_0^{-1} is diagonal matrix, the relationship of $u_t = A_0^{-1}\varepsilon_t$ implies that reduced and structural form shocks are proportional with each other.

The full recursive Wold causal chain structure, which makes the close correspondence of SEM and VAR approaches for a specific purpose, is very restrictive assumption to represent the real causal structures. Note that when the empirical study aims to understand impulse responses with respect to not only policy variable but also other structural shocks and when the covariance matrix of the estimated residuals is not a diagonal matrix, the entire causal ordering among variables Z_t in the full recursive assumption is important for the result in impulse response functions. There are $N!$ or $M!$ possible causal orders in VAR or SEM approach respectively and the results in both approaches are sensitive to the specific causal ordering among (endogenous) variables. In more general circumstances where non-recursive causal structures exist, the correspondence of SEM and VAR approaches is no longer valid, since the block recursive Wold causal chain structure, which discriminates endogenous and predetermined variables, does not by itself guarantee identification in the SEM approach. The order and rank conditions in the SEM approach to the identification problem requires: (a) Discriminating endogenous and exogenous variables such that a sufficient number of exogenous variables are identified relative to endogenous variables and (b) Imposing restrictions on not only $\alpha_j^* = 0$ but also $\beta_j^* = 0$ such that $rank\lfloor \pi_j^* \ \underline{\Pi_j^*} \rfloor = rank\lfloor \underline{\Pi_j^*} \rfloor = M_j$ for unique solution of α_j. In this respect, the SEM approach requires: (a) The causal information to discriminate endogenous and exogenous variables, (b) The causal information among endogenous variables (restrictions on α), and (c) The very specific causal information between endogenous and exogenous variables (restrictions on β) to discriminate the simultaneous relationships of the *j* th endogenous variable and other endogenous variables included in the *j* th equation by using the exogenous variables, which are not in the *j* th equation but in other equations for endogenous variables included in the *j* th equation, as the specific shifters or additional causal determinants.

Sims (1980) argues that the restrictions used in usual SEM approach are neither credible nor required. The restrictions used in usual SEM framework are incredible in a sense that they are imposed simply because they are required to attain identification, given that theories are too heterogeneous to provide a conclusive causal structure or the overall theories do not provide sufficient information to identify causal structure. Even though the exogenous variables, defined as variables determined outside the model by assuming all exogenous variables are uncorrelated, provide general bounds of causal information in SEM framework, some variables are assumed as exogenous simply because seriously explaining them would require additional extensive modeling effort in areas away from the main interests of the model-builders. In this respect, the causal information to discriminate endogenous and exogenous variables assumed in usual SEM framework is incredible, given that the presence or absence of exogeneity cannot be inferred from the data and hence is not testable, as many economists using the SEM framework admit. The very specific causal information between endogenous and exogenous variables used in usual SEM framework (restrictions on β relative to the restrictions on α) is also incredible. For example for identifying this type of restrictions, based on the typical distinction between nature and tastes in micro-economics, although it is usually assumed that the weather affects supply and not much demand, whereas the demographic structure of the population affects demand but not much supply, consumers' demand decisions can still rely on information of supply shift variables such as weather and firms' hiring decisions can still depend on forecasts of the demand shift variables such as demographic variable, especially under the rational expectation hypothesis.

All the restrictions used in usual SEM framework are not required for forecasting and/or policy analysis in a sense that the SEM approach requires too much causal information and an alternative approach is possible for forecasting and/or policy analysis. While the causal information should be very specific to meet the order and rank conditions in the SEM framework, the causal information for identification in the VAR framework, as an alternative to the SEM framework, is less demanding. Unlike the SEM approach, the VAR approach to the identification problem does not require: (a) The causal information to discriminate endogenous and exogenous variables, since all the variables in the VAR framework are considered as endogenous and treated symmetrically and (b) The causal information on the dynamic structure in the sequence of coefficient matrices $\{A_1, \cdots, A_p\}$, which is analogous to the very specific causal information between endogenous and exogenous variables used in usual SEM framework (restrictions on β relative to the restrictions on α). The causal information required for identification in the VAR framework is only for the contemporaneous coefficient matrix A_0, which controls how variables are causally linked to each other contemporaneously and relates the structural and reduced form VAR specifications. This advantage of the VAR framework, as an alternative to the SEM framework, increases the possibility of incorporating inductively inferred causal information from statistical properties of the data into the econometric model without pretending to have too much *a priori* theory and/or without demanding too much information from the data. However, given that the reduced form VAR can only be interpreted as the descriptive statistical models, which summarizes observational equivalent statistical properties of data just like correlation in dynamic setting, it is still impossible to use this descriptive statistical model to study effects on variables in the model with respect to economically meaningful structural shock. In this respect, how to determine the causal structure to relate empirical regularities captured in reduced form model to theoretical properties represented by the structural form model remains an important methodological issue to be addressed. Note that even when the covariance matrix of the estimated residuals in the reduced form VAR is almost an identity or diagonal matrix, without the assumption of the full recursive causal structure, the relationship of $Cov(u_t) = \hat{\Omega} \approx I = A_0^{-1}(A_0^{-1})^T$ implies that there can be many A_0^{-1} matrices whose columns are orthonormal (orthogonal matrices)

The identification problem can be understood in the more general context of the induction problem, where the causal structure is in general underdetermined by the statistical properties of the data. A simple but fundamental version of this induction problem is that correlation does not imply causation. In this respect, several inductive approaches to infer causal structures from data are proposed. Among them, the probabilistic approach is widely discussed, especially in the context of the VAR approach. In the probabilistic approach, Suppes (1970) defines causality such that (a) An event A causes prima facie an event B if the conditional probability of B given A is greater than B alone (prima facie causality) and (b) A occurs before B (temporal order condition). The condition of $P(B \mid A) > P(B)$ without temporal order condition is not enough to incorporate asymmetry of causality, since $P(B \mid A) > P(B)$ implies $P(A \mid B) > P(A)$, given that $P(B \mid A) > P(B) \Rightarrow \dfrac{P(A, B)}{P(A)} > P(B) \Rightarrow \dfrac{P(A \mid B)P(B)}{P(A)} > P(B)$ $\Rightarrow P(A \mid B) > P(A)$. This problem occurs due to the symmetrical property $P(A, B) = P(B \mid A)P(A) = P(A \mid B)P(B)$ in the conditional probability, just as the correlation has the symmetrical property. Note that it can be understood that the identification problem in system of equations are due to analogous symmetrical property of reduced form equations for the structural equations. Beside the statistical property, the temporal order is the additionally required condition that allows incorporating the asymmetry of causality, since $P(B_{t+1} \mid A_t) > P(B_{t+1})$ does not imply $P(A_{t+1} \mid B_t) > P(A_{t+1})$.

Based on the similar logic that: (a) A cause makes an effect more likely and (b) A cause occurs before an effect, Granger (1980) defines causality such that a (time-series) variable A causes B, if the probability of B conditional on its own past history and the past history of A does not equal the probability of B conditional on its own past history alone $P(y_t \mid \{y_{t-p}\}_{p=1}^{P}, \{x_{t-p}\}_{p=1}^{P}, I_{t-1}) \neq P(y_t \mid \{y_{t-p}\}_{p=1}^{P}, I_{t-1})$. However, this causality concept, based on the incremental predictability with the temporal order condition, is still not enough to identify the contemporaneous causal structure, which is required for the identification in the VAR approach. The relationship between Granger causality of $y_t \leftarrow x_{t-1}$ and structural contemporaneous causality of $y_t \leftarrow x_t$ can be understood by using the following simple two-variable structural and reduced form VAR example.

· Structural form VAR focusing on structural causality test of $y_t \leftarrow x_t$:

$y_t = a_0^{12} x_t + a_1^{11} y_{t-1} + a_1^{12} x_{t-1} + \varepsilon_t^y$ and $x_t = a_0^{21} y_t + a_1^{21} y_{t-1} + a_1^{22} x_{t-1} + \varepsilon_t^x$

· Reduced form VAR focusing on Granger-causality test of $y_t \leftarrow x_{t-1}$:

$$y_t = \left(\frac{a_0^{12} a_1^{21} + a_1^{11}}{1 - a_0^{12} a_0^{21}} \right) y_{t-1} + \left(\frac{a_0^{12} a_1^{22} + a_1^{12}}{1 - a_0^{12} a_0^{21}} \right) x_{t-1} + \left(\frac{a_0^{12} \varepsilon_t^x + \varepsilon_t^y}{1 - a_0^{12} a_0^{21}} \right) \text{ or } y_t = b_1^{11} y_{t-1} + b_1^{12} x_{t-1} + u_t^y$$

There are all four logically possible relations between b_1^{12} and a_0^{12}: (a) $b_1^{12} \neq 0$ if $a_0^{12} \neq 0$ and $a_1^{22} \neq -a_1^{12} / a_0^{12}$, (b) $b_1^{12} = 0$ if $a_0^{12} = 0$ and $a_1^{12} = 0$, (c) $b_1^{12} \neq 0$ if $a_0^{12} = 0$ and $a_1^{12} \neq 0$, and (d) $b_1^{12} = 0$ if either $a_0^{12} \neq 0$ and $a_0^{12} = -a_1^{12} / a_1^{22}$ or $a_0^{12} \neq 0$ and $a_1^{22} = a_1^{12} = 0$. The corresponding relationship between the two causality concepts can exist as in cases (a) and (b), but the possible non-corresponding relationship as in cases of (c) and (d) can not be excluded. There is no systematic relationship between Granger-causality of $y_t \leftarrow x_{t-1}$ and structural-causality of $y_t \leftarrow x_t$. Since structural-causality a_0^{12} neither implies nor is implied by Granger-causality b_1^{12}, it can be argued that a Granger causality test in a reduced form VAR is not enough to identify the contemporaneous causal structure in a structural form VAR. Note also that given that $b_1^{12} = 0$ implies $a_0^{12} a_1^{22} + a_1^{12} = 0$ not $a_0^{12} = a_1^{12} = 0$, it can be argued that Granger causality does not imply strict exogeneity, whereas strict exogeneity implies Granger causality, since $a_0^{12} = a_1^{12} = 0$ implies $b_1^{12} = 0$. Note also that the restriction $a_0^{12} = 0$ implies $u_t^y = \varepsilon_t^y$, thus hypothesized shock in impulse response function has a clear interpretable meaning in the structural VAR approach (Hoover, 2006).

Not only does Granger-causality not provide enough causal information to solve the induction problem, Granger-causality concept itself has some problems as a legitimate causal definition. Among them, two issues can be understood by following two examples.

$$\text{(a)} \begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} b_p^{11} & b_p^{12} & 0 \\ 0 & b_p^{22} & 0 \\ 0 & b_p^{32} & b_p^{33} \end{bmatrix} \cdot \begin{bmatrix} Z_{t-p}^1 \\ Z_{t-p}^2 \\ Z_{t-p}^3 \end{bmatrix} + \begin{bmatrix} u_t^1 \\ u_t^2 \\ u_t^3 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} Z_t^1 \\ Z_t^2 \\ Z_t^3 \end{bmatrix} = \sum_{p=1}^{P} \begin{bmatrix} b_p^{11} & b_p^{12} & 0 \\ b_p^{21} & b_p^{22} & b_p^{23} \\ b_p^{31} & b_p^{32} & b_p^{33} \end{bmatrix} \cdot \begin{bmatrix} Z_{t-p}^1 \\ Z_{t-p}^2 \\ Z_{t-p}^3 \end{bmatrix} + \begin{bmatrix} u_t^1 \\ u_t^2 \\ u_t^3 \end{bmatrix}$$

Granger-causality test is sensitive for information set I_{t-1} as in above example (a). $b_p^{12} \neq 0$ implies $Z_t^1 \leftarrow Z_{t-s}^2$, $b_p^{32} \neq 0$ implies $Z_t^3 \leftarrow Z_{t-s}^2$, and $b_p^{13} = 0$ implies $Z_t^1 \not\leftarrow Z_{t-s}^3$. However, excluding common cause Z_t^2 from information set I_{t-1} can mislead one to conclude $Z_t^1 \leftarrow Z_{t-s}^3$, since Z_{t-1}^3 has information of Z_{t-1}^2 which does have information about Z_t^1. Granger-causality concept does not guarantee transitivity of causality as in above example (b). $b_p^{12} \neq 0$ implies $Z_t^1 \leftarrow Z_{t-s}^2$ and $b_p^{23} \neq 0$ implies $Z_t^2 \leftarrow Z_{t-s}^3$ but $b_p^{13} = 0$ implies $Z_t^1 \not\leftarrow Z_{t-s}^3$. In this respect, it is conceivable that the omitted variable problem can occur in a small information set and the violation of transitivity can occur in a larger variable set. The variable selection approaches based on regression methods with several diagnostics or inclusion/deletion criteria have the similar issue. (a) When the small explanatory variable set is initially assumed and then subsequently expanded into larger selected variable set (Bottom-up approach), the omitted variable (especially common cause variable) problem in the initial (or subsequent) small model can mislead the subsequent procedures. For example, if true causal structure is $y_t \leftarrow W_t \rightarrow x_t^1$ but the initial small model $y_t = a_1 x_t^1 + \varepsilon_t$ omits the common cause variable W_t, then hypothetic test of $H_0 : a_1 = 0$ can be rejected. (b) When the large explanatory variable set is initially assumed and then subsequently reduced into smaller selected variable set (Top-down approach), the included variable (especially common effect variable) problem in initial (or subsequent) large model can mislead the subsequent procedures. For example, if true causal structure is $y_t \rightarrow W_t \leftarrow x_t^1$ but the initial large model $y_t = a_1 x_t^1 + \beta W_t + \sum a_k x_t^k + \varepsilon_t$ includes the common effect variable W_t, then the hypothetic test of $H_0 : a_1 = 0$ can be rejected. In this respect, it can be argued that the variable selection approach and the Granger's causality test have the same difficulty to decide the appropriate explanatory variable or information set. Given that asymmetry and transitivity (if cause and effect relation is effective) are two intuitive properties of the causality concept, the prima facie causality based on the conditional probability has difficulty to incorporate asymmetry and Granger's definition based on the incremental predictability has an ambiguity with respect to transitivity. The causal concept based on the temporal order does not provide enough information for the contemporaneous causal structures, which is required for the identification in the VAR approaches.

We propose to use the graphical causal models as an alternative inductive method of inferring contemporaneous causal relationships from non-temporal and non-experimental data in this study. The graphical causal models have been developed by mathematically connecting probabilistic structures to graphical concepts, which effectively and efficiently capture all the probabilistic structures. The graphical causal model or directed acyclic graph (DAG) approaches are based on several mathematical propositions. Let A, B, and C denote three disjoint subsets of variables, called vertices or nodes. When it is assumed that the cyclic or feedback causal structure does not exist (causal acyclic condition) and all the causally relevant variables can be measured (causal sufficiency condition), the probability distribution follows the Markov condition such that every variable is independent of all its causal nondescendants, conditional on its direct cause. This implies that (a) An effect is independent of its indirect causes conditional on its direct causes (causal chains of $A \rightarrow C \rightarrow B$ or $A \leftarrow C \leftarrow B$) and (b) The effect variables are independent conditional on their common causes (causal fork of $A \leftarrow C \rightarrow B$). Note that two nodes A and B in both causal chain and fork are unconditionally or marginally dependent on each other, but conditionally independent given C. This observation provides a causal interpretation for a simple but fundamental version of induction problem that (unconditional) correlation does not imply causation. In the statistical literature, the other logically possible causal structure except cyclical structure among three nodes is known as the selection bias, where observation on a common consequence of two marginally independent causes tends to make those two causes dependent conditional on common effect. This selection bias occurs because information about one of two causes tends to make the other more or less likely, given that the consequence is observed (unshielded-collider of $A \rightarrow C \leftarrow B$). Note that this causal structure of the unshielded-collider provides an "empirical clue" to address induction problem that correlation does not imply causation, since the combinational statistical information of marginal correlation (unconditionally independence of A and B) and partial correlation (conditional dependence of A and B given C) make it possible to infer the causal structure of the unshielded-collider, which is discriminated from the observational equivalent causal structures of causal chain and fork. Note also that acyclic condition allows excluding possible cyclic structures and sufficiency condition allows including the causal fork structure.

In graphical causal models, it is also assumed that all the marginal and conditional probabilistic structures are invariant to the changes of their numerical or parametric values (probabilistic stability condition). This implies (a) All the observed (un)conditional probabilistic structures are due to the underlying causal structures, not their special numerical values in probabilistic structures. (b) No (in)dependencies in probability structures can be destroyed or induced by changing probabilistic parameter values. (c) It is possible to effectively and efficiently encode (un)conditional (in)dependencies structures into graphical model without numerical probabilities. Thus, with the Markov condition, (d) It is possible to infer the underlying causal structures from the observed marginal and conditional probabilistic structures, where the observation is done through the statistical decisions based on either the Neyman-Pearson type statistical test (conditional independence test approach) or the Bayesian information criterion (goodness-of-fit scoring approach). To empirically infer the marginal and conditional probabilistic structures, two distinctive approaches have been proposed. While the accessible explanation is provided in chapter II, the conditional independence test approach is explained in Spirtes et al. (2000), the goodness-of-fit (Bayesian) scoring approach is explained in Chickering (2003), and more theoretical and conceptual aspects of graphical causal models are explained by Pearl (2000). While the PC algorithm incorporates the conditional independence test approach, GES algorithm take the goodness-of-fit Bayesian scoring approach. The PC algorithm is discussed in Spirtes et al. (2000) and the GES algorithm is originated from Meek (1997) and its optimality is proved by Chickering (2003). Spirtes et al (2000) also provide several algorithms in their operational software “Tetrad”, which can be used to implement the PC and GES algorithms.

The observed equivalence between the causal chain and the causal fork, which is again a simple version of induction problem that correlation does not imply causation, can not be discriminated based only on statistical observations without using non-observational extra causal information or manipulative (randomized) experimentation. However, the graph theory provides “logical clues” to partially address the observational equivalence problem. After the maximum information of unconditional and conditional probabilistic structures from data is obtain, (a) All the discriminative information between the true statistical relationships and spurious correlations among variables without causal orientations are summarized into the graph with undirected edges, named as the skeleton, and (b) All the information to discriminate the unshielded-collider structure from the observational equivalent causal structures of causal chain and fork are summarized into the partially oriented graph, named as the partially directed acyclic graph (PDAG) with causal orientations from independent causes to the common effect. By logically deciding causal directions for the remaining undirected edges in PDAG, the completed partially directed acyclic graph (completed PDAG or essential graph), which is maximally oriented PDAG, can be further inferred. The logical inferences about causal directions are based on the idea that orienting the remaining undirected edges in PDAG does not result in the causal structure which is inconsistent with the statistical observations, as long as the logically decided orientations do not create either the empirically unsupported new unshielded-collider structure or the cyclic causal structure excluded by the acyclic assumption.

The graphical causal models or DAG approaches have several features and assumptions. To infer causal relationship between two variables A and B, the DAG use the criterion whether a third variable C exhibits a specific pattern of dependency with A and B. In this respect, the DAG approach can be compared with the SEM approach, where the simultaneous relationships of the j th endogenous variable (A) and other endogenous variables included in the j th equation (B) are discriminated (identification or induction problem) by the assumed exogenous variables (C) excluded from the j th equation as the additional third causal determinants or specific shifters for behavioral equations of other endogenous variables included in the j th equation. However, methods to address this induction problem are quite different. In the SEM approach, the selection of exogenous variables is usually considered as maintained assumptions derived from the theory rather than something to be learned form data itself. Even when the hypothetical test approach based on regression framework is implemented, (a) The non-nested hypothetical test approaches oftentimes have a power problem related with the statistical hypotheses test, so that they have generally little power to discriminate competing specifications or causal hypotheses. (b) The nesting of hypothetical tests based on variable selection methods also faces issues, since the top-down or bottom-up approach have difficulties in dealing with common effect variables or common cause variables of dependent and explanatory variables respectively as mentioned above. In the DAG approach, all marginal and conditional probabilistic structures among all the relevant combinations of variables are efficiently checked in search procedures to obtain the maximum information of specific pattern of dependencies among variables, where relevant search spaces are logically decided based on the graph theory. In this respect, the graph theory not only provides logical orientation rules to partially discriminate observationally equivalent causal structures but also allows the full use of the maximum information of unconditional and conditional probabilistic structures from the data. Note that checking or searching all the relevant (un)conditional probabilistic structures among all the possible combinations of variables becomes infeasible without systematically and efficiently defining the relevant or entire search space.

The graph theory used in the DAG approach is based on some assumptions as discussed earlier. The acyclic assumption and the causal sufficiency assumption are required for the Markov conditions. While Richardson and Spirtes (1999) develop the Cyclic Causal Discovery

(CCD) algorithm to allow cyclic possibility and Spirtes et al. (2000) develop the Fast Causal Inference (FCI) algorithm to relax sufficient condition, these developments are not incorporated in this study, since it is still ambiguous how to distinguish between feedback and latent phenomena (Moneta and Spirtes, 2006). We hope that it is not too harmful to take the acyclic and sufficiency assumptions, given observation that these two assumptions are common to almost all the empirical models. Given the fact that while the Markov condition suggests the logical implication from the underlying causal structures to probabilistic dependency patterns, the stability condition, with the Markov condition, suggests the logical implication from probabilistic dependency patterns to the underlying causal structures, it can be argued that the stability condition, with the Markov condition, makes it possible to inductively infer the causal structures from the data. In this respect, the stability condition needs to be discussed more to use the graphical causal model in empirical study. There can be two circumstances where the stability condition can be violated, as discussed in the Tetrad II manual. One possible circumstance is that there may exists strict equality among products of parameters, so that a spurious (in)dependence in probability distribution can be destroyed or induced by changing underlying parameter values. The other possible circumstance is that there may exist deterministic or near deterministic relationships among variables so that any of the statistically observed (un)conditional probabilistic structures are due to not only the underlying causal structures but also their special numerical values. For the first case, it has been demonstrated that the strict equalities among products of parameters have very little possibility or Lebesgue measure of zero in any probability space in which parameters vary independently from one another. According to Tetrad II manual, the Tetrad program should not be used for the following cases or these second cases should be practically addressed in empirical study, where (a) There are deterministic relationships among variables or (b) There are conditional probabilities very close to 1 in the discrete case or (c) There are correlations very close to 1 in the linear case. These restrictions for using the Tetrad program can be understood based on the following reasoning. If $P(A \mid B) \approx 1$, then $P(A \mid B, C) = P(A \mid B)$ can be hold for any set of variable C, regardless of the causal structures among them. So it is not possible to infer reliable causal structure from the probabilistic dependency pattern. Note that this problem is similar to the multicollinearity problem, which makes it difficult to obtain precise estimates of the separate effects of the variables in the regression method. Given the observation that many variables in high dimensional data set oftentimes move very closely, the direct use of the graphical causal model for the high dimensional data set can be problematic, since the stability condition can be violated in its applications for high dimensional data sets. One possible way to address this problematic situation is to use aggregation method. However, before using an aggregation method, the legitimate aggregation condition should be empirically identified to consistently infer causal structures among disaggregated variables by the aggregated variables as the legitimate representatives. This issue is the next topic to be discussed.

Aggregation in Study of Monetary Policy Effect

To promote sustainable growth and stabilize inflation have been considered as main goal of macro-economic policy. While fiscal and monetary policy have been considered as two primary policy instruments to attain that goal, it is observed that monetary policy has become more emphasized than the fiscal policy, since (a) fiscal policy brings not only doubts that the tax and spending decisions can not be made in timely way, but also concerns that using fiscal policy in inappropriate ways can result in the possible persistent budget deficits, (b) it is observed that the monetary policy effects do exist over the short and mid run period, despite of the argument that the monetary policy has neutral effects on economic activity in the long run. In this respect, the understanding of how monetary policy affects the economic activity has been the primary topic for theoretical and empirical studies in macro-economics for a long time. While there have been many approaches to study the monetary policy transmission mechanism, the structural vector autoregressive (VAR) framework is widely used, since it does not require the excessive and incredible identifying restrictions in the structural equation model (SEM) framework. Sims (1980) introduces VAR approach as an alternative to SEM approach and Sims (1992) and Bernake and Blinder (1992) use these models to identify and measure the effect of monetary policy on macro-economic variables. However, beside the causal identification issue previously discussed, the relatively small information set incorporated in the standard low dimensional VAR model may imply potential problems in the empirical understanding of the monetary policy transmission mechanism based on the small number of variable VAR model, given the observation that (a) monetary authorities monitor a large number of economic variables and (b) there can be many possible channels through which the monetary policy affect the economy. Accordingly, there are research interests in moving beyond the low dimensional VAR.

First, when the central bank and the private sector have additional information not incorporated in the model, the policy innovations measured by reduced form VAR residuals of policy reaction functions is likely to be contaminated and the measured responses of economic variables to the monetary policy innovations is also likely to be misleading. The possibility that there can be missing elements in the policy reaction functions can be understood by using following example of the "price puzzle". The price puzzle is counter-intuitive impulse responses result that contractionary shocks to monetary policy lead to persistent price increases in a VAR of output, prices, money, interest rate and perhaps some more variables. When the policy rule i_t (the federal fund rate, for example) is represented as the function of the inflation expectations $E_t(\pi_{t+1})$, the effect of other variables $g(X'_t)$, and the policy shock ε_t^s but the expected inflation is actually some function of not only some variables I_t included in VAR but also some other variables W_t omitted in the VAR, the VAR residual for policy variable u_t^s is actually some function of not only the policy shock ε_t^s but also omitted variables $\pi_t^w(I_t, W_t)$, which have information about the expected inflation. If these omitted variables dominate the policy shock, then a primary component of the monetary policy shocks measured from the reduced form VAR residual is actually the omitted information about the expected inflation, which can lead to high future inflation.

· The underlying policy reaction function:

$i_t = \beta E_t(\pi_{t+1}) + g(X'_t) + \varepsilon_t^s$ by assumption of $E_t(\pi_{t+1}) = \pi_t(I_t) + \pi_t^w(I_t, W_t)$

$i_t = \beta \pi_t(I_t) + g(X'_t) + \beta \pi_t^w(I_t, W_t) + \varepsilon_t^s$

· The measured policy innovations:

$i_t = \beta \pi_t(I_t) + g(X'_t) + u_t^s$ where $u_t^s = \beta \pi_t^w(I_t, W_t) + \varepsilon_t^s$.

One possible solution to this price puzzle would be to include the omitted information about the expected inflation in VAR, which makes $\pi_t^w(I_t, W_t)$ term vanished. A large number of possible variables are studied and it is demonstrated that broad commodity price indices and some of financial data seems to be successful, whereas individual commodity prices have very small effects (Sims, 1992). However, there is difficulty in this approach to address the price puzzle. Although additional variables W_t, which represent the omitted information about the expected inflation in VAR to solve the price puzzle, must have incremental predictive power for future inflation over the Ω_t, it is not easy to find the empirical support for this argument. For example, Hansen (2004) compares several commodity price indices and other indicators and finds very little correlation between the ability to forecast inflation and to solve the price puzzle.

Second, when there are additional channels not incorporated in the VAR model, the measured responses of economic variables to monetary policy shocks can be misleading. The possibility that there can be missing elements in monetary transmission mechanisms or channels not captured in the VAR model is based on the observation that macro-economic responses to policy-induced interest rate changes are considerably larger than those implied by the conventional estimates of the interest rate elasticities of consumption and investment (Bernanke and Gertler, 1995). The theoretical descriptions of the monetary transmission mechanism are based on the following arguments that (a) The monetary policy affects the short and long term nominal as well as real interest rates. Short-term nominal and real interest rates are assumed to move in the same directions by the nominal rigidities of general price level. On the other hand, short-term and long-term interest rates are also assumed to move in the same directions by the rational expectation hypothesis of the term structure, which states that the long-term interest rate is an average of expected future short-term interest rates. Thus, hereafter interest rates (i) denote all the general interest rates. (b) The size of economy (Y) can be measured by the expenditure method, which states that the market value of final goods and services or the sum of value added at every stage of production within a country in a given period of time can be measured by planned investment (I), consumer spending (C), government spending (g) and net exports (NX). There are several monetary policy transmission channels described in the literatures including one based on the traditional macro-economic models (Mishkin, 1995).

In the traditional ISLM macro-models, the monetary transmission mechanism can be described as follows. (a) The general interest rate (i) moves in the same direction as the required rate of return (cost of capital r) or the discount rate (r'). While investment spending is affected through the influence on the required rate of return of investments (cost of capital r), consumer spending is affected through the relative price of current and future consumption (discount rate r'). Both investment and consumer spending are also affected by the lending and borrowing activities. (b) The relative attractiveness of domestic currency to foreign currency due to domestic interest rate change affects the relative value of domestic currency to foreign currency E (exchange rate). The exchange rate affects the relative price competitiveness of domestic goods to foreign goods, which influence the net export. The exchange rate also affects the domestic debt burden denominated in foreign currency (Mishkin, 1995).

Descriptions of (other) asset market channels are based on the following several alternative propositions, where (other) assets markets are represented by the financial assets prices (P_S) and physical assets prices (P_H). While the common stocks usually represent financial assets or wealth, the residential housing and durable goods represent physical asset or real capital. (a) The contractionary monetary policy decreases money supply and increases interest rate. Decreased money supply induces public to spend less, decreasing the demand for financial and physical assets. Increased interest rates makes bonds more attractive relative to other assets, decreasing financial and physical assets prices. (b) Based on the Tobin's q theory of investment, where q is defined as the ratio of market value of asset to the replacement cost of capital, it is argued that when asset price is decreased and thus q is decreased, spending on asset become expensive relative to asset market value and thus investment spending on asset decreased. Just like firms' decisions about business investment, consumers' decisions about residential housing and durable goods are considered as investment decisions. (c) Based on the Modigliani's life cycle model, it is argued that the consumption spending is also determined by the lifetime resources of consumers, which consist of human capital, real capital, and financial wealth. Decreased asset price reduces lifetime resources, which leads to decline in consumption (Mishkin, 2001).

Descriptions of the bank credit channel are based on how bank assess borrowers, especially borrowers' balance sheets (BS). The contractionary monetary policy deteriorates not only the borrowers' debt service burden or cash-flows (CF) by raising interest rate but also the borrowers' collateral value (CV) by decreasing asset prices. The deteriorated balance sheet makes banks' willingness to lend decreased, which implies a decrease in the bank dependent borrowers' investment or consumption. More detailed descriptions are pursued. In the bank side, it is argued that the small banks' willingness to lend is restricted more than the large banks, since small banks are not able to substitute deposits funding with other sources of funds. In the firm side, it is argued that and the small or medium size firms more depend on banks than the large firms for external funds, since small firms can not directly access the credit markets such as stock and bond market. This implies that the monetary policy affects the overall economy through its effects on the small banks or firms. Note that size is used for proxy variable for this argument. On the consumer side, given that financial assets are considered more liquid than physical assets, the change in liquidity affects the willingness to hold non-liquid assets. For example, decreased stock price induced by monetary policy makes consumers' financial position less secure, reducing consumers' expenditure on physical assets, which in turn implies decreased willingness to lend and borrow (Bernanke and Gertler, 1995).

The monetary transmission mechanisms described above can be summarized for the contractional monetary policy (MP) as follows:

MP => $i\uparrow$ => r & r' $\uparrow$ => I & $C\downarrow$ => $Y\downarrow$

MP => $i\uparrow$ => $E\uparrow$ => $NX\downarrow$ => $Y\downarrow$

MP => $i\uparrow$ => P_S & $P_H\downarrow$ => (Tobin's q and/or Wealth) $\downarrow$ => I & $C\downarrow$ => $Y\downarrow$

MP => $i\uparrow$ => (CF and/or P_S & P_H => CV)$\downarrow$ => $BS\downarrow$ => Credit$\downarrow$ => I & $C\downarrow$ => $Y\downarrow$

The existence of additional channels other than the narrow interest rate channel implies that the standard small number of variable VAR model based on the traditional ISLM macro-model can underestimate the monetary policy effects, since the stock and house market, for example, suggest possible amplified indirect monetary effects more than direct interest rate effects. However, the theoretical and empirical descriptions and understandings of monetary transmission mechanism are still incomplete, thus they can not provide clear guidelines for the choice of variables to enter the VAR system. For example, (a) The change in interest rates induced by monetary policy can affect the overall economic activity through the expectations such as inflationary expectations and confidence about the future outlook of the economy. However, the direction in which such effects work can vary from time to time and is hard to predict. (b) The relative importance and their total effect of different transmission mechanisms or channels depends on the different structures and the nature of the economy such as the history of business cycles, differences in depth and diversity of financial markets, different nature and size of firms and/or consumers and their financial structures, the elasticity of demand for exports and imports, relative openness of the economy, relative amount of national debt denominated in foreign currency, and so on. Some of these issues can only be address based on the detailed micro-level data, rather than aggregate data (see Juks, 2004 and references in there).

Third, besides the potential problems due to possible omitted variables in both measuring policy shocks in monetary policy reaction functions and fully capturing monetary transmission mechanism channels, there is a more fundamental issue for choosing variable in empirical models. Watson (2000, page 88) argues "The main problem to be solved when constructing a small model is to choose the correct variables to include in the equation. This is the familiar problem of variable selection in regression analysis. Economic theory is of some help, but it usually suggests large categories of variables (money, interest rates, wages, stock prices, etc.), and the choice of a specific subset of variables then becomes a statistical problem.

The large-model approach is again guided by economic theory for choosing categories of variables, and the statistical problem then becomes how to combine the information in this large collection of variables". The variable selection approach based on regression method can be problematic, since (a) The top-down or bottom-up approach has some difficulties to deal with common effect variables or common cause variables of dependent and explanatory variables respectively as discussed in the context of causality issues and (b) The variable selection approach requires an unrealistic assumption that the very specific observable measures precisely corresponds to some theoretical constructs. The observed variables may be subject to a variety of errors such as (a) The observed variables are likely to be contaminated by measurement errors. Most macroeconomic data may be subject to multiple rounds of revisions and are never free of measurement error. For example, various biases are involved in the measurement of inflation such as the inherent difficulty of full adjustment for quality improvement and (b) There is a conceptual ambiguity in linking each theoretical variable to a specific observed variable. The choice of a specific data series to represent a general economic concept is often arbitrary to some degree and thus a specific measured variable is likely not to correspond to a theoretical variable. For example, output in the theoretical model may correspond more closely to a latent measure of economic activity than to a specific data series such as real GDP. Considering only the common components of observed variables is one way to eliminate measurement errors and treating theoretical variables as unobserved in empirical analysis is one way to acknowledge these underlying problems (Bernanke, Boivin, and Eliasz, 2005). In this respect, an alternative approach to variable selection methods is to use statistical dimensional reduction methods such as factor and principal component analyses, which treat theoretical constructs as unobserved factors revealing their information by their multiple observable indicators. In addition, factors or principal components can be used to combine the information in large collection of variables into empirical models.

As an alternative to the SEM approach, which requires a large number of identifying restrictions for system estimation by two- or three- stage least square methods for either forecasting or policy analysis, the VAR approach requires one to identify the contemporaneous coefficient matrix only in order to infer the structural economic shocks from the reduced form innovations. However, the inference based on the VAR approach can be misleading, unless the reduced form innovations span the space of the structural shocks or the VAR model does not have an omitted variable problem. The main issue to address this possible misspecification problem is how to increase the amount of information in the VAR model so that the reduced form innovations span the space of the structural economic shocks, given that econometric considerations such as degrees-of-freedom and multicollinearity require the economy of parameters in empirical models. In this respect as well as the related problems of the observable measurements with respect to theoretical constructs, the statistical factor model is proposed to span the space of structural shocks, when there can be hundreds of economic variables that potentially contain information about the underlying shocks. Two approaches, commonly named as the dynamic factor model, are suggested to generalize the standard static factor models based on the static covariance or correlation matrix to incorporate the possible distributed lag effect of factors on observed variables. While Forni et al (2000) use the spectral density matrix in a frequency-domain framework, Stock and Watson (2002) use cross-covariance matrix, which includes auto-covariance matrix in a time-domain framework. Since both approaches apply the singular value decomposition theorem to their generalized covariance or correlation matrix to derive eigen-vectors as weighting schemes, the dynamic factor model can be understood as the generalized approximate factor model based on generalized principal component methods. The dynamic factor model approach is based on the propositions that (a) There are small numbers of unobserved common dynamic factors that produce the observed co-movement of economic time series, (b) These common dynamic factors are driven by the underlying common structural economic shocks, (c) these underlying structural shocks are only revealed by distilling the small numbers of common sources of co-movement from a very large number of observed variables. These plausible propositions of dynamic factor models, with the observed co-movement of many economic time series variables, have motivated recent advances in VAR modeling on how to best integrate this factor method into VAR and SVAR analysis for either forecasting or policy analysis (Stock and Watson, 2005).

For the forecasting purpose, Stock and Watson (2002) propose to use an approximate dynamic factor model, where the information of a large numbers of time series variables is summarized by relatively small number of estimated factors, called diffusion indexes. They show using forecasting simulations that forecasts based on estimated factors outperform univariate autoregressive models, small number of variable VAR models and leading indicator models. Let y_{t+1} be the variable to be forecast based on the number N_X of variables X_t through the number N_F of latent factors F_t. Their approach can be understood as follows: If

$y_{t+1} = \beta \cdot X_t + \varepsilon_{t+1}$ and $X_t = \Lambda F_t + u_t$, then $Y_{t+1} = \alpha F_t + e_{t+1}$, where $E\left(e_{s+1} \mid \left\{Y_{s+1}, X_s, F_s\right\}_{s=-\infty}^{t}\right) = 0$. If y_{t+1} is the variable to be forecast based on the vector of variables X_t but the comovement among variables X_t can be summarized by a small number of latent factors F_t, then a three step process can be used for forecasting y_{t+1}: (a) Estimate latent factors $\hat{F}_t$ from the observed variables X_t, (b) Estimate coefficients $\hat{\alpha}$ in $Y_{t+1} = \alpha F_t + e_{t+1}$, and (c) Forecast Y_{T+1} based on $\hat{Y}_{T+1} = \hat{\alpha}\hat{F}_T$. Note that they found that only a few observed variables have predictive power, since most of β in $y_{t+1} = \beta \cdot X_t + \varepsilon_{t+1}$ are zero. However, all the observed variables turn out to have predictive power through representative common factors, since all of the elements of β in factor model are non zero in general, even though each of them is small. This means that much more information can be incorporated for analysis by using the dynamic factor model approach. The β in factor model are derived as follows. When $X_t = \Lambda F_t + u_t$ can be written as $E\left(F_t \mid X_t\right) = \gamma \cdot X_t$ (regression of F_t onto X_t is linear), then $Y_{t+1} = \alpha F_t + e_{t+1}$ implies that $E\left(y_{t+1} \mid X_t\right) = \alpha \cdot E\left(F_t \mid X_t\right) = \alpha \cdot \gamma \cdot X_t$ and $\beta = \alpha \cdot \gamma$.

Bernanke, Boivin, and Eliasz (2005) extend this dynamic factor model for the structural VAR approach and propose to use factor-augmented vector autoregressive models (FAVARs) based on the idea that if large amounts of information about the economy can be effectively incorporated in the model by a small number of estimated factors, then augmenting standard VARs with estimated factors can be natural way to incorporate large information set into the structural VAR model. Note that when the number of factors N_F is much smaller than the number of observed variables N_X, the amount of information incorporated in the model drastically increases by using FAVAR framework. Incorporating similar amount of information by directly using observed variables without factor framework would be both inefficient due to possible multicollinearity problem and impractical due to the degree of freedom problem. Their approach can be understood as follows: $\begin{bmatrix} F_t \\ y_t \end{bmatrix} = \Phi(L) \cdot \begin{bmatrix} F_{t-1} \\ y_{t-1} \end{bmatrix} + v_t$, where $X_t = \Lambda_f \cdot F_t + \Lambda_y \cdot y_t + e_t$. When the information structure is assumed such that the central bank and the econometrician observe only the policy instrument (nominal interest rate) y_t with a large set of noisy macroeconomic indicators X_t but the comovement among variables X_t can be summarized by a small number of latent factors F_t, a three step process can be used to study monetary policy effect: (a) Estimate latent factors $\hat{F}_t$ from the observed variables X_t, (b) Estimate impulse response functions of factor augmented VAR, and (c) Obtain impulse response functions of individual macroeconomic indicators based on $X_t = \Lambda_f \cdot F_t + \Lambda_y \cdot y_t + e_t$. Note that assuming a full recursive causal structure among factors and policy variable, they use a Cholesky identification scheme where the policy variable, federal fund rate, is ordered last. Note also that they construct factors from the observed variables' information space not spanned by policy variable $\hat{C}(X_t) - \hat{b} \cdot y_t$, where $\hat{C}(X_t)$ denote principal components of entire observed variables X_t and $\hat{b} \cdot y_t$ is obtained through a multiple regression of $\hat{C}(X_t) = \hat{a} \cdot \hat{C}(X_t^{Slow}{}_t) + \hat{b} \cdot y_t + e_t$ when $\hat{C}(X_t^{Slow}{}_t)$ denote principal components of slow-moving observed variables only. This is based on the assumption of block recursive causal structure among observed variables such that observed variables are divided into slow-moving and fast-moving variables, where the slow moving variables such as real variables are assumed not to respond to policy shock and fast moving variables such as financial asset prices are allowed to respond to a policy shock in contemporaneous time.

Although it is demonstrated that the dynamic factor models are useful approach to incorporate a broad range of information in empirical macroeconomic modeling for either forecasting or policy analysis, it is also observed that there remain several issues to be addresses as Bernanke, Boivin, and Eliasz (2005) discussed. First, there is some ambiguity in the choice of observed variables X_t. For example, Boivin and Ng (2003) using simulation and empirical data demonstrated that expanding the dataset by adding more variables without considering data structure can be not always desirable in the context of forecasting. They show that it is possible to forecast equally well and perhaps marginally better by pre-screening observed variables into smaller dataset, although their pre-screening method is considered as a largely ad hoc procedure. Second, there is some ambiguity in choosing the number of factors F_t. Although some (information) criteria are proposed to determine the number of factors present in the data set X_t (see Stock and Watson, 2002 and Bai and Ng, 2002, for examples), it is argued that these criteria do not necessarily address the question of how many factors should be included in the VAR. For example, Stock and Watson (2005, page 33) argue that "for the purposes of forecasting, it may suffice to use a small number of dynamic factors but for the purpose of structural VAR modeling the dimension of the space of dynamic factor innovations appears to be larger." Third, although the large amounts of information can be effectively incorporated by a small number of estimated factors to improve forecasting performances or empirical plausibility of structural analysis, there is difficulty to provide economic interpretations for the estimated factors F_t . Given that structural VAR is widely used to study the monetary policy transmission mechanisms, it is not enough to mitigate some puzzles such as price puzzles by augmenting estimated factors when the estimated factors can not be economically interpreted. Stock and Watson (2005, page 33) suggest possible economic interpretations for the estimated factors based on the relative size of factor loadings and variance decomposition and argue that "additional dynamic factors account for additional movements of the remaining series, which are mainly financial series such as interest rates, stock returns, and exchange rates". Given that these financial series are the possible monetary transmission mechanism channels identified in the literature as discussed, it can be argued that incorporating additional dynamic factors is important for the structural understanding of the monetary transmission mechanisms. However, it is still not easy to provide clear economic interpretations for the estimated factors, when the estimated factors are linear combinations of the entire data set. Fourth, there remains an ambiguity to use the full recursive contemporaneous causal structure among factors and policy variable. The estimated factors are independent with each other by construction so that the covariance matrix of the estimated residuals of FAVAR is almost diagonal matrix and thus the reduced and structural form shocks are proportional with each other when a Cholesky identification scheme is used. However, given that the estimated factors are linear combinations of the entire data set of observed variables, it is not easy to connect the estimated factors to the underlying structural shocks except the policy variable shock. Bernanke, Boivin, and Eliasz (2005) eschew such difficult issue on how to decide the causal orderings among the estimated factors by using the fact that the orderings within the before or after policy variable block are not important to understand the monetary policy effects when the full recursive causal structure is assumed. In their identification scheme, the policy variable federal fund rate is ordered last so that all the estimated factors are in the higher order placed block. Note that how to construct factors from the observed variables' information space not spanned by policy variable in the first step also depends on the specific identifying assumption used in the second step. Bernanke, Boivin, and Eliasz (2005) use the block recursive assumption of slow variables' block and fast variables' block in the first step and the full recursive in the second step. Given that the full recursive causal structure is considered as a very restrictive assumption to represent the causal structure in the real data, other general identification schemes need to be considered in the FAVAR model. However, there is some ambiguity to use other general identification schemes in FAVAR model, since more general identification schemes would require that the estimated factors to be identified as specific economic concepts.

All the issues discussed above for using FAVAR model are related with the data structure of observed variables X_t and interpretation of estimated factors F_t and these two issues are related each other. The intuitively suggested (Bernake et al., 2005) approach is extracting principal components from blocks of data corresponding to different dimensions of the economy. Mathematically this approach can be explained as follows:

$$\begin{bmatrix} F_t^1 \\ F_t^2 \\ \vdots \\ F_t^I \\ y_t \end{bmatrix} = \Phi(L) \cdot \begin{bmatrix} F_{t-l}^1 \\ F_{t-l}^2 \\ \vdots \\ F_{t-l}^K \\ y_{t-l} \end{bmatrix} + v_t \text{, where } \begin{bmatrix} X_t^1 \\ X_t^2 \\ \vdots \\ X_t^{K-1} \\ X_t^K \end{bmatrix} = \begin{bmatrix} \Lambda_f^1 & 0 & \cdots & 0 & 0 \\ 0 & \Lambda_f^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \Lambda_f^{K-1} & 0 \\ 0 & 0 & \cdots & 0 & \Lambda_f^K \end{bmatrix} \cdot \begin{bmatrix} F_t^1 \\ F_t^2 \\ \vdots \\ F_t^{K-1} \\ F_t^K \end{bmatrix} + e_t$$

If we assume that each block of observed data X_t^k is explained by only the corresponding one factor F_t^k ($\forall\ k = 1, \cdots, K$) and each of the variables in the entire data set X_t is affected each other only through the corresponding factors, then each of estimated factors can be interpreted based on the assumed group of observed variables and the contemporaneous causal structures among the estimated factors can be meaningfully imposed. This approach is empirically used in Belviso and Milani (2005), where two types of deductive assumptions are used. In their empirical study, the classification of observed variables and the contemporaneous causal structure are chosen based on researchers' subjective intuition. Their classification will be discussed in the below empirical section, when our inductive classification is discussed along with another subjective classification of Leeper, Sims, and Zha (1996). Their identification scheme is based on the full recursive restriction where the different causal orderings are tried. Although the results of Belviso and Milani (2005) are generally successful and they call their method as the structural FAVAR (SFAVAR), the deductive approach for aggregation and causality issues can result in ambiguity in empirical studies, given that theory does not provide definitive or sufficient information for these two issues.

The possibility of inductively inferring data structure from observed variables X_t and obtaining interpretable estimated factors F_t from the observational data is discussed from the aggregation theory and statistical dimensional reduction methods in chapter II. The main result can be summarized as follows. If the observed data has the special data structure of $X_t^k = \Lambda_f^k \cdot F_t^k$, then the correlation matrix of observed data set X_t have the special block diagonal structure Σ^H such that variables within each block are highly correlated but variables across blocks are nearly uncorrelated. Thus, if we can identify an approximate block diagonal structure $\hat{\Sigma}^H$ in the correlation matrix of observed data set X_t, then we can inductively infer empirical classification in the form of $X_t^k = \Lambda_f^k \cdot F_t^k + e_t^k$. Given that the standard static correlation $Corr(X)$ only measures synchronous or contemporaneous co-movement among variables, the dynamic correlation $DynCorr(X)$ is also used to measure co-movement among observed variables.

$$\Sigma = Corr(X) \text{ or } DynCorr(X)$$

$$= \Sigma^H = \begin{bmatrix} \Sigma_{11}^H & 0 & \cdots & 0 \\ 0 & \Sigma_{22}^H & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_{KK}^H \end{bmatrix}, \text{ where } \Sigma_{kk}^H = \begin{bmatrix} 1 & \rho_k & \cdots & \rho_k \\ \rho_k & 1 & \cdots & \rho_k \\ \vdots & \vdots & \ddots & \vdots \\ \rho_k & \rho_k & \cdots & 1 \end{bmatrix}$$

$$\approx \hat{\Sigma}^H = \begin{bmatrix} \hat{\Sigma}_{11} & 0 & \cdots & 0 \\ 0 & \hat{\Sigma}_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\Sigma}_{KK} \end{bmatrix}, \text{ where } \hat{\Sigma}_{kk}^H = \begin{bmatrix} 1 & \rho_{k,12} & \cdots & \rho_{k,1N} \\ \rho_{k,21} & 1 & \cdots & \rho_{k,2N} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k,N1} & \rho_{k,N2} & \cdots & 1 \end{bmatrix}, \ \forall k = 1,.....,K .$$

The dynamic correlation is proposed from the frequency domain framework and defined as follows: $\rho_{xy}(\lambda) = \dfrac{C_{xy}(\lambda)}{\sqrt{S_x(\lambda) \cdot S_y(\lambda)}}$ for the frequency λ and $\rho_{xy}(\Lambda) = \dfrac{\int_\Lambda C_{xy}(\lambda) d\lambda}{\sqrt{\int_\Lambda S_x(\lambda) d\lambda \cdot \int_\Lambda S_y(\lambda) d\lambda}}$ for the frequency band $\Lambda = [\lambda_1, \lambda_2)$ where $-\pi \le \lambda \le \pi$, $0 \le \lambda_1 < \lambda_2 \le \pi$, x and y are two zero-mean real stochastic processes, $S_x(\lambda)$ and $S_x(\lambda)$ are the spectral density functions of x and y, and $C_{xy}(\lambda)$ is the co-spectrum. The dynamic correlation has useful properties such as: (a) The dynamic correlation measures different degrees of co-movement which varies between -1 and 1 just as standard static correlation. (b) The dynamic correlation over the entire frequency band is identical to static correlation after suitable pre-filtering and it is also related to stochastic co-integration. (c) The dynamic correlation can be decomposed by frequency and frequency band, where the low or high frequency band in spectral domain have implication for the long-run or short-run in time domain respectively (Croux, Forni, and Reichlin, 2001). Note that Forni et al

(2000) also propose the dynamic factor model based on the spectral density matrix in a frequency-domain framework due to the similar issue of standard correlation or covariance matrix. We use the standard static correlation as well as the dynamic correlation defined to measure the close co-movements of disaggregated variables within a group and near independences of disaggregated variables across groups.

The use of aggregate variables or estimated factors to study dynamic relationships and to infer causal relationships among observed disaggregate variables can be theoretically justified based on the compositional stability condition derived from the aggregation theory as discussed in chapter II. The identified block diagonal pattern of correlation matrixes implies that the observed disaggregate variables approximately satisfy the consistent aggregation condition of compositional stability condition. This condition in turn implies that there exists not only the possibility of obtaining interpretable macro-variables as the representative aggregate of homogeneous disaggregate micro-variables, but also the possibility of yielding interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for the subsequent analysis. This means that when the disaggregate variables can be legitimately grouped and represented by aggregate variables, it is possible to use aggregate variables to capture (causal) relationship among disaggregate variables through the (causal) relationship among aggregate variables as the legitimate representatives as long as the compositional stability conditions hold among disaggregate variables. Moreover, given that the VAR approach is proposed and used as an inductive method as an alternative to the deductive SEM approach, it is better to pursue inductive methods, where the classification/aggregation and causality issues are addressed based on the inductively inferred information from the data itself, rather than based on the maintained assumptions derived from deduction and/or researchers' intuition.

Summary and Proposed Method

There are significant advances in macro-econometric study from the methodological and empirical perspective. In methodological perspective, the vector autoregressive (VAR) model approach is proposed and used as an alternative to the structural equation model (SEM) approach. Given that the SEM approach requires too much causal information for the identification problem, the VAR approach provide the possibility of inferring causal information from statistical properties of the data without pretending to have too much *a priori* theory and/or without demanding too much information from the data. Given that such possibility to inductively infer the causal structure of the VAR approach, compared to the SEM approach, is not fully used within the full recursive causal assumption, the use of the graphical causal model approach is proposed to address the remaining issue of how to inductively infer the causal structure to relate empirical regularities captured in reduced form model to theoretical properties represented by the structural form model. On the other hand, recent advances in data processing capabilities have brought the possibility of analyzing larger number of detailed variables. The macro-economic panel data have brought forth research potentials for significant advances in the macro-economic analysis of monetary policy effects. The factor augmented VAR (FAVAR) approach is proposed to use such research potentials and to address informational issue in the small size VAR approach. For the full use of the inductive possibility of structural understanding of macro-economy, the use of the approximate form of the compositional stability condition is proposed. This method provides inductive classification of macro-economic panel data and thus makes it possible to obtain meaningfully interpretable estimated factors, which in turn allow the use of the graphical causal model for the FAVAR approach.

Given the observation that many variables in this high dimensional data move very closely, the compositional stability condition as the consistent aggregation condition provides an inductive way to pursue the possibility of obtaining not only (a) interpretable aggregate macro-variables as the representative aggregate of homogeneous disaggregate micro-variables but also (b) interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for the subsequence analysis. This implies that when the micro-variables can be legitimately grouped and represented by macro-variables, it is possible to use aggregation methods (a) to incorporate broad range of information into the empirical models with minimizing econometric issues such as the multicollinearity and degrees of freedom, (b) to capture (causal) relationships among disaggregated variables through (causal) relationships among aggregated variables as the legitimate representatives. This compositional stability condition is used (a) to provide an inductive way of forming suitable partitions before conducting any empirical test to justify those classifications based on the empirical data patterns rather than on researchers' intuition and (b) to address the possible violation of the (probabilistic) stability condition to use the graphical causal models for the high dimensional data. Note that it is conceivable and oftentimes observed that the (probabilistic) stability condition for the graphical causal models is violated for using high dimensional data in empirical study, given the observation that there exist close co-movements and thus near deterministic relations among variables in high dimensional data.

More specific procedure we propose is as follows: (a) Both standard static correlation matrix and dynamic correlation matrix over identified frequency band are used to measure co-movement among original variables. Based on these similarity measure of disaggregate micro-variables, the modified k-nearest neighbor algorithm is used to sort the variables such that the highly correlated variables are near each other along the main diagonal in reordered correlation matrix. The block-diagonal pattern of reordered or sorted static and dynamic correlation matrixes are used to identify homogeneous group of variables, based the approximate form of the compositional stability condition. (b) Based on identified classifications of original variables, the statistical dimensional reduction method are used for actual aggregation procedure to decide weighting schemes for aggregating disaggregated micro-variables into representative macro-variables within each identified group. The principal component method applied onto each of groups is used as the best dimensional reduction method with as little loss of information as possible in the mean squared error sense. (c) Given that the inference based on the small size VAR can be misleading unless the reduced form innovations span the space of the structural shocks or the VAR model does not have the omitted variables problem, the estimated factors are augmented in the VAR (FAVAR) framework to increase the amount of information in the empirical model so that the reduced form residuals span the space of the structural economic shocks. (d) Based on the residuals of reduced form FAVAR, the contemporaneous causal structure among innovations is inferred by the graphical causal model. The identified compositional stability condition in the data makes it possible to infer causal structures among micro-variables through relationships among representative aggregated macro-variables. The PC algorithm or GES algorithm is used to infer causal structures among macro-variables as the legitimate representative causal relationships among micro-variables for the subsequent analysis. (e) Based on the contemporaneous causal structure used for identification of FAVAR, structural relationships of the macro-economy are studied in the two types of the moving average representations. The impulse response functions of all the observed variables with respect to shocks in the monetary policy variable as well as each of the estimated factors are estimated and interpreted. The forecast error variance in each factor is decomposed into the parts attributable to each of a set of innovations processes in the FAVAR. Note that inductive properties are emphasized in every sequence of the proposed method, since any types of deductive properties can bring subjectivities or ambiguities into the empirical results. The proposed method is illustrated with the applications for retail checkout scanner data as an example of the high dimensional data.

Empirical Analysis and Results

The proposed methodological procedure is illustrated with U.S. macro-economic panel data. Given that the vector autoregressive (VAR) model approach is proposed and used as an alternative to the deductive structural equation model (SEM) approach, inductive properties are emphasized in every step of empirical procedures. First, the data used for this study are described. Second, based on the identified common frequency for the estimated spectrum of variables in the data set, static and dynamic correlations among variables are measured. Third, based on the block diagonal pattern of the correlation matrixes identified by the modified k-nearest neighbor algorithm, the variables are classified and classified groups are interpreted, where variables within each group move together closely. Fourth, based on the classified groups, the latent factors are estimated and augmented in the VAR (FAVAR) framework. Based on the residuals of reduced form FAVAR, the contemporaneous causal structure among innovations is inferred by the graphical causal model. Fifth, based on the causal structure used for identification, the impulse response functions with respect to shocks in the monetary policy variable as well as each of the estimated factors are estimated and interpreted. The forecast error variance in each factor is decomposed into the parts attributable to each of a set of innovations processes in the FAVAR. The empirical results are summarized and further issues to be studied are discussed.

Data Description

The data set consists of monthly observations on 103 U.S. macro-economic time series panel data from 1959:1 through 2003:12 with the sample size of 526. All the data are from the data set used in Stock and Watson (2005). According to these authors, all series are from the Global Insights Basic Economics Database, the Conference Boards' indicators Database, and their own calculations. The data represent a broad range of macro-economic activity. Stock and Watson intuitively grouped the time series variables in the data set as following categories: 1. Real output and income, 2. Employment and hours, 3. Real retail, manufacturing and trade sales, 3. Consumption, 4. Housing starts and sales, 5. Real inventories, 6. Orders, 7. Stock prices, 8.

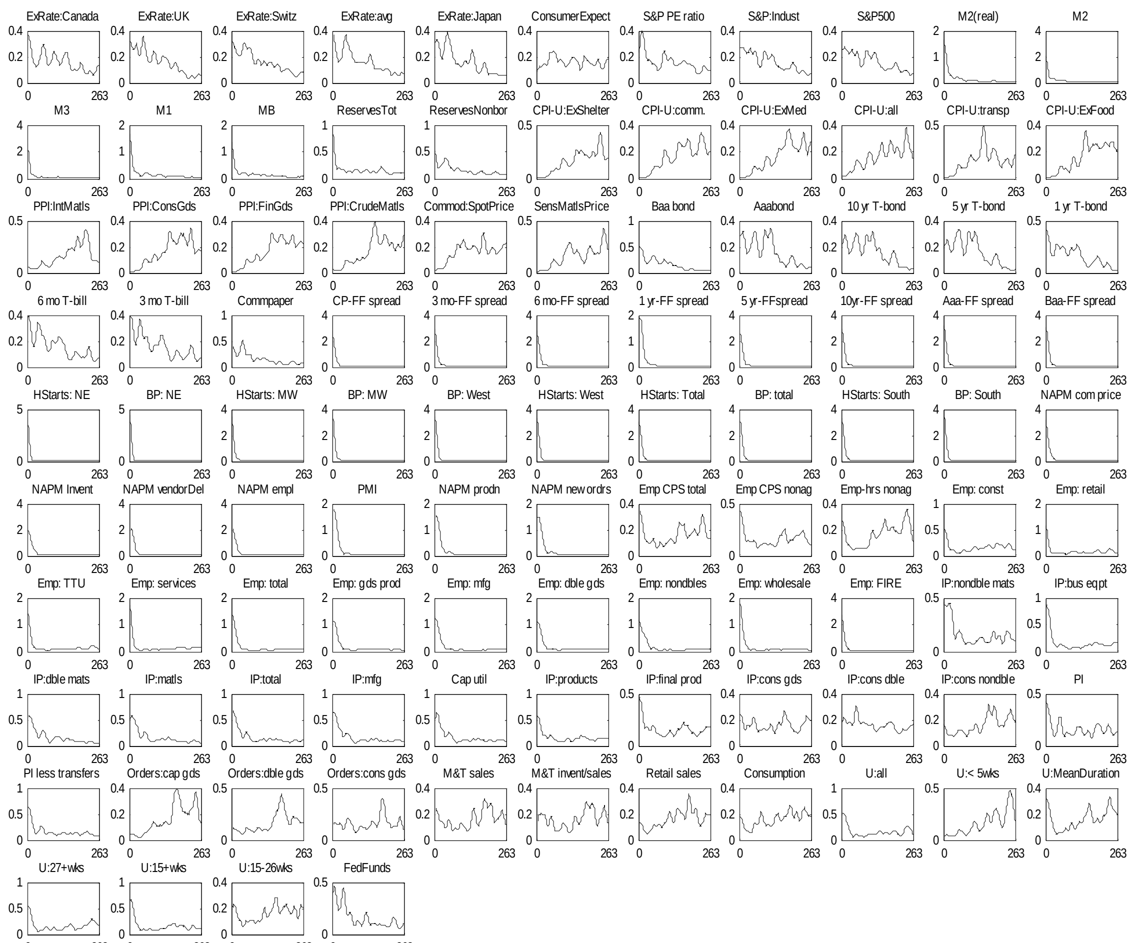
Exchange rates, 9. Interest rates, 10. Spreads, 11. Money aggregates, 12. Price indexes, and 13. Miscellaneous.

The data are transformed in four ways. First, many of the series are seasonally adjusted by the reporting agency. Second, the series are transformed by taking logarithms and/or differencing so that the transformed series are approximately stationary. In general, the first difference of logarithms (growth rates) is used for real variables, the second difference of logarithms (changes in growth rates) is used for price series, and the first differences are used for nominal interest rates. Third, outliers contained in some of the transformed series are identified as absolute median deviations larger than 6 times the inter quartile range and adjusted by replacing those observations with the one-sided median value of the preceding 5 observations. Fourth, the series are demeaned and standardized (Stock and Watson, 2005). The list of variables with detailed descriptions and their transformations are given in Appendix E. The grouping and ordering of the variables are based on the empirical results of this study.

Classification and Aggregation

One of objectives of this study is to propose an inductive procedure for the construction of appropriate grouping of variables. Given that theory does not provide sufficient and conclusive information for classification, an inductive property is emphasized due to the empirical implausibility of attempting all possible partitions. In this respect, it is better to pursue inductive classifications related with legitimate aggregation conditions, which is based on the empirical data pattern itself rather than researchers' subjective intuition. Based on the compositional stability conditions derived from the aggregation theory, our inductive procedure is based on the idea that homogeneity or similarity of group of variables can be identified through their dynamic movements. When original disaggregate variables within a group have the similar dynamic movements so that they co-move each other very closely, their high co-movements reveal their underlying similarity.

Given that the standard static correlation only measures synchronous or contemporaneous co-movements between variables and it is desirable to allow possible leads and/or lags in dependency among the time-series data in dynamic setting, both the standard static correlation matrix and the dynamic correlation matrices estimated over identified frequency bands are used to measure co-movement among the original variables. For the dynamic correlations, several different frequency bands are chosen as the non-overlapping bands or regions, based on the estimated spectrums of all the time series variables in the data. They approximately centered at peak λ_k so that $\{\ \Lambda = \lfloor \lambda_i, \lambda_j) \cup \lfloor -\lambda_j, -\lambda_i) : 0 \le \lambda_i < \lambda_k < \lambda_j \le \pi\ \}$, where the frequency λ_k is specified as $\{\ \lambda_k = 2\pi \cdot k/T : k = 1, \cdots, (T/2)\ \}$ and T is the sample size (Rodrigues, 1999). Note that if the frequency of a cycle is λ, the period of the cycle is $2\pi/\lambda$. Thus, a frequency of $\lambda_k = 2\pi \cdot k/T$ corresponds to a period of $2\pi/\lambda_k = T/k$. We choose common frequency bands to measure co-movement among variables with possible leads and lags, based on the estimated spectrums of variables, which capture dynamics of variables in terms of their cyclic properties with long or short run trends (Hamilton, 1994). The estimated spectrums of all the time series variables are presented in Figure 4.1.



* See Appendix E for the description of variables, where variables are in the same order.
* The x- axis is the frequency in terms of k and the y-axis is the estimated spectrum.

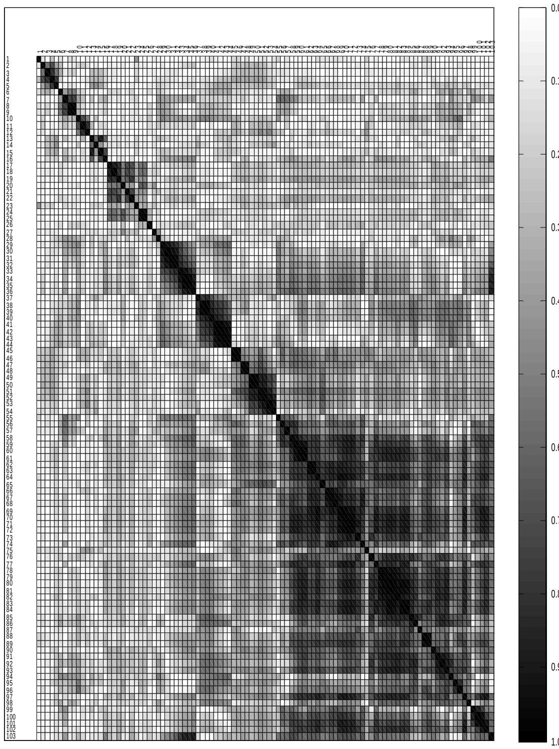
Figure 4.1. Estimated Spectrums of Macroeconomic Variables

The x-axis is the frequency in terms of k and the y-axis is the estimated spectrum. We use five frequency bands: 0-30, 31-80, 81-160, 161-220, and 221-263 in terms of k, which correspond to period more than 17.53 months (frequency Band 01), period of 16.97 to 6.58 months (frequency Band 02), period of 6.49 to 3.29 months (frequency Band 03), period of 3.27 to 2.39 months (frequency Band 04), and period less than 2.38 months (frequency Band 05) ranges respectively. These ranges approximately correspond to 2.5 year and 12, 6, and 3 months and short period ranges, where the 2.5 year is known as a business cycle frequency, given dates of the economic recessions (Hamilton, 1994).

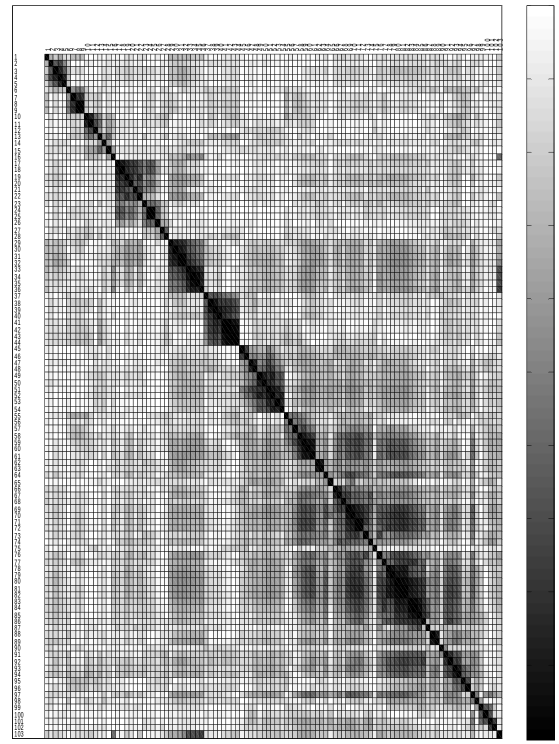
Based on the similarity measures of disaggregate micro-variables, the modified k-nearest neighbor algorithm is used to sort or reordered the variables such that the highly correlated variables are near each other along the main diagonal in the final correlation matrix. The final result of the sorted static correlation matrix and dynamic correlation matrixes for different frequency bands are presented in Figure 4.2. The black/white color scheme is used to represent the absolute value of measured correlations, where the darkest black represents the correlation of 1 and the brightest white represents the correlation of 0. The sorted static correlation matrix with the color scheme is presented the Appendix F. Note that the frequency Band 00 is the entire frequency region. It is demonstrated that dynamic correlation over entire frequency band is equivalent to the static correlation of pre-filtered data, where the following two-sided filter is used: $A_\Lambda(L) = \dfrac{\lambda_2 - \lambda_1}{\pi} + \sum_{k=1}^{\infty} \dfrac{\sin k\lambda_2 - \sin k\lambda_1}{k\pi}\left(L^k + L^{-k}\right)$, where L is lag operator and $\Lambda = \Lambda_+ \cup \Lambda_-$. This dynamic correlation of the entire frequency band represents the idea, similar to that used in correlations of band-pass filtered data, that the synchronic cyclical components of variables can be measured by looking at the correlation over the extracted cycles from the variables (Croux, Forni, and Reichlin, 2001).

The main feature of the compositional stability condition is that each aggregate variable is composed of grouped disaggregate variables with a "stable" compositional factor over time, so the ratios of disaggregate variables over aggregate variables are near constant and stable over time. In this respect, when we can identify that the correlation matrixes of observed data set X_t have the special block diagonal structure such that variables within each block are highly correlated but variables across blocks are nearly uncorrelated, we can inductively infer an empirical classification and thus we can use the block form of factor model of $X_t^i = \Lambda_f^i \cdot F_t^i + e_t^i$

Frequency Band 01



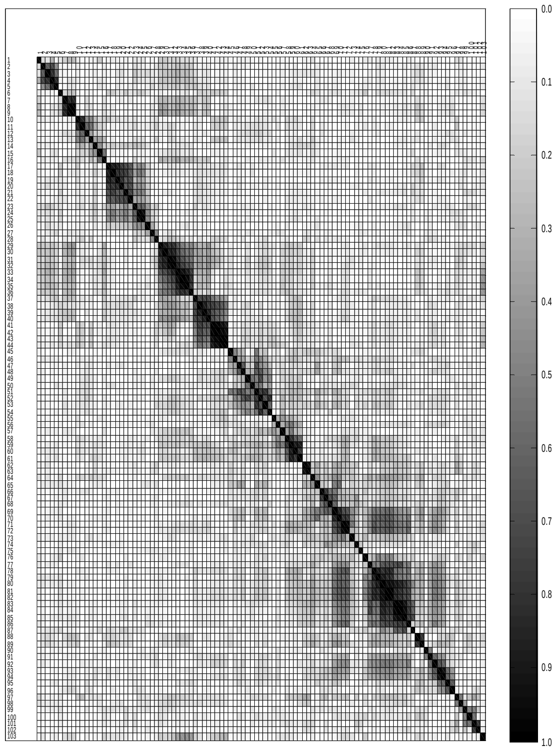
Frequency Band 02



* The black/white color scheme is used to represent the absolute value of measured correlation, where the darkest black represents the correlation of 1 and the brightest white represents the correlation of 0.
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.2. Sorted Static and Dynamic Correlation Matrix of Macroeconomic Variables

Frequency Band 03



Frequency Band 04

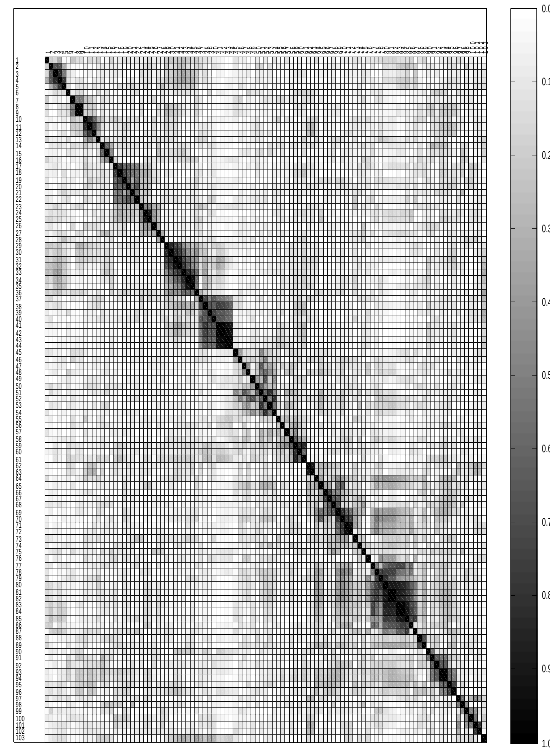
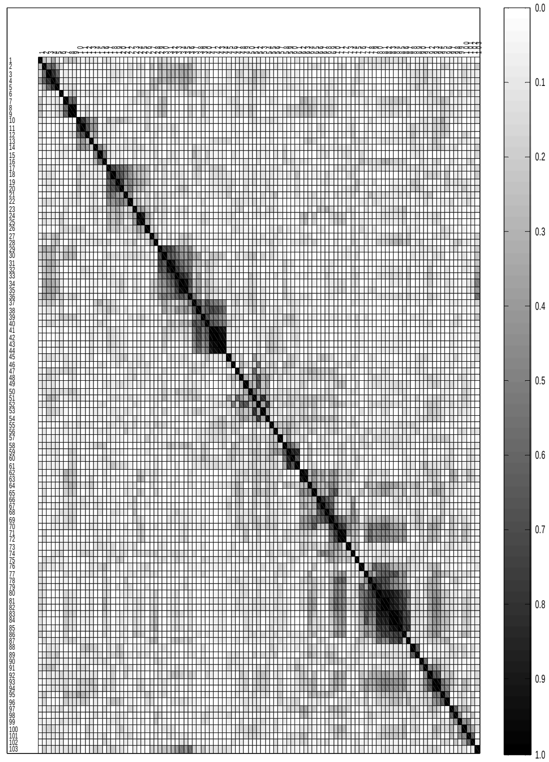


Figure 4.2. (Continued)

Frequency Band 05



Frequency Band 00

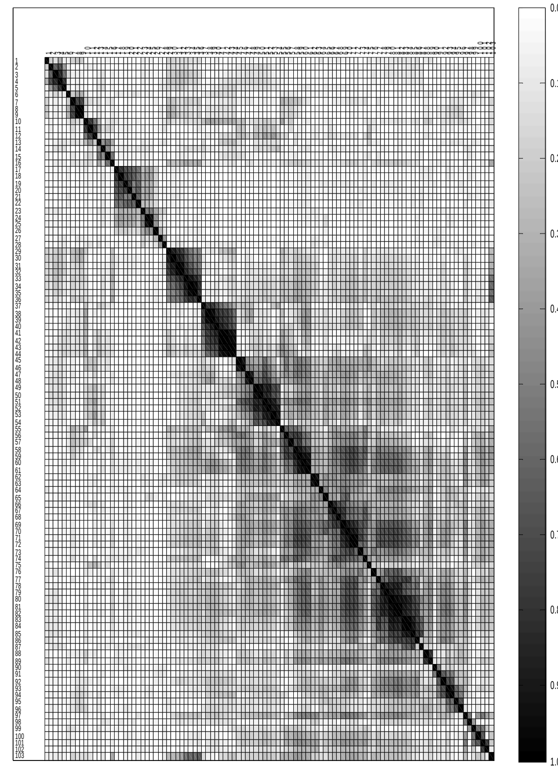


Figure 4.2. (Continued)

Standard Static Correlation Matrix

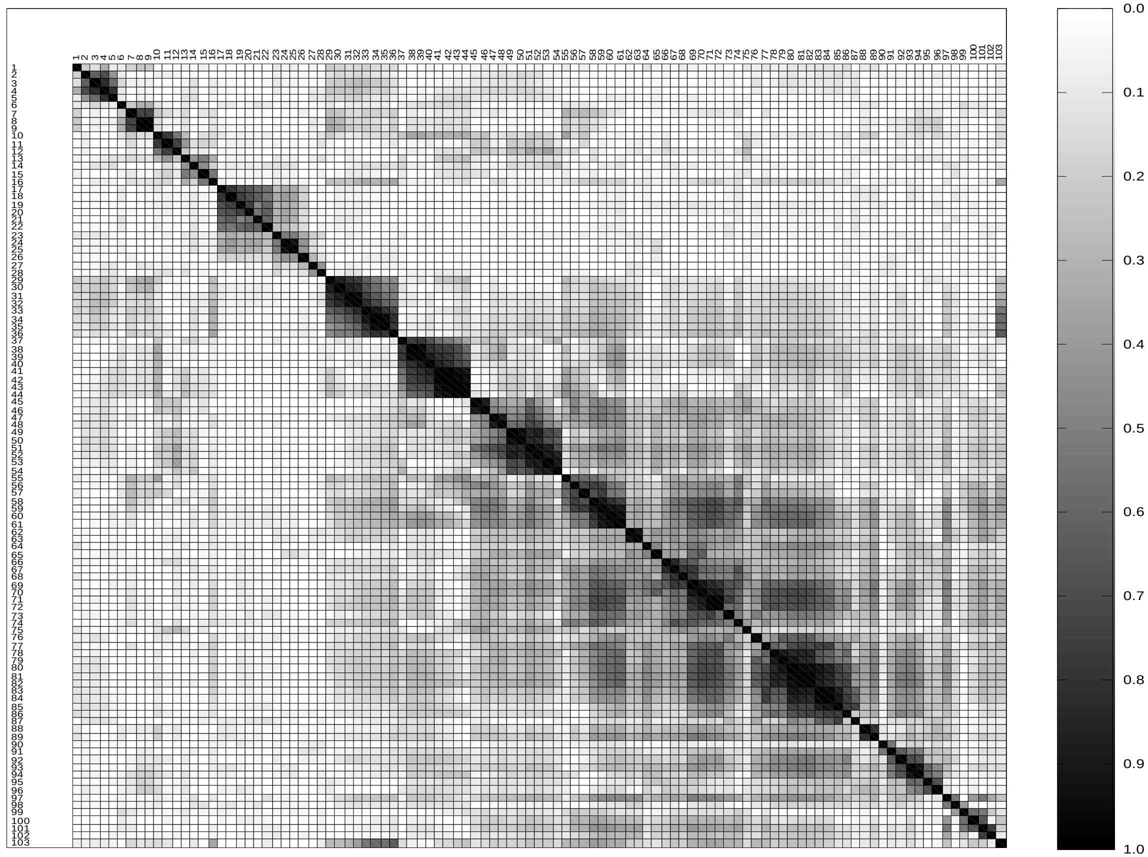


Figure 4.2. (Continued)

where $i = 1, \cdots, I$ is the classified group index. When data reveals this special block diagonal structure, extracting the estimated factors from each block of variables, rather than obtaining the estimated factors from the entire data set, can provide better representative aggregates with clear interpretations for each aggregate variables based on the grouped disaggregate variables. In Figure 4.2 of the sorted static and dynamic correlation matrixes, we can identify that this special block diagonal structure commonly exists over all the different frequency bands and static correlation matrix, although the correlations of pair-wise variables across different groups show somewhat different degrees of correlation over the different frequency bands.

Based on the sorted static correlation matrix and the dynamic correlation matrixes over the different frequency bands, the following groups of macro-economic variables are identified as homogeneous groups, which are commonly identified in both the static correlation matrix and the dynamic correlation matrix over the different ranges of frequency bands.

Exchange Rate Variable Group: Variables of 001 to 005.

Several foreign exchange rates for different countries such as Canada, Japan
with average foreign exchange rate

Stock Market Variable Group: Variables of 006 to 009

Several S&P composite stock price indexes with S&P composite,
stock price-earning ratio, and consumer expectation index variables

Money Aggregate Variable Group: Variables of 010 to 016

Several monetary stock indexes such as M1, M2, M3
with money supply and several deposits, bank reserves variables

Price Variable Group: Variables of 017 to 028

Several consumer and producer price indexes
with spot market price and sensitive materials price indexes

Interest Rate Variable Group: Variables of 029 to 036

Several interest rates of different maturities
with several bond yields and commercial paper rate

Spread Variable Group: Variables of 037 to 044

Several spread between federal fund rates
with interest rates variables of different maturities included above group

Housing Market Variable Group: Variables of 045 to 054

Several variables on housing starts and houses permitted of total and different regions such as northeast, south areas.

NAPM Variable Group: Variables of 055 to 061

Several National Association of Purchasing Management (NAPM) indexes such as production, new order indexes.

Employment Variable Group: Variables of 062 to 075

Several employment on non-farm payrolls of total and different areas with employed labor force variables.

Output Variable Group: Variables of 076 to 089

Several industrial production indexes of total and different areas with personal income variables and capacity utilization variable.

Consumption/Investment Variable Group: Variables of 090 to 096

Several Manufacturers' new order of different sectors and sales of different sectors with consumption variables

Unemployment Variable Group: Variables of 097 to 102

Several unemployment variables of different durations with total unemployment rate variable.

Federal Funds Rate Variable: Variable 103

the effective federal funds rate.

The complete variable names and their detailed description for each group is given in Appendix E, where variables are grouped and in the same order in the sorted correlation matrix. While this classification result has its own interpretations for each group of variables in terms of corresponding macro-economic theoretical variables, this classification is the inductive one using the empirical data itself based on the following observed patterns. First, the different degrees of correlation across identified groups are observed. The correlations across group in the long run period ranges (frequency bands 01 and 02) are relatively high, compared with those in the short run period ranges (frequency bands 04 and 05). This correlation pattern across different groups can be interpreted based on the fact that each frequency band represents a different cyclic period. As the dynamic correlation matrix is based on the more long run range of period, it measures more long run relationships among variables. And the relationships among variables generally increase as they are measured in the longer period range, when there are certain stability or equilibrium relationships among variables. In this respect, the close co-movement among variables is expected more in the long run range period dynamic correlation matrix than in the short run range one. Second, although the correlations of pair-wise variables across different groups show somewhat different degrees of correlation over the different frequency bands, the common groups of variables are identified over all the different frequency bands. The Exchange Rate, Stock Market, Money Aggregate, Price, Interest Rate, Spread, and Unemployment variable groups are distinct homogenous groups. The money group variables are homogenous especially in the frequency bands of 02 and 03, although they are somewhat separated as the monetary aggregate variables and the reserve variables. The price group variables are homogenous especially in the frequency bands of 01, 02, and 03, although they are somewhat separated as the CPI variables and the PPI and commodity price variables. Third, the degrees of correlations among the variables in the NAPM, employment, output, and consumption/investment groups are high, especially when dynamic correlations are measured in longer period range (frequency bands 01 and 02) rather than shorter period range (frequency bands 04 and 05). The Housing Market group and NAPM group variables are discriminated by their relatively different relationships with variables in the Employment and Output groups, given that the variables in NAPM group have higher correlation with the variables in Employment and Output groups. The NAPM group and Employment group variables are discriminated by their relatively different relationships with variables in the Housing Market and Output groups, given that the variables in NAPM group have higher correlation with the variables in Housing Market group, whereas the variables Employment have higher correlation with the variables in Output groups. The Employment group and Output group variables are discriminated by their relatively different relationships with variables in the NAPM and Consumption/Investment groups, given that the variables in Employment group have higher correlation with the variables in NAPM (and Housing Market) group, whereas the variables Output have higher correlation with the variables in Consumption/Investment groups. The Output group and Consumption/Investment group variables are discriminated by their relatively different relationships with variables in the Employment (and NAPM) group, given that the variables in Output group have higher correlation with the variables in Employment (and NAPM) group.

This classification result can be interpreted in the context of monetary policy transmission mechanism literature. In demand side of economy, overall size of economy (output group) consists of consumption and investment (consumption/investment group). On the other hand, total labor force can be divided into employment and unemployment components. The money and price groups represent two important components affecting real economic activities. The interest rate group of variables corresponds to narrow interest rate channel of the monetary transmission channel. The exchange rate group of variables can be understood based on the traditional ISLM macro-models for the open economy. The stock market and housing market groups can approximately represent the corresponding asset market channels in the transmission mechanisms, given that stock and house represent financial and physical assets respectively. The NAPM or spread groups can approximately represent the expectation channel suggested in some monetary transmission mechanism literature.

The resulting classification based on the inductive procedure can be compared with other deductive classifications, which rely on the researchers' intuitive choices. For example, Leeper, Sims, and Zha (1996) implicitly classify macro-economic variables into real gross domestic product, real private non-residential fixed investment, and real residential fixed investment with some selected variables such as unemployment, several monetary aggregates, several interest rates, several price indexes, exchange rate in their Bayesian structural VAR model. Their classification has some distinctive features. (a) The real product group consists of total industrial production, employment, retail sales, personal consumption, and NAPM indexes. (b) The non-residential investment group consists of several variables related with industrial structures, equipment component and manufacturers' shipments to capital goods industries. (c) The residential investment group consists of variables related with housing starts and construction. (d) They individually select several similar variables. For example, M1 and total reserve variables are selected individually, not aggregated. Note that their aggregation is based on the Chow-Lin procedure, where national income and product accounts quarterly series are combined with each group of monthly time series variables. For another example, Belviso and Milani (2005) explicitly classify macro-economic variables into real activity, inflation, interest rates, financial market, money, credit, expectation groups for their structural factor augmented VAR model (SFAVR). In their classification, (a) The real activity group consists of almost all the variables except variables included in other groups. (b) The expectation group consists of NAPM group variables and spread group variables. (c) The financial market represents the stock market.

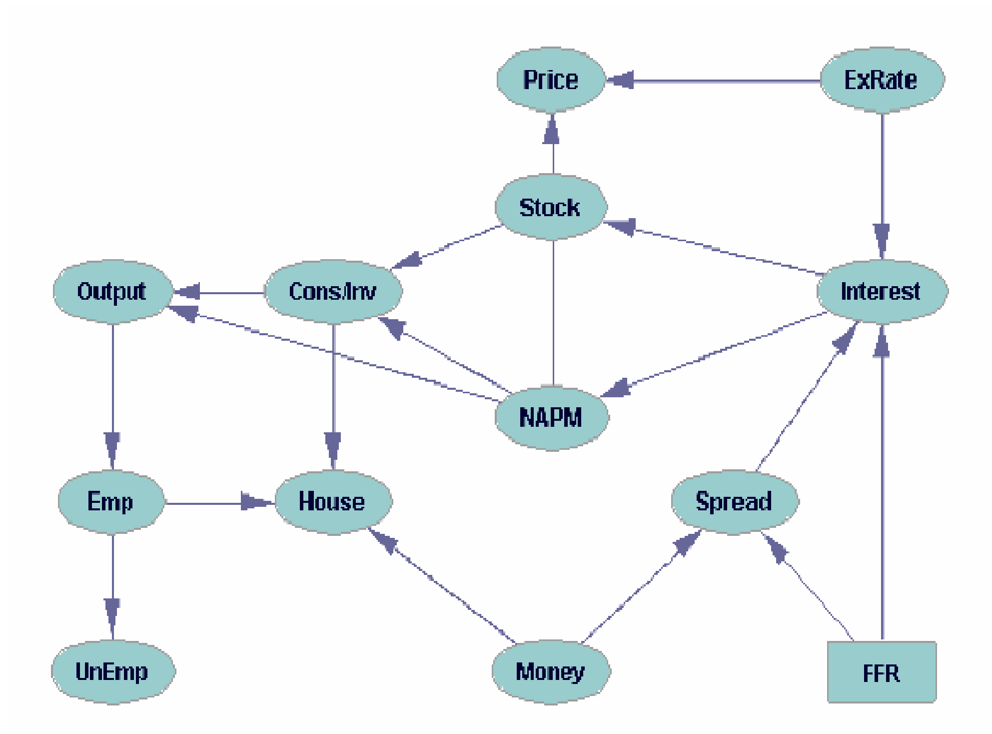
Comparing with two classifications mentioned above and other implicitly suggested deductive classifications, an inductive classification of this study has following distinctive features: (a) The house, consumption/investment, employment, unemployment, and production groups are separately identified. This separation can be observed in other empirical studies. For example, Leeper, Sims, and Zha (1996) separate real activity group of variables into real gross domestic product, real private non-residential fixed investment, and real residential fixed investment groups of variables in their empirical study. The non-residential investment group variables approximately correspond to the invest component of consumption/investment group variables. On the other hand, the residential investment group variables approximately correspond to the house group variables. The consumption or sale related variables are classified as consumption/investment group with investment related variables. The employment group is separated from the output group, since the employment group shows higher degrees of correlation with NAPM and housing market groups than the output group as discussed. (b) The spread group is separated from the NAPM group. Although Belviso and Milani (2005) identify these two groups as one homogeneous group, their explanations for their expectation group provide clues for interpret this empirically found separation. The NAPM surveys indexes are relatively more related with expectations about real activity such as production, employment, inventories, and new orders. On the other hand, the interest rate spreads are relatively more related with expectations about the future short-term rates and future inflations. Note that Leeper, Sims, and Zha (1996) include NAPM group variables into the real gross domestic product group variables.

Causality for Identification

Based on the classification results, a five step procedure is used to study monetary policy effects. First, the latent factors are estimated from the observed variables based on the principal component method. Given that dynamic factor approach is based on the proposition that the observed co-movements of variables are produced by the underlying common dynamic factors, which are in turn driven by underlying common structural economic shocks, the block diagonal pattern of static and dynamic correlation matrixes imply that (i) Co-movements among variables exist within blocks rather than across blocks. (ii) As common sources of comovement, there can be each of dynamic factors common for each specific block rather than for the entire data set. (iii) The underlying structural economic shocks can be revealed by estimating each of common sources of comovement from each block rather than from the entire data set. Based on these reasoning, it can be better to estimate each latent factor from each block rather than to estimate factors from the entire data set. And thus each latent factor $\hat{F}_t^i$ is estimated from each block of variables in the block form of factor model framework of $X_t^i = \Lambda_f^i \cdot F_t^i + e_t^i$, where $i = 1, \cdots, I$ is index for the classified groups. Second, estimate reduced form VAR augmented with estimated factors (FAVAR) and obtain the covariance among innovations from the estimated reduced form FAVAR $Cov(u_t) = \hat{\Omega}$. Given the monthly data is used, 13 lags are used to incorporate sufficient dynamics into model following Bernanke, Boivin, and Eliasz (2005). Third, based on the system of equations $\hat{\Omega} = A_0^{-1}(A_0^{-1})^T$, the unknown elements in A_0 coefficient matrix are solved or recovered in terms of the estimated elements of $Cov(u_t) = \hat{\Omega}$ covariance matrix. Since there are $N \cdot (N+1)/2$ equations in $Cov(u_t) = \hat{\Omega}$ and N^2 unknown parameters in A_0, at least $N \cdot (N-1)/2$ restrictions in A_0 need to be imposed for the existence of a solution for A_0. The contemporaneous coefficient matrix A_0, which relates the structural and reduced form VAR specifications, specifies how variables are causally linked to each other contemporaneously.

The causal information (in the form of restriction on A_0 matrix) required for identification in the FAVAR framework can be inductively inferred from data based on the graphical causal models or the DAG approach. The use of aggregate variables or estimated factors to study dynamic relationships and to infer causal relationships among observed disaggregate variables can be justified based on the compositional stability condition derived from the aggregation theory. The identified block diagonal pattern of correlation matrixes discussed in aggregation section implies that the observed disaggregate variables approximately satisfy the consistent aggregation condition of compositional stability condition. This condition in turn implies that there exists not only the possibility of obtaining interpretable macro-variables as the representative aggregate of homogeneous disaggregate micro-variables but also the possibility of getting interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for the subsequence analysis. This means that when the disaggregate variables can be legitimately grouped and represented by aggregate variables, it is possible to use aggregate variables to capture (causal) relationship among disaggregate variables through (causal) relationship among aggregate variables as the legitimate representatives as long as the compositional stability conditions hold among disaggregate variables. Note that in the preliminary study for causal structures in the disaggregated original level data set, many reasonable causal relationships among disaggregate micro-variables are not statistically observed. It is because high correlation among x_1 and x_2 can induce $P(x_1 \mid x_2, x_3) = P(x_1 \mid x_2)$ through $P(x_1 \mid x_2) \approx 1$ regardless of the causal structures among them. So it is not possible to infer reliable causal structure from the probabilistic dependency pattern. The (probabilistic) stability condition of the graphical causal model is violated and thus DAG method can not be legitimately used for disaggregate level data set. Note that this problem is similar to the multicollinearity problem, which makes it difficult to obtain precise estimates of the separate effects of the variables in the regression method. The GES algorithm is used to infer contemporaneous causal structures among innovations of FAVAR as the legitimate representative causal relationships among observed disaggregate variables. Note that the PC algorithm results in several undecided causal orientations in the similar non-spurious statistical dependencies (skeleton) with the GES algorithm and thus only the result of GES algorithm is used in this study. The inductively inferred contemporaneous causal structure by the GES algorithm is presented in Figure 4.3. The covariance\correlation matrix among innovations of reduced Form FAVAR is presented in Table 4.1.

The contemporaneous causal structure, which is inductively inferred by the GES algorithm without any deductive information, can be interpreted as follows. (a) There is observational equivalence between stock market innovations and NAPM innovations. This means that the causal direction can not be decided based on statistical observations only or either direction between them is statistically equivalent (Chi-Square(59) value is 68 with the significant level of 0.2082 for the likelihood ratio test of both over-identifications). Empirical results based on the causal direction from stock to NAPM are presented, given that the empirical results for the subsequent analyses including impulse responses are not sensitive to either orientation. (b) There are several first causes (causal roots) and last effects (causal sinks). The federal fund rate variable and monetary aggregate and exchange rate factors turn out to be causal root innovations. On the other hand, the price, unemployment and housing factors turn out to be causal sink innovations. Note that the observed policy variable represented by the federal fund rate is not causally ordered last, as Bernanke, Boivin, and Eliasz (2005) assumed. This result is also found in the contemporaneous causal structure inferred by the PC algorithm.



* See Appendix E for the description of representative aggregates

Figure 4.3. Contemporaneous Causal Structure Inferred by GES Algorithm

Table 4.1. Covariance\Correlation Matrix among Innovations of Reduced Form FAVAR

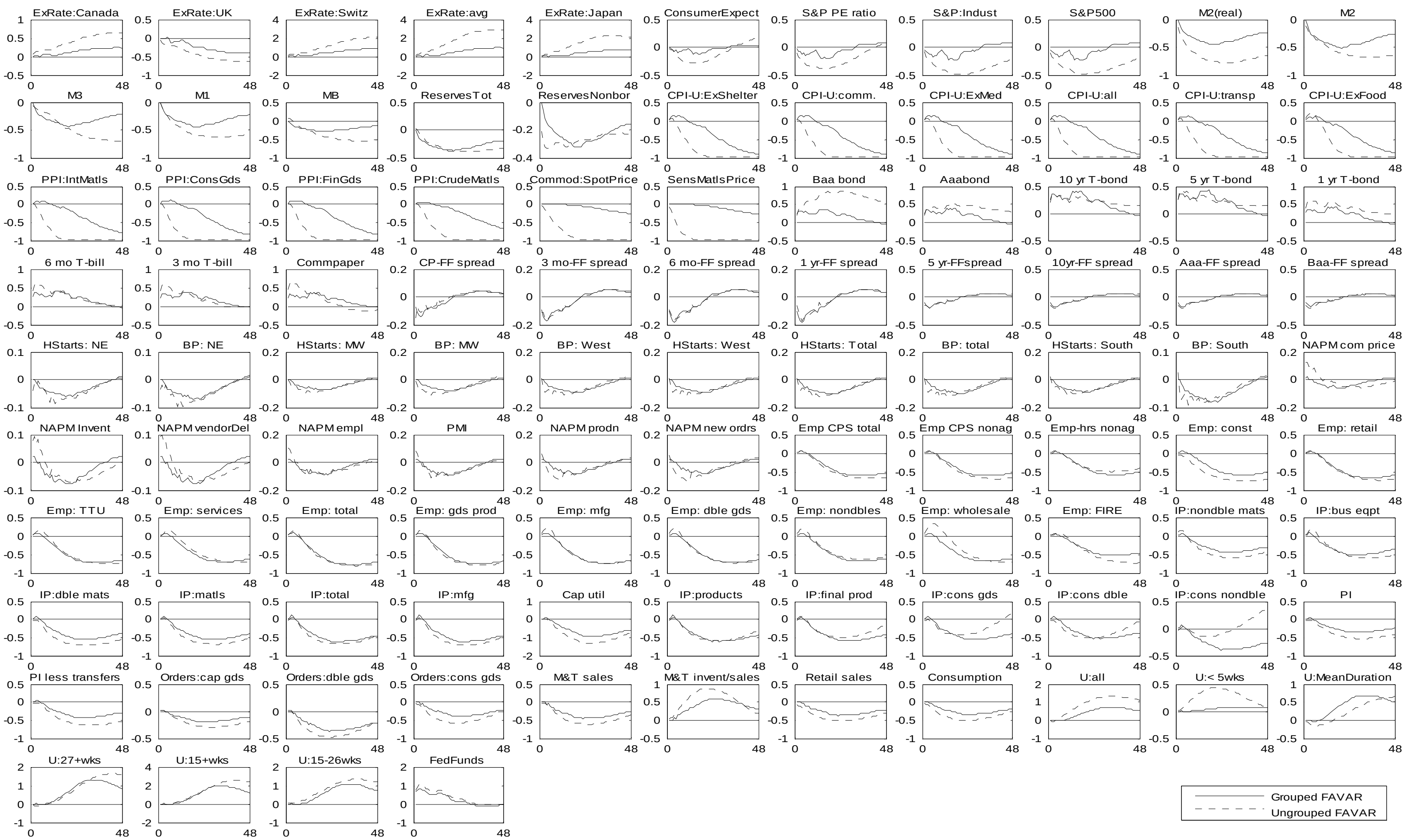
	ExRate	Stock	Money	Price	Interest	Spread	House	NAPM	Emp	Output	Cons/Inv	UnEmp	FFR
ExRate	0.31773	−0.02079	−0.02540	−0.11210	0.19158	−0.01092	−0.07857	0.03551	−0.00131	0.07946	0.07082	−0.02903	0.06757
Stock	−0.00731	0.38890	−0.02901	−0.12784	−0.13024	−0.04496	0.03959	0.15225	0.03544	0.02279	0.16917	0.04482	−0.02200
Money	−0.00550	−0.00694	0.14735	−0.02526	−0.10232	−0.11504	0.10809	−0.00558	0.01903	−0.01099	−0.07206	−0.00771	0.00281
Price	−0.02635	−0.03325	−0.00404	0.17390	0.06807	0.08844	0.02109	−0.01037	−0.01320	−0.04479	−0.07110	0.05640	−0.03041
Interest	0.05789	−0.04355	−0.02106	0.01522	0.28743	0.28539	−0.02222	0.24185	0.08712	0.09416	0.09840	−0.07446	0.45529
Spread	−0.00145	−0.00662	−0.01043	0.00871	0.03614	0.05579	0.00024	0.16497	0.01408	0.03802	0.04701	0.01771	−0.40940
House	−0.00707	0.00394	0.00662	0.00140	−0.00190	0.00001	0.02547	0.05888	0.23645	0.16068	0.22311	0.03414	0.02908
NAPM	0.00358	0.01697	−0.00038	−0.00077	0.02318	0.00696	0.00168	0.03195	0.18262	0.25212	0.26275	0.01497	0.06110
Emp	−0.00027	0.00820	0.00271	−0.00204	0.01732	0.00123	0.01399	0.01210	0.13752	0.50298	0.34197	−0.18065	0.08344
Output	0.02277	0.00723	−0.00215	−0.00949	0.02566	0.00456	0.01303	0.02291	0.09482	0.25843	0.54527	−0.12593	0.00666
Cons/Inv	0.02000	0.05286	−0.01386	−0.01486	0.02644	0.00556	0.01784	0.02353	0.06354	0.13890	0.25109	−0.05235	0.01925
UnEmp	−0.00700	0.01196	−0.00127	0.01006	−0.01707	0.00179	0.00233	0.00114	−0.02865	−0.02738	−0.01122	0.18295	−0.09804
FFR	0.02356	−0.00849	0.00067	−0.00784	0.15099	−0.05982	0.00287	0.00676	0.01914	0.00209	0.00597	−0.02594	0.38262

* The lower triangular is for covariance values and the upper triangular is for correlation values.
* See Appendix E for the description of representative aggregated variables, where variables are in the same order.

The entire causal structure can be understood as three parts for a convenient explanation. The first part is the real economy sector, which consists of consumption/investment, output, employment, unemployment, and house factors. The second part is money/interest sector, which consists of federal fund rate, monetary aggregate, interest rate spread, and interest rate factors. The third part consists of exchange rate, price, stock, and NAPM factors. In the real economy component, the contemporaneous causal order is consumption/inventory, output/production, employment, and unemployment factors. The housing factor is directly affected by the monetary aggregate, employment, and consumption/investment factors. In the money/interest component, the federal fund rate and monetary aggregates affect interest rates either directly or through interest rate spread. Interest rate factor is also affected by the exchange rate factor. In the third component, the influences of the money/interest part, summarized by interest rate factor, on the price factor are transmitted by the NAPM and financial market (stock) factors. The price factor is also affected by the exchange rate factor. On the other hand, the effects of the monetary policy, summarized by interest rate factor, on the real economy part, more directly consumption/investment and output factors, are transmitted by the NAPM and financial market (stock) factors. In this respect, it can be argued that the monetary transmission mechanisms identified in this causal structure are interest rate, financial market (stock), and expectation (NAPM) channels. The financial market (stock) and expectation (NAPM) factors turn out to be crucial channels to transmit the causal influences from money/interest part into the rest parts of the overall economy. Note that all the causal interpretations are based on the contemporaneous causal structure among innovations from a reduced form FAVAR.

Empirical Results of the Structural FAVAR

Based on the identified structural coefficient matrix $\hat{A}_0$, the estimated impulse response functions of FAVAR are used to study the responses of the system to particular initial shocks. The impulse response functions of individual macroeconomic indicators are obtained by using the impulse responses of FAVAR and the estimated coefficient of $\hat{\Lambda}^i_f$ based on the relationship $X^i_t = \Lambda^i_f \cdot F^i_t + e^i_t$. The resulting impulse response functions describe the effects of variables to one standard deviation shock to the federal fund rate variable and each of the estimated factors. The impulse response functions of all the variables with respect to a initial shock in the federal fund rate are presented in Figure 4.4.



* Straight lines represent IRF estimates based on proposed Grouped FAVAR method and dotted lines straight lines represent those based on previous Ungrouped FAVAR method.

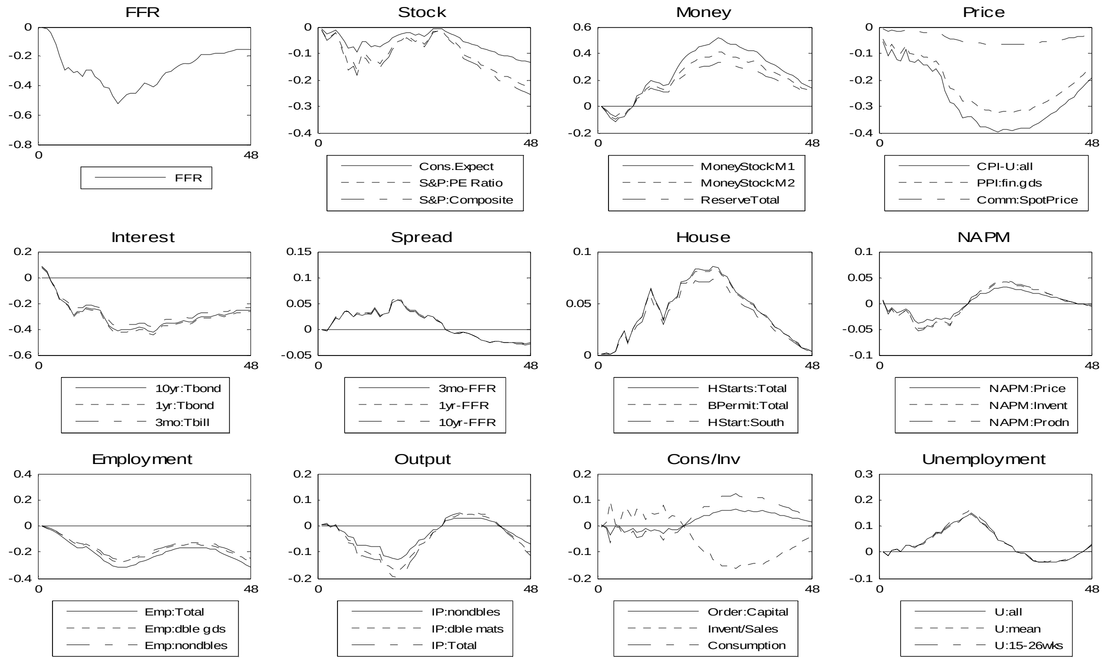
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.4. Impulse Responses to Federal Fund Rate Shock

For comparison purpose, the impulse responses obtained from the ungrouped FAVAR, which follows the methods of Bernanke, Boivin, and Eliasz (2005), are also presented. As discussed previously, their methods are based on the estimated factors from the entire data set (ungrouped FAVAR) and the assumed full recursive restrictions. On the other hand, the method used in this study is based on the estimated factors from the inductively classified groups of variables (grouped FAVAR) and inductively inferred causal structures. The results of the ungrouped FAVAR are used as the baseline with which our result is compared, since the FAVAR methods and empirical results of Bernanke, Boivin, and Eliasz (2005) are generally accepted as the benchmark for the study of the monetary policy effects (see Stock and Watson, 2005 for example).Both ungrouped and grouped FAVAR models result in similar impulse responses with respect to federal fund rate shock, except the grouped FAVAR model generally has smaller magnitude of responses than the ungrouped FAVAR. (a) The exchange rates appreciate and eventually fall. (b) The stock markets, money aggregates decline. (c) Given that the price puzzle found in the literature remains beyond several years, the price puzzle is considerably reduced and prices eventually go down. The different movement among CPI, PPI, and spot market price index and sensitive material price index can be explained by the fact that the posted prices (CPI) adjust more slowly to the production cost shock induced by the federal fund rate shock. (d) The interest rates increase, whereas the interest rate spreads, housing market, and NAPM decline. (e) The real activity measures (employments, output, and consumption/investment group variables) decline and eventually return toward zero (long-run money neutrality). (f) The inventory and unemployment variables increase. The counter-intuitive results such as increase stock market and decrease exchange rates, found in another ungrouped FAVAR model applied to U.K data (Lagana and Mountford, 2005), are not found in our impulse responses. Given that these results appear to be sensible measures of the effect of monetary policy, the similar results obtained from the grouped FAVAR with graphical causal model approach used in this study may well be interpreted as an empirically plausible specification.

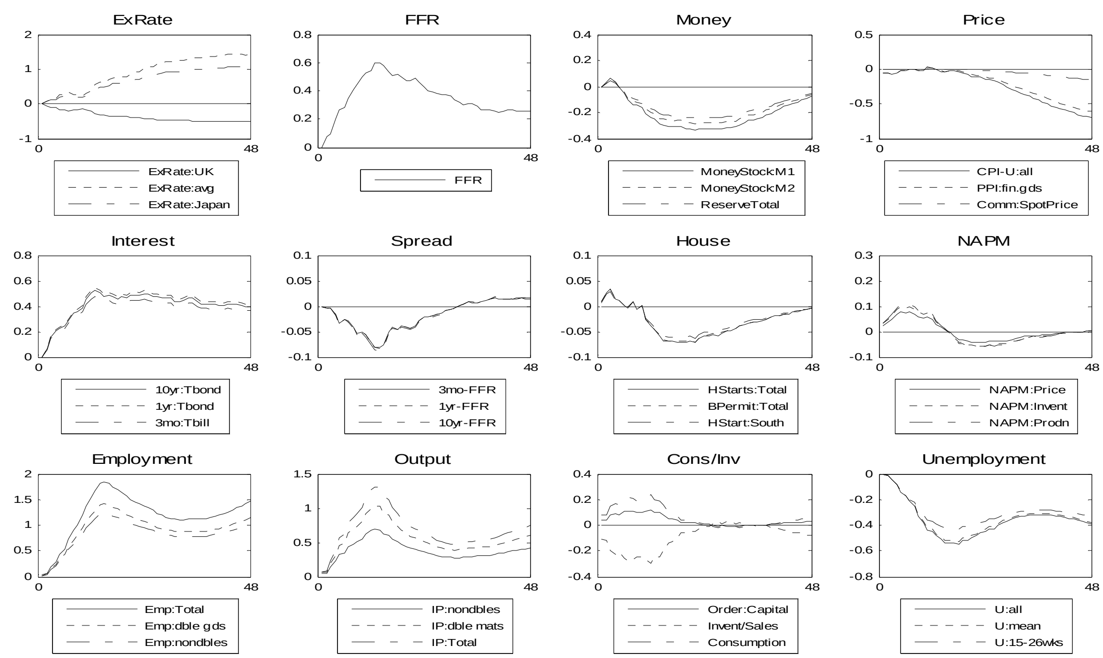
Compared to the usual small size VAR approach, the FAVAR approach has an advantage to obtain impulse responses for a large number of variables, that is, for any variables included in the data set. However, in the previous applied (ungrouped) FAVAR approaches, the latent factors are estimated from the entire data set and thus the estimated factors are linear combinations of all the variables in the data set. In this case, the advantage of the FAVAR model is restricted to study of impulse responses with respect to a shock of the observed (policy) variable, since it is not easy to provide economic interpretations for the impulse responses with respect to a shock of each of augmented factors, except the observed policy variable. Note also that it is not easy to provide clear economic meanings for the estimated factors and thus not easy to use other general identification schemes, except the full recursive one with the first or last ordered policy variable. Note that when the policy variable is in the middle of the causal order, even the full recursive assumption itself is not easy to use in the previous used (ungrouped) FAVAR approaches. Compared to the ungrouped FAVAR approach, the grouped FAVAR framework with compositional stability condition introduced in this study makes it possible to obtain meaningful factors and thus meaningful additional impulse responses with respect to shock of each of augmented factors as well as the observed (policy) variable. These additional impulse responses from the grouped FAVAR provide more structural information and allow the additional comprehensive checks on the empirical plausibility of the grouped FAVAR with a DAG specification.

The additional impulse responses of selected variables with respect to shocks in each of augmented factors are presented in Figure 4.5. to 4.16. For the interpretation of the results, it is convenient to describe the results based on the two general observations: (i) The movements of the real activity measures (employment, output, and consumption/investment variables) are generally in opposite direction to the movements of the inventory and unemployment variables. And thus these two opposite movements of variables for the real economy part are described by the movement of the real economy. (ii) The movements of the federal fund rate and interest rates are generally in opposite direction to the movements of the monetary aggregate variables, except for the housing market and unemployment shocks, which will be mentioned separately. And thus these two opposite movements of variables for the monetary economy part are described by the movement of the general interest rate. Based on these two general observations, (i) The movement of the general interest rate can be understood as the result of the monetary policy in the context of the relative movements of the real economy and the prices variables. For example, high inflation or excessive boom can induce contractionary monetary policy. (ii) The movements of the exchange rates and two asset markets of stock and housing markets and two types of expectations of the NAPM indices and interest rate spreads can be interpreted as the monetary policy transmission mechanisms channels based on the movement of the general interest rate, as well as the relative movements of the real economy and the prices variables. For example, high interest rates, induced by contractionary monetary policy, can decrease asset prices to stabilize



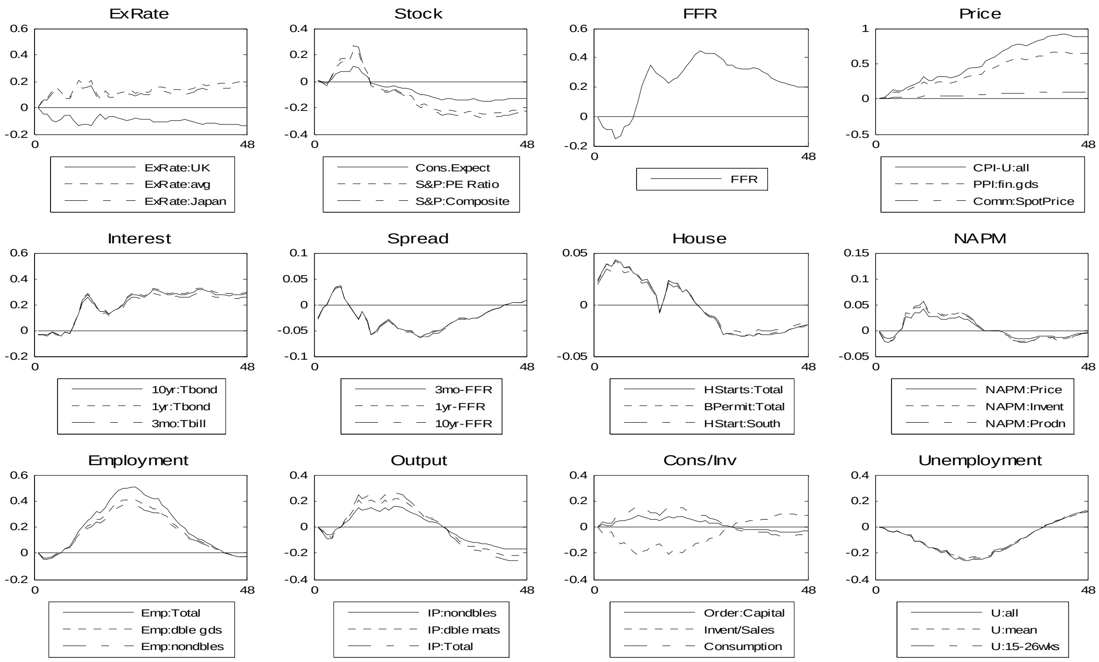
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.5. Impulse Responses to Exchange Rate Factor Shock



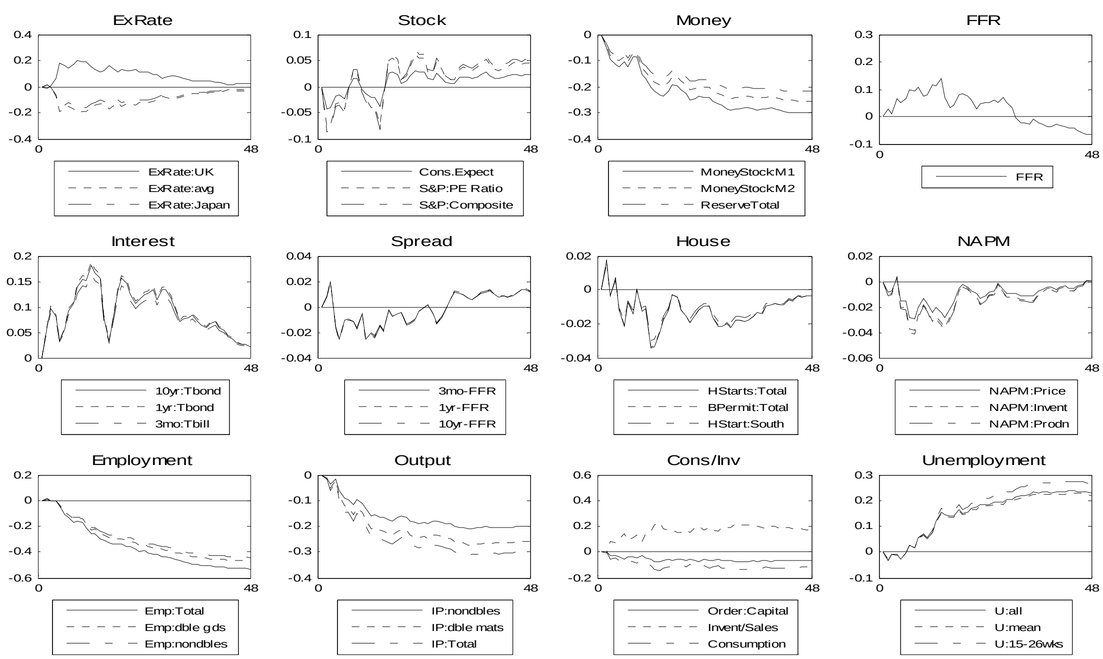
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.6. Impulse Responses to Stock Factor Shock



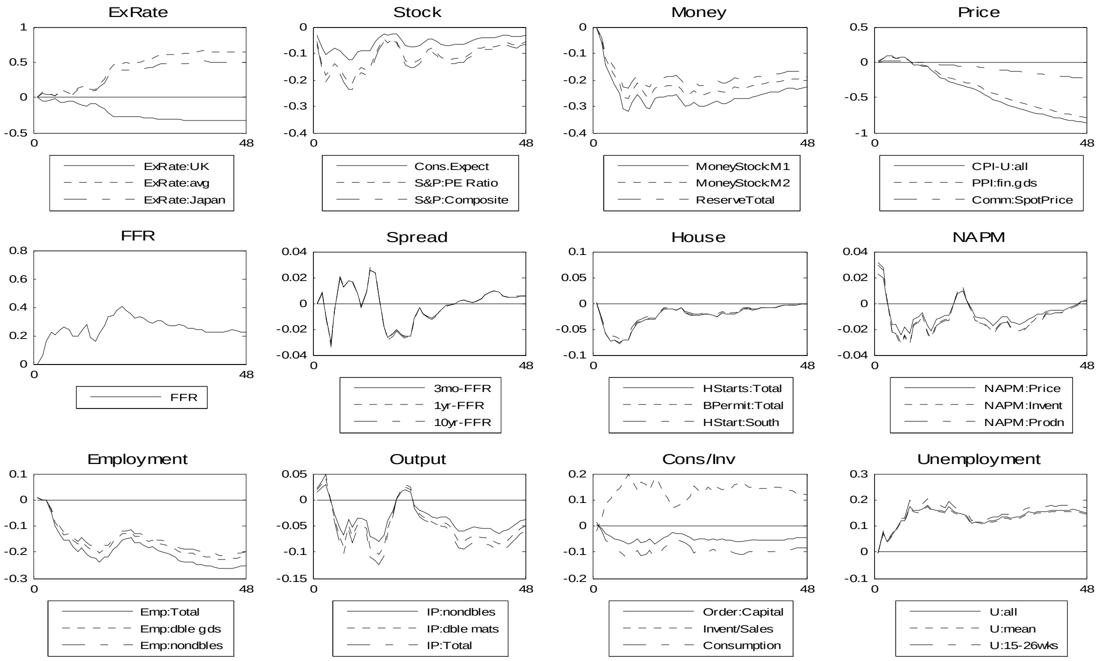
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.7. Impulse Responses to Money Factor Shock



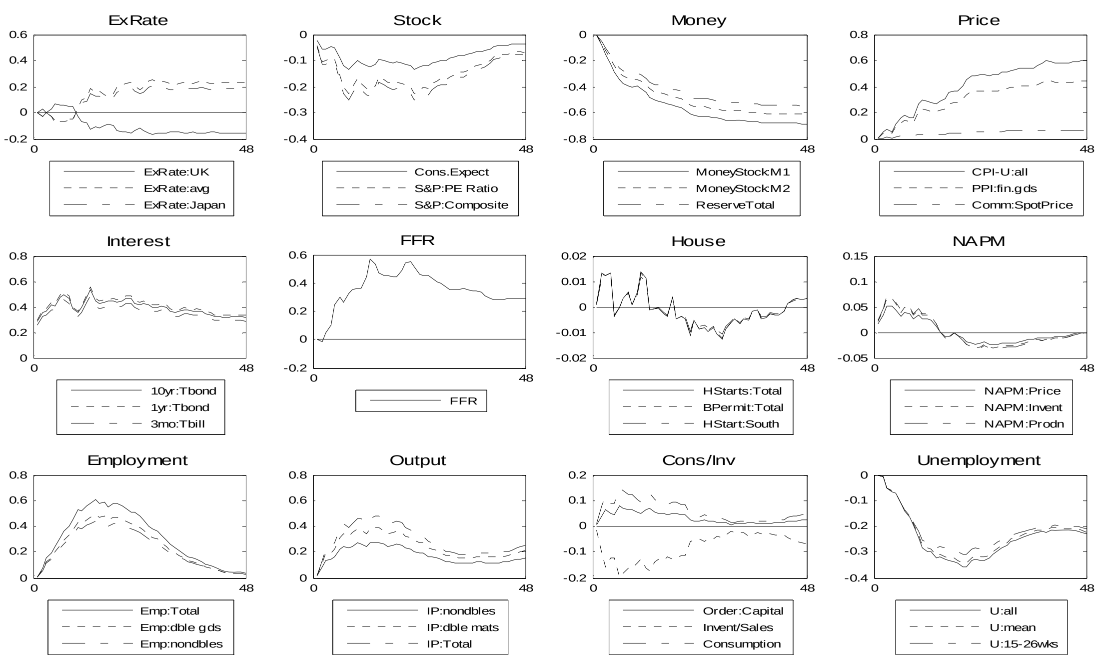
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.8. Impulse Responses to Price Factor Shock



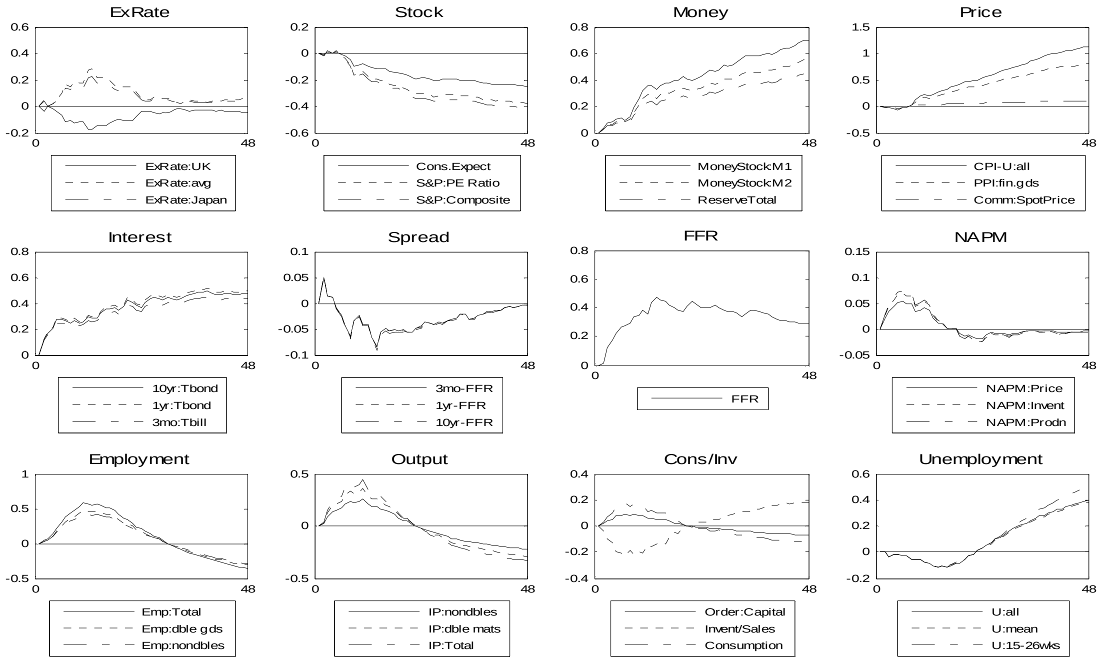
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.9. Impulse Responses to Interest Factor Shock



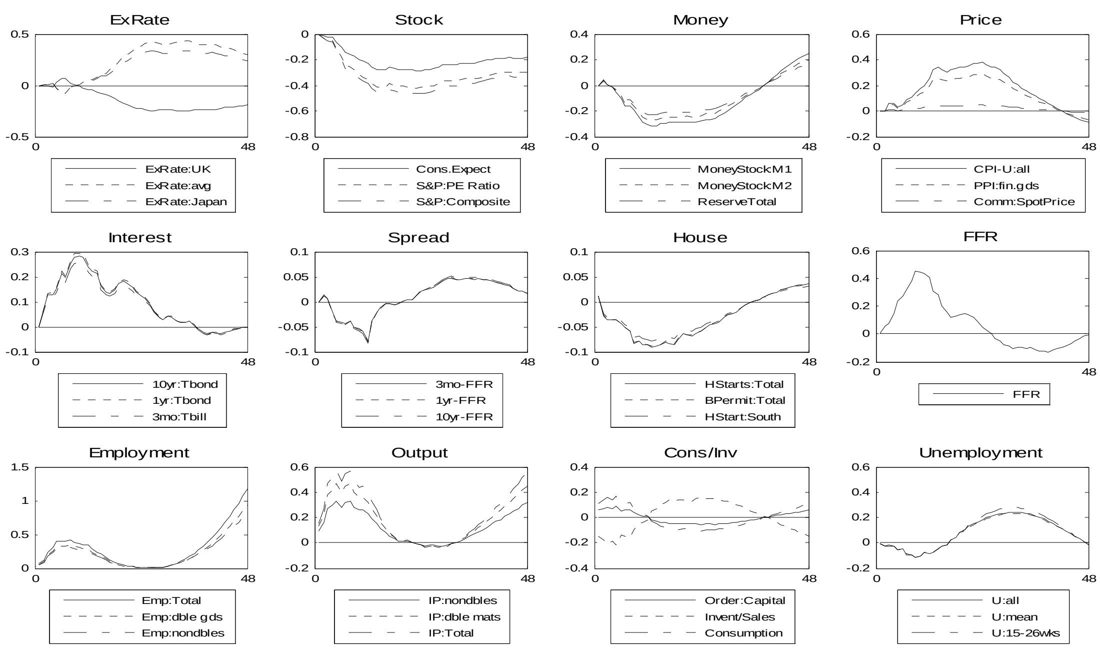
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.10. Impulse Responses to Spread Factor Shock



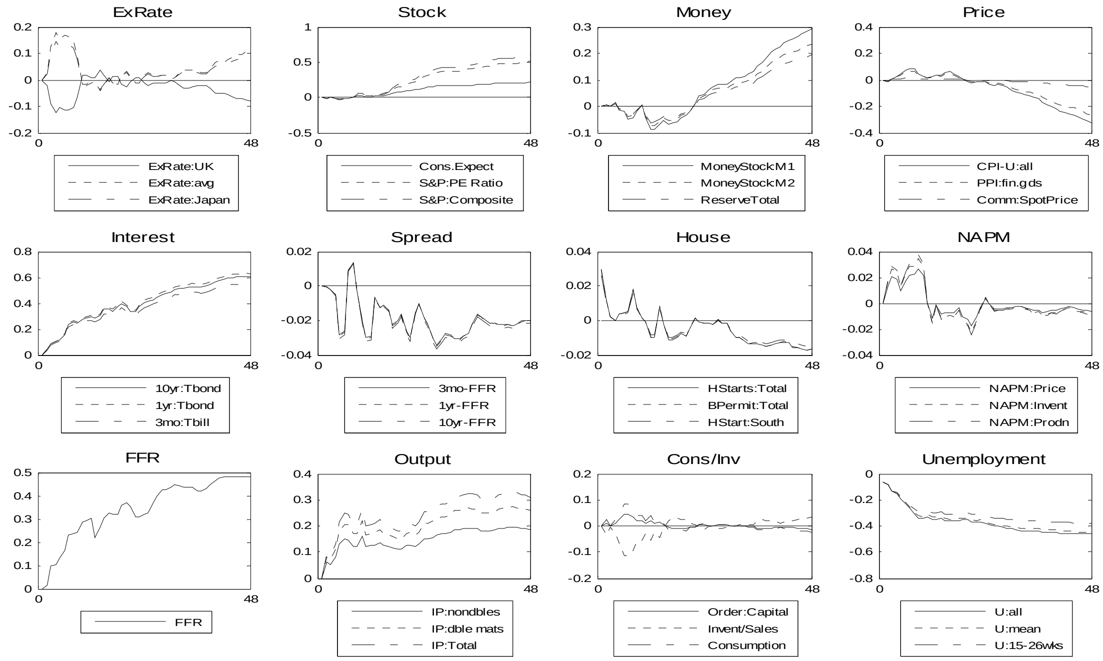
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.11. Impulse Responses to House Factor Shock



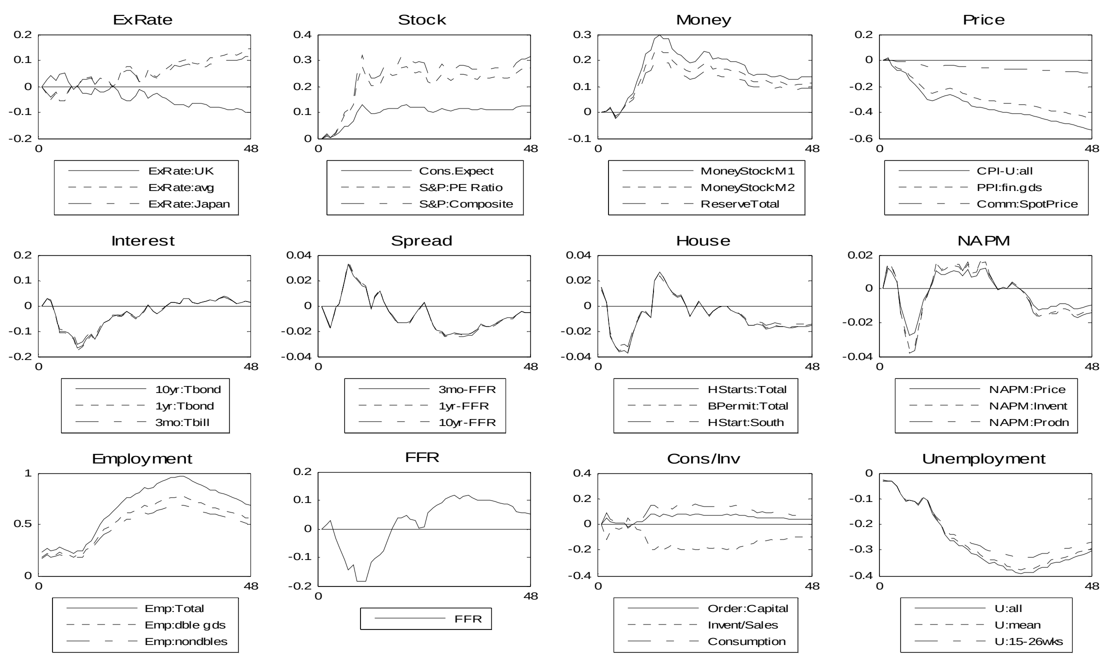
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.12. Impulse Responses to NAPM Factor Shock



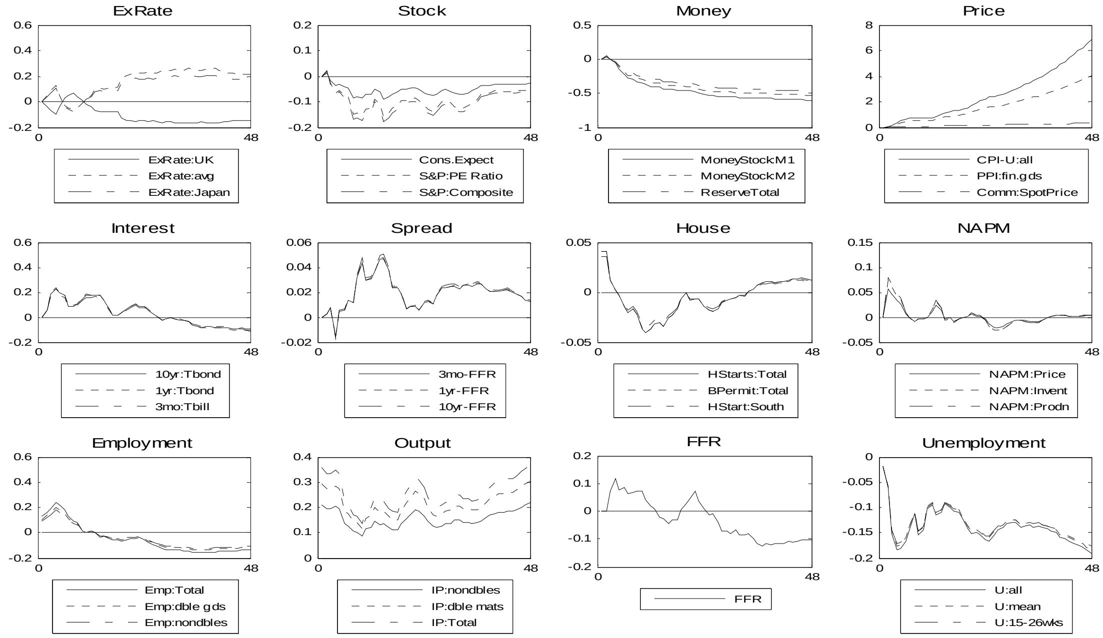
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.13. Impulse Responses to Employment Factor Shock



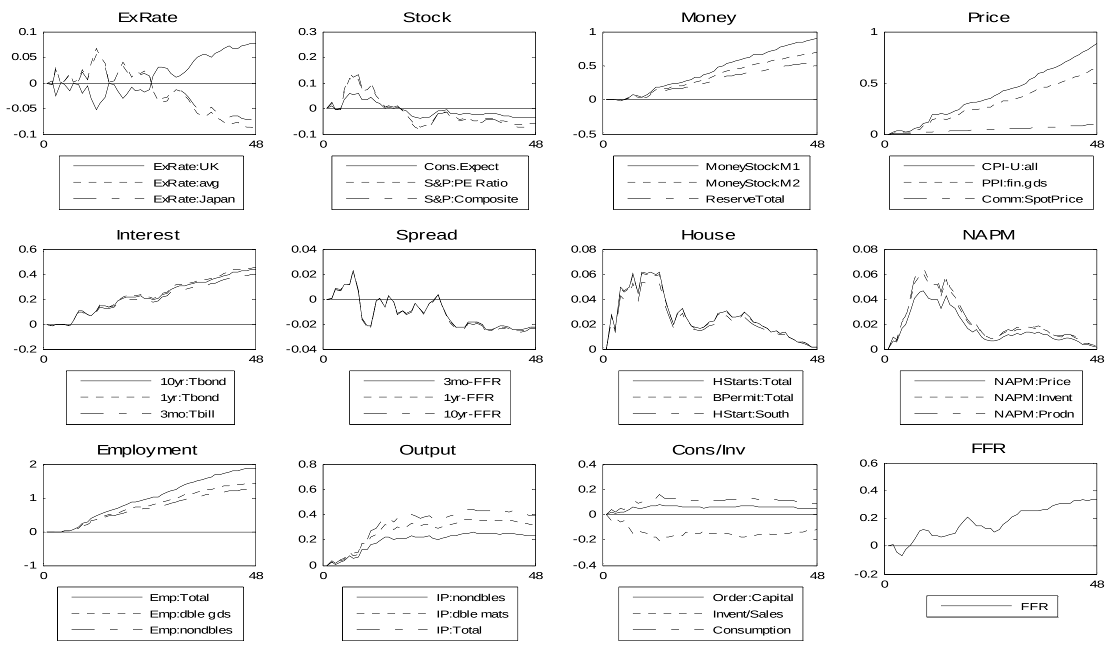
* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.14. Impulse Responses to Output Factor Shock



* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.15. Impulse Responses to Consumption/Investment Factor Shock



* See Appendix E for the description of variables, where variables are in the same order.

Figure 4.16. Impulse Responses to Unemployment Factor Shock

the high inflation or excessive boom economy. Note that the above descriptions only provide one possible convenient interpretation to check the empirical plausibility of model and thus they are understood in such a limited context. We do not have sufficient information on the complete causal structures among variables of the overall economy over the full dynamics interactions beyond contemporaneous time. The main advantage of VAR approach is that it does not require the kinds of the complex and full structural causal information. All we need for the identification problem in the VAR approach is the contemporaneous causal relationship among innovations, which can be inductively identified by the graphical causal model approach of the GES algorithm.

The contemporaneous causal relationships among innovations, however, do not provide sufficient information on the complete and fully dynamic causal structures among variables of overall economy. For example, the above description implicitly involves feedback causal structure beyond contemporaneous time: overall economic conditions at time $t-1$ or t $\rightarrow$ monetary policy at time t $\rightarrow$ movements of the monetary policy channels at time t or $t+1$ $\rightarrow$ overall economic conditions at time $t+1$ or $t+2$. For another caveat, the monetary policy channels of exchange rates, asset market, and expectations can be affected not only by the policy induced interest rates but also by the overall economic conditions as well as expectations. In fact, the general interest rates itself can be affected not only by the policy induced interest rates but also by the overall economic conditions as well as expectations. In this respect, the descriptions of the resulting impulse responses offered below are restricted to association without causal directions among variables.

General descriptions for the additional impulse responses are as follows: (a) For the exchange rate shock (Figure 4.5), the general interest rate slightly decreases with the decreased real economy and the decreased price variables. The stock market and NAPM variables decrease, whereas the house and spread variables slightly increase. (b) For the stock market shock (Figure 4.6), the general interest rate slightly increases with the increased real economy and the slightly decreased price variables. The exchange rates and NAPM variables increase, whereas the house and spread variables decrease. (c) For the money aggregates shock (Figure 4.7), the general interest rate drops initially but increases after some delay with the slightly increased real economy and the increased price variables. The exchange rate, house, NAPM variables slightly increase, whereas spread and stock market decreases after small jump. (d) For the prices shock (Figure 4.8), the general interest rate increases with the decreased real economy. The exchange rate, the stock market, spread, house, NAPM variables decrease. (e) For the interest rates shock (Figure 4.9), the general interest rate increases with the decreased real economy and the decreased price variables. The exchange rate increases, whereas the stock, house, and NAPM variables decrease. (f) For the interest rates spread shock (Figure 4.10), the general interest rate increases with the slightly increased real economy and the increased price variables. The exchange rate increases, whereas the stock, house, and NAPM variables decrease. (g) For the house market shock (Figure 4.11), the general interest rate increases with the slightly increased real economy and the increased price variables. The monetary aggregate and exchange rates variables slightly increase, whereas the stock, spread, and NAPM variables decrease. (h) For the NAPM shock (Figure 4.12), the general interest rate slightly increases with the slightly increased real economy and the increased price variables. The exchange rates slightly increases, whereas the stock, spread, house decrease. (i) For the employment shock (Figure 4.13), the general interest rate slightly increases with the slightly increased real economy and the decreased price variables. The exchange rates, house, NAPM variables shortly increase and return to normal, whereas the stock market slightly increases. (j) For the output shock (Figure 4.14), the general interest rate decreases with the slightly increased real economy and the decreased price variables. The exchange rates, spread, house, NAPM, and stock market slightly increase. (k) For the consumption/investments shock (Figure 4.15), the general interest rate increases with the slightly increased real economy and the increased price variables. The exchange rate, spread, NAPM variables slightly increase, whereas stock market and house slightly decrease after short jump. (l) For the unemployment shock (Figure 4.16), the general interest rate initially drops but slightly increases after short delay with the slightly increased real economy and the increased price variables. The monetary aggregates increase. Note that all the impulse response functions trace the effect to one time shock under the condition that all other innovations remain unchanged and thus the resulting impulse responses need to be interpreted under such cetris paribus condition. For example, an output innovation shock (technological advance for example), not followed by adverse movements of fundamentals of overall economy, can induce the slightly increased real economy and the decreased price variables and thus the stable general interest rate.

To study overall relationships among factors, the one-step forecast error variance in each factor is decomposed into the parts attributable to each of a set of innovations processes in the FAVAR. The results of forecast error variance decomposition are presented in Table 4.2.

Table 4.2. Forecast Error Variance Decomposition

	period	Money	FFR	Interest	Spread	ExRate	Stock	NAPM	House	Cons/Inv	Output	Emp	UnEmp	Price	Real	Channel
Money	0	100.000 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	6	63.186 1	7.123 3	5.213 5	7.919 2	1.612 8	2.035 7	2.713 6	0.721 11	6.363 4	0.898 10	0.355 13	0.493 12	1.372 9	8.109	14.999
	12	52.650 1	6.883 3	5.577 5	8.175 2	3.191 7	2.814 9	3.997 6	2.253 10	6.604 4	2.047 11	1.255 13	1.652 12	2.900 8	11.558	20.432
	36	46.183 1	7.300 3	5.738 5	9.256 2	4.398 8	3.324 7	4.636 6	2.564 11	6.972 4	2.266 10	1.695 13	2.463 12	3.203 9	13.397	24.178
FFR	0	0.000	100.000 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	6	2.373 8	67.722 1	3.231 6	5.097 2	3.781 4	4.101 3	3.395 5	3.050 7	1.480 10	1.369 11	2.146 9	1.329 12	0.926 13	6.324	19.425
	12	6.680 3	54.635 1	4.475 6	7.453 2	3.898 8	4.843 5	5.110 4	3.913 7	1.383 12	2.150 10	2.766 9	1.663 11	1.031 13	7.962	25.218
	36	6.741 3	47.258 1	5.374 5	7.690 2	5.165 6	4.867 7	5.893 4	4.081 8	2.102 13	2.659 10	3.417 9	2.547 11	2.208 12	10.725	27.695
Interest	0	0.417 5	20.058 3	50.832 1	26.353 2	2.340 4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	28.693
	6	0.741 12	14.247 3	41.019 1	17.584 2	4.642 5	3.292 7	2.364 9	4.278 6	5.512 4	1.643 11	2.377 8	0.509 13	1.793 10	10.041	32.160
	12	4.653 6	13.841 3	35.100 1	17.400 2	4.914 5	4.408 7	2.822 9	3.958 8	4.941 4	1.905 12	2.118 10	2.074 11	1.864 13	11.039	33.502
	36	5.740 5	13.480 3	29.871 1	15.663 2	5.534 6	4.295 8	3.337 10	4.535 7	5.973 4	2.350 13	3.042 11	2.432 12	3.749 9	13.796	33.364
Spread	0	1.297 3	16.739 2	0.000	81.963 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	6	1.224 5	52.520 1	0.586 9	37.472 2	1.038 7	1.040 6	2.039 3	1.839 4	0.192 13	0.658 8	0.563 10	0.301 12	0.527 11	1.715	5.957
	12	2.134 6	48.825 1	0.697 12	26.158 2	1.922 7	5.759 3	5.610 4	4.335 5	1.557 8	0.787 11	0.924 9	0.476 13	0.816 10	3.744	17.626
	36	8.122 4	36.834 1	1.173 11	16.887 2	3.942 7	7.763 5	6.831 6	9.656 3	3.049 8	1.383 10	2.560 9	1.001 12	0.800 13	7.992	28.191
ExRate	0	0.000	0.000	0.000	0.000	100.000 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	6	1.439 8	3.225 3	1.662 7	0.687 11	79.912 1	2.432 5	0.730 10	1.431 9	2.594 4	0.573 12	1.677 6	0.412 13	3.226 2	5.256	5.280
	12	2.029 10	3.238 4	2.533 7	2.469 8	70.148 1	4.981 2	1.389 11	2.244 9	2.953 6	0.916 12	3.109 5	0.681 13	3.311 3	7.658	11.083
	36	3.320 7	4.875 3	4.380 4	2.849 9	61.312 1	5.803 2	2.104 11	2.833 10	3.092 8	1.522 12	3.354 6	1.093 13	3.463 5	9.062	13.590
Stock	0	0.007 6	0.336 4	0.852 2	0.442 3	0.039 5	98.323 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.481
	6	1.852 5	1.778 7	3.472 2	2.787 4	1.539 8	79.436 1	3.279 3	0.722 11	0.913 10	0.530 12	0.360 13	1.500 9	1.832 6	3.303	8.327
	12	5.087 2	2.730 6	3.611 4	3.012 5	2.475 8	67.561 1	4.186 3	2.640 7	2.007 11	2.059 10	0.693 13	1.856 12	2.082 9	6.616	12.313
	36	5.296 2	4.382 4	4.707 3	3.266 7	3.133 8	59.311 1	4.372 5	3.023 9	2.840 10	2.191 12	1.681 13	2.222 11	3.575 6	8.934	13.795
NAPM	0	0.024 7	1.160 5	2.940 3	1.524 4	0.135 6	3.437 2	90.779 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	5.096
	6	0.943 11	0.817 13	1.637 7	7.263 4	0.898 12	17.465 2	53.013 1	7.501 3	4.927 5	1.054 9	1.241 8	2.253 6	0.987 10	9.476	33.127
	12	3.159 9	5.537 6	1.757 11	6.814 4	3.548 7	18.757 2	38.047 1	8.388 3	3.530 8	1.207 13	1.592 12	5.812 5	1.852 10	12.140	37.507
	36	3.814 8	16.024 3	2.058 11	6.506 4	5.954 7	18.756 2	26.829 1	6.157 5	2.981 9	1.231 13	1.386 12	6.028 6	2.277 10	11.626	37.372

Table 4.2. (Continued)

	period	Money	FFR	Interest	Spread	ExRate	Stock	NAPM	House	Cons/Inv	Output	Emp	UnEmp	Price	Real	Channel
House	0	1.357 4	0.003 9	0.007 7	0.003 8	0.000	0.161 6	0.360 5	90.951 1	4.310 2	0.539 4	2.308 3	0.000	0.000	7.157	0.525
	6	4.183 6	5.287 3	13.423 2	0.244 13	0.405 12	1.006 9	4.704 5	59.815 1	1.901 8	2.827 7	0.500 10	5.230 4	0.476 11	10.458	6.358
	12	3.215 6	12.288 3	10.070 4	0.214 13	3.098 7	1.605 10	12.874 2	43.708 1	2.525 8	1.732 9	0.395 12	7.345 5	0.931 11	11.997	17.791
	36	2.641 8	24.014 1	5.533 6	0.216 13	11.984 4	9.588 5	14.471 3	22.193 2	1.764 9	1.112 11	0.369 12	4.976 7	1.140 10	8.220	36.258
Cons/Inv	0	0.001 8	0.034 6	0.086 4	0.045 5	0.004 7	3.100 3	5.340 2	0.000	91.390 1	0.000	0.000	0.000	0.000	0.000	8.489
	6	1.172 12	3.145 6	1.187 11	3.161 5	3.084 7	3.792 4	5.053 2	1.762 9	68.659 1	4.052 3	1.078 13	1.605 10	2.250 8	6.736	16.852
	12	1.859 10	4.309 4	1.617 12	3.076 8	3.232 7	3.836 5	5.973 2	2.449 9	60.990 1	5.748 3	1.556 13	1.836 11	3.520 6	9.140	18.565
	36	2.920 10	5.055 4	2.627 11	3.350 8	3.880 7	4.847 5	5.847 2	3.331 9	54.243 1	5.773 3	1.810 13	1.948 12	4.369 6	9.531	21.255
Output	0	0.001 9	0.049 7	0.124 5	0.064 6	0.006 8	1.260 4	5.302 3	0.000	24.197 2	68.997 1	0.000	0.000	0.000	24.197	6.632
	6	1.828 10	2.078 8	1.224 12	4.323 5	0.290 13	10.251 3	9.519 4	3.611 6	16.493 2	44.519 1	2.666 7	1.306 11	1.893 9	20.465	27.994
	12	3.005 9	6.091 5	1.872 12	3.982 7	0.658 13	10.160 3	8.888 4	4.135 6	14.260 2	38.719 1	3.098 8	2.614 10	2.519 11	19.971	27.823
	36	4.400 8	8.225 5	2.520 11	3.931 7	1.695 13	10.320 3	9.399 4	5.381 6	12.744 2	33.238 1	2.924 9	2.711 10	2.513 12	18.379	30.725
Emp	0	0.000	0.012 8	0.031 6	0.016 7	0.001 9	0.319 5	1.341 4	0.000	6.121 3	17.455 2	74.702 1	0.000	0.000	23.577	1.678
	6	0.941 11	1.102 10	1.846 9	3.865 7	0.741 13	12.436 2	5.761 4	4.664 6	5.101 5	10.446 3	50.151 1	0.869 12	2.078 8	16.416	27.467
	12	2.230 11	6.418 4	1.839 12	3.861 9	1.222 13	15.016 2	4.768 7	4.273 8	4.786 6	8.686 3	38.930 1	4.888 5	3.083 10	18.360	29.140
	36	3.339 11	14.012 2	2.089 12	3.910 9	1.958 13	12.282 3	5.415 7	5.748 5	4.269 8	7.822 4	30.155 1	5.435 6	3.569 10	17.525	29.312
UnEmp	0	0.000	0.000	0.001 7	0.001 8	0.000	0.010 6	0.044 5	0.000	0.200 4	0.570 3	2.438 2	96.737 1	0.000	3.207	0.055
	6	0.546 12	0.948 10	3.882 4	2.042 6	0.517 13	3.652 5	1.555 7	0.872 11	5.168 2	1.350 9	4.950 3	73.071 1	1.446 8	11.469	8.638
	12	1.604 11	3.486 7	4.375 5	3.529 6	0.640 13	7.887 2	1.752 10	1.250 12	5.717 3	2.730 8	5.226 4	59.249 1	2.554 9	13.674	15.058
	36	3.381 9	7.971 2	4.151 6	3.623 8	1.555 13	7.949 3	3.634 7	3.370 10	5.094 4	3.334 11	4.574 5	48.439 1	2.926 12	13.001	20.131
Price	0	0.000	0.006 6	0.014 4	0.007 5	1.260 3	1.667 2	0.000	0.000	0.000	0.000	0.000	0.000	97.046 1	0.000	2.934
	6	1.780 8	3.123 3	1.269 10	1.952 6	5.820 2	2.856 4	1.970 5	0.671 11	1.786 7	1.581 9	0.431 12	0.173 13	76.587 1	3.971	13.270
	12	3.410 6	3.825 4	1.909 9	4.171 3	6.029 2	3.463 5	2.041 8	1.653 12	2.127 7	1.858 10	0.677 13	1.680 11	67.156 1	6.343	17.356
	36	4.042 5	4.296 3	2.245 9	4.233 4	7.144 2	3.795 6	2.485 8	1.715 12	3.109 7	2.014 10	1.088 13	1.888 11	61.946 1	8.099	19.373

* Each cell in the table contains the percentage of the forecast error accounted for by each innovation with relative ranking at each period.
* The Real column is the sum of Consumption/Investment, Output, Employment, and Unemployment factors except own factor.
* The Channel column is the sum of Spread, Exchange Rate, Stock, NAPM, and House factors except own factor.

The table gives the percentage of the forecast error uncertainty explained by each of innovations with the relative ranking at period of 0, 6, 12, and 36 months. It is helpful to understand the estimated factors or innovations as following several categories: the real economy category consists of the consumption/investment, output, employment, and unemployment innovations and the monetary policy transmission channel category consists of the interest, spread, exchange rate, stock, NAPM, and house market innovations. The remaining innovations are from the money and price factors with the federal fund rate variable. The last two columns in the Table 4.2 are the sum (except own contribution) of attributable parts of the real economy and channel except interest rate factors. In general, the percentage of the forecast error uncertainty explained by each of innovations is not much different each other and thus there is no dominant innovation for explaining each of the forecast error uncertainty of all the factors. Given this general observation, the relative importance of each innovation is not interpreted easily. The overall results, however, can be interpreted as follows. The overall contribution of the monetary policy transmission channel except interest rate channel is approximately 20 %. This result suggests the importance of incorporating this set of variables into the empirical model in addition to the interest rate channel. The overall contribution of the real economy category factors is approximately 10 %. Especially their overall contributions for the money, price, and interest rate factors and federal fund rate variable are 13.40, 8.10, 13.80, and 10.73 %. Given that the money and interest rate innovations explain the price forecast error about 4.04 and 2.25 % and the price and interest rate innovations explain the money forecast error about 3.20 and 5.74 %, this result suggests that a dichotomy between the real and nominal variables is not observed.

Other individual results can be described based on the 6 and 36 month horizons as representing the short run and long run relationships except own contributions. The own contribution for each of factors are in the diagonal positions in the table. (a) For the money aggregate factor, the spread (7.92), federal fund rate (7.12), consumption/investment (6.36), and interest rate (5.21) innovations appear to be important in the short run and the spread (9.26), federal fund rate (7.30), consumption/investment (6.97), and interest rate (5.74) innovations appear to be important in the long run. (b) For the federal fund rate, the spread (5.10), stock market (4.10), exchange rate (3.78), NAPM (3.40), and money (2.37) innovations appear to be important in the short run and the spread (7.69), money (6.74), NAPM (5.89), interest rate (5.37), and exchange rate (5.17) innovations appear to be important in the long run. (c) For interest rate factor, the spread (17.58), federal fund rate (14.25), consumption/investment (5.51) innovations appear to be important in the short run and the spread (15.66), federal fund rate (13.48), consumption/investment (5.97), money (5.74), and exchange rate (5.53) innovations appear to be important in the long run. (d) For interest rate spread factor, the federal fund rate (52.53) and NAPM (2.04) innovations appear to be important in the short run and the federal fund rate (36.83), housing market (9.66), money (8.12), stock market (7.76), and NAPM (6.83) innovations appear to be important in the long run. (e) For the exchange rate factor, the price (3.226) and federal fund rate (3.225) innovations appear to be important in the short run and the stock market (5.88), federal fund rate (4.88), interest rate (4.38), and price (3.46) innovations appear to be important in the long run. (f) For the stock market factor, the interest rate (3.47) and NAPM (3.28) innovations appear to be important in the short run and the money (5.30), interest rate (4.71), federal fund rate (4.38), and NAPM (4.37) innovations appear to be important in the long run. (g) For the NAPM indices factor, the stock market (17.47), house market (7.50), spread (7.26), and consumption/investment (4.93) innovations appear to be important in the short run and the stock market (18.76), federal fund rate (16.02), spread (6.51), house market (6.16), and unemployment (6.03) innovations appear to be important in the long run. (h) For the housing market factor, the interest rate (13.42), federal fund rate (5.29), unemployment (5.23), NAPM (4.70), and money (4.18) innovations appear to be important in the short run and the federal fund rate (24.01), NAPM (14.47), exchange rate (11.98), stock market (9.59), and interest rate (5.53) innovations appear to be important in the long run. (i) For the consumption/investment factor, the NAPM (5.05), output (4.05), stock market (3.79), spread (3.16), and federal fund rate (3.15) innovations appear to be important at the short run and the NAPM (5.85), output (5.77), federal fund rate (5.06), stock market (4.85), and price (4.369) innovations appear to be important in the long run. (j) For the output factor, the consumption/investment (16.49), stock market (10.25), NAPM (9.52), spread (4.32), and housing market (3.61) innovations appear to be important in the short run and the consumption/investment (12.74), stock market (10.32), NAPM (9.40), federal fund rate (8.23), and housing market (5.30) innovations appear to be important in the long run. (k) For the employment factor, the stock market (12.44), output (10.45), NAPM (5.76), consumption/investment (5.10), and housing market (4.66) innovations appear to be important in the short run and the federal fund rate (14.01), stock market (12.28), output (7.82), housing market (5.75), unemployment (5.44), NAPM (5.41), and consumption/investment (4.27) innovations appear to be important in the long run. (l) For the unemployment factor, the consumption/investment (5.17), employment (4.95), interest rate (3.88), and stock market (3.65) innovations appear to be important in the short run and the federal fund rate (7.97), stock market (7.95), consumption/investment (5.09), employment (4.57), and interest rate (4.15) innovations appear to be important in the long run. (m) For the price factor, the exchange rate (5.82), federal fund rate (3.12), and stock market (2.86) innovations appear to be important in the short run and the exchange rate (7.14), federal fund rate (4.30), spread (4.23), money (4.04), and stock market (3.80) innovations appear to be important in the long run.

Given that there is no dominant innovation to explain each of the forecast error variance, the overall results can be summarized as follows. The federal fund rate innovation is important for each forecast error uncertainty of almost all the factors, whereas the spread, money, NAPM, interest rate innovations are important to explain the forecast error variance of the federal fund rate variable. The stock market innovation is important for each forecast error uncertainty of the channel and real category factors, whereas the money, interest rate, federal fund rate, and NAPM innovations are important to explain the forecast error variance of the stock market factor. The consumption/investment innovation is important for each forecast error uncertainty of the real category factors, whereas the NAPM, output, federal fund rate, stock market, and price innovations are important to explain the forecast error variance of the consumption/investment factor.

Summary and Discussion

The proposed methodological procedure to address two methodological issues in the study of monetary policy effect is illustrated by using macro-economic panel data of time series variables. The two methodological issues are the informational issue and the causal identification issue. For the informational issue to incorporate broad information into empirical model, the aggregation method based on the compositional stability condition is used. The legitimate classification is inductively identified among macro-economic variables and the empirical evidence is provided based on the approximate form of the compositional stability condition. The following groups with the Federal Funds Rate variable are used for subsequent analyses: Exchange Rate, Stock Market, Money Aggregate, Price, Interest Rate, Spread of interest rate, Housing Market, NAPM indices, Employment, Output, Consumption/Investment, Unemployment groups.

Given that the disaggregate original variables approximately satisfy the consistent aggregation condition of compositional stability condition, the use of aggregated variables and their relationships can be justified as legitimate representatives of disaggregate variables and their relationships for the following subsequent analyses. The estimated latent factors for each classified group are used in the factor augmented vector autoregressive (FAVAR) framework. As the homogeneity of variables in each group allows meaningful interpretations of each estimated factor, the contemporaneous causal structure among innovations of FAVAR is inductively inferred by using the GES algorithm. Based on the identified casual structure by the graphical causal model, the impulse response functions with respect to a shock in each of the estimated factors as well as the monetary policy variable are estimated. While the estimated impulse response functions of FAVAR are used to study the responses of the system to particular initial shocks, to study overall relationships among factors, the forecast error variance in each factor is decomposed into the parts attributable to each of a set of innovations processes in the FAVAR. The empirical results suggest the importance of incorporating a broad range of information into an empirical model. The empirical findings imply that the informational issue in the small size VAR can explain the so called the price puzzle phenomenon and the monetary policy transmission channels such as stock market, spread of interest rate and NAPM indices in addition to the interest rate channel are important to understand the overall macro-economy. Compared to the previously used ungrouped FAVAR with the recursive assumption or deductively grouped FAVAR with the recursive assumption, the empirically grouped FAVAR with inductively inferred causal structure used in this study is more consistent with the fact that the VAR approach emphasizes the inductively inferred information from the data itself rather than the deductively maintained information from the researchers' intuition.

As future research directions, several methodological issues to be studied can be suggested. A first issue is how to incorporate the non-stationarity in the original data and capture the possible co-integration relationships into the grouped FAVAR framework. The dynamic correlation and the principal component methods used in this study are based on the stationarity condition, which require transformations of the original data. The main issue is to find inductive classification and aggregation methods, which allow the possible non-stationarity of the original data. A second issue is how to incorporate the possible non-linearity such as structural changes. While the observed co-movements among macro-economic time series variables provide empirical foundation for the proposed non-parametric methods of classification and aggregation, the non-linearity phenomenon is oftentimes observed in macro-economic time series variables. One possible approach is to use the state space framework with the Gibbs sampler method in the

Bayesian perspective (Kim and Nelson, 1997). The main issue is how to inductively decide the parametric value, given that the state space framework is the full parametric approach. For example, the parameter value to capture the distributed lag effect of factors on the individual variables is not easily identified. A third issue is how to decide the boundary of the variables included in the entire data set. While the issue of what variables are included in a particular group can be inductively addressed by the proposed classification methods, the issue of what variables should be included in the entire data set can only be addressed based on the researchers' intuition or the theory. This issue is related with the causal sufficiency issue in the graphical causal models. The main issue is how to satisfy or how to relax the causal sufficiency conditions in the analysis, especially in the GES algorithm with discriminating the possible cyclic phenomenon. A fourth issue is how to decide the number of classified groups and estimated factors for each group. For an example of the number of classified groups, the empirical testing of the compositional stability condition, illustrated in the micro-econometric analysis in chapter III, requires the identification of instrumental variables. One possible way to pursue is to use the graphical causal model to identify instrumental variables, as Chalak and White (2006) propose. The main issue is how to use causal structure among observed variables to identify the validity condition of the instrumental variables, which involve the unobserved causal factors. A fifth issue is how to study the complete causal structures among variables over the full dynamic interactions beyond contemporaneous time. While the VAR framework only require the contemporaneous causal structure among innovations, identifying the complete causal structure such as feedback phenomena over full dynamic period can allow more precise understanding of macro-economic phenomena. One possible way is to apply the graphical causal model onto the dynamically separate variables based on the possible lag. For example, the N vector of time series variables with P lag of X_t, $X_{t-1}, \cdots$, X_{t-P} can be separately defined and then the graphical causal model is applied for this extended $N \cdot P$ dimensional data set. The full dynamic causal information can be incorporated into the VAR framework or the final form of dynamic SEM framework. The main issue is how to handle the complexity in the extended $N \cdot P$ dimensional data. A sixth issue is how to study macro-economic phenomena at the original disaggregate level beyond the aggregate level used in this study, given that close co-movement among variables implies that the (probabilistic) stability condition is violated and multicollinearity problem is severe. While this issue is partially addressed based on the factor analysis framework, alternative approach is to use the mixed estimator. The main issue is how to combine aggregate level information into the mixed estimator to study disaggregate level. Although there remain many other methodological issues to be addressed in empirical study, this study provides one plausible inductive procedure for the understanding of macro-economic structure, while minimizing the deductive properties or ambiguities. The remaining subjectivities in our proposed method are left as further research topics, with the hope that the remaining subjectivities bring fewer ambiguities relative to the previously used methods.

CHAPTER V
CONCLUSION

Economic studies have experienced significant advances in the theoretical, methodological, and empirical perspectives. From the empirical perspective, recent advances in data processing capabilities have brought the possibility of analyzing a large number of detailed variables. In many areas of economics, high dimensional panel data are now available. For example, retail checkout scanner data are available for thousand of products at firm, regional and national levels at various frequencies. And central banks and statistical institutes produce a large number of macro-economic time series data. These data have brought forth research potentials for significant advances in the micro-econometric analysis of consumer behavior and the macro-econometric study of monetary policy effects. From the methodological perspective, empirical studies in economics have been developed to unify the theoretical-quantitative approach with the empirical-statistical approach. For this purpose, the structural equation model (SEM) approach has been proposed and used in economics. However, instead of using the full simultaneous equation approach, several alternative theoretical and methodological approaches have been proposed and used widely for several areas in economics. These phenomena are due to the fact that: (a) the instrumental variables needed to identify each equation in the SEM framework are not easy to find and/or (b) the restrictions for identification problem in the usual SEM approach are argued as neither credible nor required (Sims, 1980). In the study of the consumer behavior, the system-wise approach has been widely used to study interrelationships among related commodities demanded. Within this framework, full spectrums of direct, inverse, and mixed demand system of equations have been developed from the theoretical perspective of consumer behavior. On the other hand, in the study of the macro-economy, the structural vector autoregressive (VAR) model approach is widely used to study the effects of structural economic shocks. From the methodological perspective, the VAR framework provides the possibility of inferring causal information from statistical properties of the data without pretending to have too much *a priori* theory and/or without demanding too much information from the data.

The availability of high dimensional data, however, raises several methodological issues for the full realization of the research potentials brought by the large information set. This study pursued one plausible procedure to address two methodological issues of how to infer the causal structure from empirical regularities and how to incorporate the large information set into empirical model. To address the issue of how to infer the causal structure from empirical regularities, the graphical causal models are proposed as one plausible method to inductively infer causal structure from non-temporal and non-experimental data. However, the (probabilistic) stability condition for the graphical causal models can be violated for high dimensional data, when close co-movements and thus near deterministic relations exist among variables in high dimensional data. Aggregation methods are proposed as one possible way to address this issue, allowing one to infer causal relationship among disaggregated variables based on aggregated variables. The aggregation methods have been demonstrated to be helpful to address issue of how to incorporate a large information set into an empirical model, given that econometric considerations, such as degrees-of-freedom and multicollinearity, require an economy of parameters in empirical models. The weighting schemes to aggregate disaggregate micro-variables into aggregate macro-variable can be empirically decided, based on either index number theory or principal component approach. However, the actual aggregation procedures or decisions on weighting schemes require the legitimate classifications or sufficient conditions for the interpretable and consistent aggregation. In this respect, identifying legitimate aggregation conditions is found to be a primary consideration for both causal inference and actual aggregation.

We interpret theory as an inductive causal averaging procedure to deal with methodological issues at the beginning of this study. When we follow an inductive causal averaging procedure that concentrates only on similar tendencies to highlight a few common factors by ignoring many more individual differences and idiosyncrasies, we need to identify empirically justifiable conditions that allow us to legitimately define common tendencies and individual idiosyncrasies. We studied possible legitimate conditions for the interpretable and consistent aggregation based on both aggregation theory framework and statistical dimensional reduction methods with minimizing any deductive assumptions such as micro-homogeneity of micro-parameters, separability, and homogeneity of utility (production) function. From both the aggregation theory and the statistical dimensional reduction methods, we identify the similar generalized forms of the compositional stability condition. Based on the generalized condition for the consistent aggregation, we propose one possible methodological procedure to consistently address the two related issues of causal inference and actual aggregation procedures for the full use of research potentials brought by high dimensional data.

Given the observation that many variables in this high dimensional data move very closely, the compositional stability condition as the consistent aggregation condition provides an inductive way to pursue the possibility of obtaining not only (a) interpretable aggregate macro-variables as the representative aggregate of homogeneous disaggregate micro-variables but also (b) interpretable macro-parameters as the representative aggregate of corresponding micro-parameters for the subsequence analysis. This implies that when the micro-variables can be legitimately grouped and represented by macro-variables, it is possible to use aggregation methods (a) to incorporate broad range of information into the empirical models with minimizing econometric issues such as the multicollinearity and degrees of freedom, (b) to capture (causal) relationships among disaggregated variables through (causal) relationships among aggregated variables as the legitimate representatives. This compositional stability condition is used (a) to provide an inductive way of forming suitable partitions before conducting any empirical test to justify those classifications based on the empirical data patterns rather than on researchers' intuition and (b) to address the possible violation of the (probabilistic) stability condition to use the graphical causal models for the high dimensional data. Note that it is conceivable and oftentimes observed that the (probabilistic) stability condition for the graphical causal models is violated for using high dimensional data in empirical study, given the observation that there exist close co-movements and thus near deterministic relations among variables in high dimensional data. In this respect, we argue that the (probabilistic) stability condition for an "inductive causal" procedure requires the compositional stability condition for an "inductive averaging" procedure.

For the micro-econometric analysis of the consumer behavior, following methodological procedure is proposed and illustrated in chapter III: (a) Both a standard static correlation matrix and dynamic correlation matrices over identified frequency bands are used to measure co-movement among original variables. Based on these similarity measures of disaggregate micro-variables, the modified k-nearest neighbor algorithm is used to sort the variables such that the highly correlated variables are near each other along the main diagonal in the reordered correlation matrices. The block-diagonal pattern of reordered or sorted static and dynamic correlation matrices are used to identify homogeneous groups of variables based the approximate form of the compositional stability condition. (b) Based on identified classifications of the original variables, index number theory is used for the actual aggregation procedure. The Tornqvist-Theil index is the primary method to decide weighting schemes on aggregating disaggregated micro-variables into representative macro-variables within each identified group. (c) The identified classification and aggregation of micro-variables into macro-variables can be tested, as long as appropriate instrumental variables can be identified. A Hausman type misspecification test of $H_0 : \gamma_n = 0$ in the equation $x_n = X H_n + IV \cdot \gamma_n + \varepsilon_n^{IV}$, where x_n and X are micro- and macro-variables respectively and IV are Instrumental Variables such that IV is closely correlated with X and independent of d_n, provides a statistical test framework for the generalized form of the compositional stability condition of independence between d_n and X in the set of equations $x_n = X H_n + d_n$. (d) Given the observed phenomena of close co-movements and thus near deterministic relations among variables in high dimensional data, it is conceivable and oftentimes observed that the (probabilistic) stability condition for the graphical causal models is violated for using high dimensional data in empirical study. When this is the case, it is still possible to infer causal structures among micro-variables through relationships among representative aggregated macro-variables as long as the compositional stability conditions hold among micro-variables. PC algorithm or GES algorithm are used to infer causal structures among macro-variables as the legitimate representative causal relationships among micro-variables are used for the subsequent analysis. (e) Based on the local causal structure between price and quantity variables for a particular commodity, the AIDS type dependent variable synthetic functional forms for the direct, inverse, and mixed demand systems are estimated. (f) The Rotterdam, AIDS, NBR, and CBS type constant and/or variational parameterizations and synthetic model are statistically compared and the parameterizations for expenditure (scale) elasticities (flexibilities) and Slutsky (Antonelli) coefficients are chosen within each of direct, inverse, and mixed specifications. Based on the chosen parameterization, the direct, inverse, and mixed demand system are compared based on the model selection approaches, such as the Akaike information, Schwarz information, and the likelihood dominance criteria.

As future research directions for the micro-econometric analysis of the high dimensional data, several methodological issues can be suggested. A first issue is how to fully use the overall empirical findings. The model averaging approach, rather than model selection approach used in this study, can provide more precise understanding of consumer behavior. One possible approach for the model averaging method is to use the relative log-likelihood values of the direct, inverse, and mixed demand systems. The main issue is how to decide relative weights among competing models. A second issue is how to fully use the causal information inferred by the graphical causal models. Although only the local causal structure between the price and quantity variables are used in this study, other causal information can provide the possibility of a more complete understanding of the interactions in the market, which in turn allow a more precise measurements of consumer behavior. The main issue is how to combine the full causal information into the theoretical properties of demand functions while maintaining flexible and estimable functional form specification. A third issue is how to decide the boundary of the variables included in the empirical models. For example, there can be latent causal structures or interactions with other (size) commodities, although the size of 6/12 oz is used to decide what commodities are included in the study. The causal structure identified by the PC algorithm suggests that there may be latent causal variables among the price variables. The main issue is how to satisfy or how to relax the causal sufficiency conditions in the analysis, especially in the GES algorithm with discriminating the possible cyclic phenomenon. A fourth issue is how to incorporate the possible dynamic interactions and non-linearity in consumer behavior. Although the differential functional form approach provides a useful framework to deal with the possible non-stationarity of variables, incorporating the possible lagged interaction and structural change can provide more precise understanding of consumer behavior. The main issue is how to capture the possible dynamic interactions and non-linearity phenomena without sacrificing the theoretical properties of demand functions, while maintaining flexible and estimable functional form specification. A fifth issue is how to study consumer behavior at the original disaggregate level beyond the aggregated level used in this study, given that close co-movement among variables implies that the (probabilistic) stability condition is violated and multicollinearity problem is severe. One possible way is to use the mixed estimator. The main issue is how to combine aggregate level information into the mixed estimator to study disaggregate level.

For the macro-econometric analysis of the macro-economy, following methodological procedure is proposed and illustrated in chapter IV: (a) Both a standard static correlation matrix and dynamic correlation matrices over identified frequency bands are used to measure co-movement among original variables. Based on these similarity measures of disaggregate micro-variables, the modified k-nearest neighbor algorithm is used to sort the variables such that the highly correlated variables are near each other along the main diagonal in reordered correlation matrix. The block-diagonal pattern of reordered or sorted static and dynamic correlation matrixes are used to identify homogeneous group of variables, based the approximate form of the compositional stability condition. (b) Based on identified classifications of original variables, the statistical dimensional reduction method are used for actual aggregation procedure to decide weighting schemes for aggregating disaggregated micro-variables into representative macro-variables within each identified group. The principal component method applied onto each of groups is used as the best dimensional reduction method with as little loss of information as possible in the mean squared error sense. (c) Given that the inference based on the small size VAR can be misleading unless the reduced form innovations span the space of the structural shocks or the VAR model does not have the omitted variables problem, the estimated factors are augmented in the VAR (FAVAR) framework to increase the amount of information in the empirical model so that the reduced form residuals span the space of the structural economic shocks. (d) Based on the residuals of reduced form FAVAR, the contemporaneous causal structure among innovations is inferred by the graphical causal model. The identified compositional stability condition in the data makes it possible to infer causal structures among micro-variables through relationships among representative aggregated macro-variables. The PC algorithm or GES algorithm is used to infer causal structures among macro-variables as the legitimate representative causal relationships among micro-variables for the subsequent analysis. (e) Based on the contemporaneous causal structure used for identification of FAVAR, structural relationships of the macro-economy are studied in the two types of the moving average representations. The impulse response functions of all the observed variables with respect to shocks in the monetary policy variable as well as each of the estimated factors are estimated and interpreted. The forecast error variance in each factor is decomposed into the parts attributable to each of a set of innovations processes in the FAVAR.

As future research directions for the macro-econometric analysis of the high dimensional data, several methodological issues are suggested. A first issue is how to incorporate the non-stationarity in the original data and capture the possible co-integration relationships into the grouped FAVAR framework. The dynamic correlation and the principal component methods used in this study are based on the stationarity condition, which require transformations of the original data. The main issue is to find inductive classification and aggregation methods, which allow the possible non-stationarity of the original data. A second issue is how to incorporate the possible non-linearity such as structural changes. While the observed co-movements among macro-economic time series variables provide empirical foundation for the proposed non-parametric methods of classification and aggregation, the non-linearity phenomenon is oftentimes observed in macro-economic time series variables. One possible approach is to use the state space framework with the Gibbs sampler method in the Bayesian perspective (Kim and Nelson, 1997). The main issue is how to inductively decide the parametric value, given that the state space framework is the full parametric approach. For example, the parameter value to capture the distributed lag effect of factors on the individual variables is not easily identified. A third issue is how to decide the boundary of the variables included in the entire data set. While the issue of what variables are included in a particular group can be inductively addressed by the proposed classification methods, the issue of what variables should be included in the entire data set can only be addressed based on the researchers' intuition or the theory. This issue is related with the causal sufficiency issue in the graphical causal models. The main issue is how to satisfy or how to relax the causal sufficiency conditions in the analysis, especially in the GES algorithm with discriminating the possible cyclic phenomenon. A fourth issue is how to decide the number of classified groups and estimated factors for each group. For an example of the number of classified groups, the empirical testing of the compositional stability condition, illustrated in the micro-econometric analysis in chapter III, requires the identification of instrumental variables. One possible way to pursue is to use the graphical causal model to identify instrumental variables, as Chalak and White (2006) propose. The main issue is how to use causal structure among observed variables to identify the validity condition of the instrumental variables, which involve the unobserved causal factors. A fifth issue is how to study the complete causal structures among variables over the full dynamic interactions beyond contemporaneous time. While the VAR framework only requires the contemporaneous causal structure among innovations, identifying the complete causal structure such as feedback phenomena over full dynamic period can allow more precise understanding of macro-economic phenomena. One possible way is to apply the graphical causal model onto the dynamically separate variables based on the possible lag. For example, the N vector of time series variables with P lag of X_t, $X_{t-1}, \cdots$, X_{t-P} can be separately defined and then the graphical causal model is applied for this extended $N \cdot P$ dimensional data set. The full dynamic causal information can be incorporated into the VAR framework or the final form of dynamic SEM framework. The main issue is how to handle the complexity in the extended $N \cdot P$ dimensional data. A sixth issue is how to study macro-economic phenomena at the original disaggregate level beyond the aggregate level used in this study, given that close co-movement among variables implies that the (probabilistic) stability condition is violated and multicollinearity problem is severe. While this issue is partially addressed based on the factor analysis framework, alternative approach is to use the mixed estimator. The main issue is how to combine aggregate level information into the mixed estimator to study disaggregate level.

In summary, this study provides one plausible inductive procedure for the full realization of the recently available high dimensional data, while minimizing the use of deductive or subjective assumptions. Although there remain other methodological issues to be addressed in empirical studies, inductive properties are emphasized in every sequence of the proposed method, since any types of subjective properties can bring ambiguities into the empirical results. While theory as the inductive causal averaging procedure can allow some deductive elements in its developments, empirical methodologies need to be based more on inductive properties to maintain their objectivity. The remaining subjectivities in our proposed method are left as further research topics, with the hope that the remaining subjectivities bring fewer ambiguities relative to the previously used methods.

REFERENCES

Akaike, H. 1973. “Information Theory and an Extension of the Maximum Likelihood Principle.” In N.B. Petrou and F. Caski, eds. *Proceedings of 2nd International Symposium on Information Theory*, pp. 267-281.

Al-Kandari, N.M., and I.T. Jolliffe. 2001. “Variable Selection and Interpretation of Covariance Principal Components.” *Communications in Statistics Simulation and Computation* 30:339-354.

______ . 2005. “Variable Selection and Interpretation of Correlation Principal Components.” *Environmetrics* 16:659-672.

Bai, J. 2003. “Inferential Theory for Factor Models of Large Dimensions.” *Econometrica* 71:135-171.

Bai, J., and S. Ng. 2002. “Determining the Number of Factors in Approximate Factor Models.” *Econometrica* 70:191-221.

Balk, B.M. 2005. “Divisia Price and Quantity indices: 80 years after.” *Statistica Netherlandica*, 59(2):119-158.

Barnett, W.A. 1984. “Recent Monetary Policy and the Divisia Monetary Aggregates.” *The American Statistician* 38(3):165-172.

Barnett, W.A., and S. Choi. 1989. “A Monte Carlo Study of Tests of Blockwise Weak Separability.” *Journal of Business and Economic Statistics* 7:363-377.

Barten, A.P. 1993. “Consumer Allocation Models: Choice of Functional Form.” *Empirical Economics* 18:129-158.

Barten, A.P., and L.J. Berrndorf. 1989. "Price Formation of Fish: An Application of an Inverse Demand System." *European Economic Review* 33:1509-1525.

Belviso, F., and F. Milani. 2005. "Structural Factor-Augmented VAR (SFAVAR) and the Effects of Monetary Policy." Department of Economics, Princeton University, Working Paper.

Berkson, J. 1946. "Limitations of the Application of Fourfold Table Analysis to Hospital Data." *Biometrics Bulletin* 2:47-53.

Bernanke, B.S., and A.S. Blinder. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82:901-921.

Bernanke, B.S., and M. Gertler. 1995. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives* 9:27-48.

Bernanke, B.S., J. Boivin, and P. Eliasz. 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics* 120(1):387-422.

Blockorby, C., D. Primont, and R.R. Russell. 1995. "Separability." In P. Hammond, and C. Seidel, eds. *Handbook of Utility Theory*. Boston: Kluwer Academic Press.

Blundell, R., and J.M. Robin. 1995. "Latent Separability: Grouping Goods without Weak Separability." IFS Working Paper No. W95/09.

Blundell, R., and T.M. Stoker. 2005. "Heterogeneity and Aggregation." *Journal of Economic Literature* 43(2):347-391.

Boivin, J., and S. Ng. 2003. "Are More Data Always Better for Factor Analysis?" NBER Working Paper No. 9829.

Bouckaert, R.R. 1993. "IDAGs: A Perfect Map for Any Distribution." *Symbolic and Quantative Approaches to Reasoning and Uncertainty*, New York: Springer-Verlag, pp. 49-56.

Brown, M.G., J.Y. Lee, and J.L. Seale, Jr. 1995. "Family of Inverse Demand Systems and Choice of Functional Form." *Empirical Economics* 20:519-530.

Buntine, W.L. 1996. "A Guide to the Literature on Learning Probabilistic Networks from Data." *IEEE Transactions on Knowledge and Data Engineering* 8:195-210.

Capps, Jr., O., and H.A. Love. 2002. "Econometric Considerations in the Use of Electronic Scanner Data to Conduct Consumer Demand Analysis." *American Journal of Agricultural Economics* 84: 807-861.

Chalak, K., and H. White. 2006. "An Extended Class of Instrumental Variables for the Estimation of Causal Effects." Department of Economics, University of California, San Diego, Discussion Paper.

Chamberlin, G., and M. Rothschild. 1983. "Arbitrage, Factor Structure, and Mean-Variance Analysis in Large Asset Markets," *Econometrica* 51:1305-1324.

Chavas, J.P. 1984. "The Theory of Mixed Demand Functions." *European Economic Review* 24:321-344.

Chickering, D.M. 2002. "Optimal Structure Identification with Greedy Search." *Journal of Machine Learning Research* 3:507-554.

Christiano, L., M. Eichenbaum, and C. Evans. 1999. "Monetary Policy Shocks: What have we learned and to what end?" In J.B. Taylor, and M. Woodford, eds. *Handbook of Macroeconomics*, Amsterdam: Elsevier.

Connor, G., and R.A. Korajczyk. 1986. "Performance Measurement with the Arbitrage Pricing Theory: A New Framework for Analysis." *Journal of Financial Economics* 15:373-394.

Croux, C., M. Forni, and L. Reichlin. 2001. "A Measure of Comovements for Economic Indicators: Theory and Empirics." *The Review of Economics and Statistics* 83:232-241.

Davis G.C. 1999. "The Science and Art of Promotion Evaluation." *Agribusiness* 15(4):465-483.

Dhar, T., J.P. Chavas, and B.W. Gould. 2003. "An Empirical Assessment of Endogeneity Issues in Demand Analysis for Differentiated Products." *American Journal of Agricultural Economics* 85: 605-617.

Diewert, W.E. 1978. "Superlative Index Numbers and Consistency in Aggregation." *Econometrica* 46:886-900.

______ . 2001. "The Consumer Price Index and Index Number Purpose." *Journal of Economic and Social Measurement* 27:167-248.

______ . 2004. "A New Axiomatic Approach to Index Number Theory." Department of Economic, University of British Columbia, Discussion Paper No. 04-05.

Diewert, W.E., and T.J. Wales. 1995. "Flexible Functional Forms and Tests of Homogeneous Separability." *Journal of Econometrics* 67(2):259-302.

Eales, J.S., C. Durham, and C.R. Wessells. 1997. "Generalized Models of Japanese Demand for Fish." *American Journal of Agricultural Economics* 79:1153-1163.

Eichenbaum, M. 1992. "Comment on 'Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy' by Christopher Sims." *European Economic Review* 36:1001-1011.

Eichhorn, W. 1978. *Functional Equations in Economics*, London: Addison-Wesley.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin. 2000. "The Generalized Dynamic Factor Model: Identification and Estimation." *The Review of Economics and Statistics* 82: 540-554.

Forni, M., and M. Lippi. 2001. "The Generalized Factor Model: Representation Theory." *Econometric Theory* 17: 1113-1141.

Granger, C.W.J. 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Method." *Econometrica* 37(3):424-438.

______ . 1980. "Testing for Causality, A Personal Viewpoint." *Journal of Economic Dynamic and Control* 2(4):329-352.

Green, W.H. 2003. *Econometric Analysis*, 5th ed. New Jersey: Prentice-Hall.

Griliches, Z. 1972. "Effects of Aggregation over Time: comments" In B.G. Heckman, eds. *Econometric Models of Cyclical Behavior*, *Vol. II.*, New York: Columbia University Press.

Grunfeld, Y., and Z. Griliches. 1960. "Is Aggregation necessarily Bad?" *Review of Economics and Statistics* 42:1-13.

Hamilton, J.D. 1994. *Time Series Analysis*, Princeton: Princeton University Press.

Hanson, M.S. 2004. "The Price Puzzle Reconsidered." *Journal of Monetary Economics* 51:1385-1413.

Hausman, J.A. 1978. "Specification Tests in Econometrics." *Econometrica* 46:1251-1271.

Heaton, C., and V. Solo. 2006. "Estimation of Approximate Factor Models: Is it Important to have a Large Number of Variables?" Department of Economics, Macquarie University, Research Papers No. 0605.

Heckerman, D. 1996. “A Tutorial on Learning Bayesian Networks.” Microsoft Research, Technical Report MSR-TR-95-06.

Hill, R.J. 2006. ”Superlative Index Numbers: Not All of them are Super.” *Journal of Econometrics* 130:25-43.

Hillinger, C. 2002. “A General Theory of Price and Quantity Aggregation and Welfare Measurement.” CESIFO Working Paper No. 818.

Holland, P.W. 1986. “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81(396):945-960.

Hoover, K.D. 2006. “Causality in Economics and Econometrics,” forthcoming In S. Durlauf, eds. *The New Palgrave Dictiontionary of Economics*, 2nd ed. New York: Macmillan.

Houck, J.P. 1966. “A Look at Flexibilities and Elasticities.” *Journal of Farm Economics* 50:225-232.

Huang, K.S. 1994. “A Further Look at Flexibilities and Elasticities.” *American Journal of Agricultural Economics* 76:313-317.

Hume, D. 2000. *An Enquiry Concerning Human Understanding*, 1711-1776. Kitchener, Ontario, Canada: Batoche Books.

Johnson R.A., and D.W. Wichern. 1988. *Applied Multivariate Statistical Analysis*. 2nd ed. New Jersey: Prentice Hall.

Jolliffe, I.T. 1972. “Discarding Variables in a Principal Component Analysis I: Artificial Data.” *Applied Statistics* 21:160-173.

Juks, R. 2004. “Monetary Policy Transmission Mechanisms: A Theoretical and Empirical Overview.” In G.M. David, eds. *The Monetary Transmission Mechanism in the Baltic States*.

Keller, W.J., and J. van Driel. 1985. “Differential Consumer Demand Analysis.” *European Economic Review* 27:375-390.

Kim, C.J., and R.C. Nelson. 1997. “Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model with Regime Switching.” *The Review of Economics and Statistics* 80(2):188-201.

Kim, J.H., and J. Pearl. 1983. “A Computational Model for Combined Causal and Diagnostic Reasoning in Inference Systems.” Cognitive Systems Laboratory, University of California, Los Angeles, Technical Report R-25.

Kocka, T., R.R. Bouckaert, and M. Studeny. 2001. “On Characterizing Inclusion of Bayesian Networks.” In Breese, J. and D. Koller, eds. *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pp. 261-268.

Koopmans, T.C. 1947. “Measurement Without Theory.” *The Review of Economic Statistics* 29(3):161-172.

Lagana, G., and A. Mountford. 2005. “Measuring Monetary Policy in the U.K.: A Factor-Augmented Vector Autoregression Model Approach.” *Manchester School* 73(1):77-98.

Leeper E.M., C.A. Sims, and T. Zha. 1996. “What Does Monetary Policy Do?” *Brookings Papers on Economic Activity* 2:1-78.

Lewbel, A. 1996. “Aggregation Without Separability: A Generalized Composite Commodity Theorem.” *American Economic Review* 86:524-543.

Lewis, D. 1986. *Philosophical Papers*, New York: Oxford University Press.

Matsuda T. 2005. “Forms of Scale Curves and Differential Inverse Demand Systems.” *American Journal of Agricultural Economics* 87:786-795.

______ . 2004. “Incorporating Generalized Marginal Budget Shapes in a Mixed Demand System.” *American Journal of Agricultural Economics* 86:1117-1126.

McCabe, G.P. 1984. “Principal Variables.” *Technometrics* 26:137-144.

Meek, C. 1995a. “Causal Inference and Causal Explanation with Background Knowledge.” In S. Hanks and P. Besnard, eds. *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, pp. 403-418.

______ . 1995b. “Strong Completeness and Faithfulness in Bayesian Networks.” In S. Hanks and P. Besnard, eds. *Proceedings of 11th Conference on Uncertainty in Artificial Intelligence*, pp.411-418.

______ . 1997. Graphical Models: Selecting Causal and Statistical Models. Ph.D. thesis, Carnegie Mellon University.

Mishkin, S.F. 1995. “Symposium on the Monetary Transmission Mechanism.” *Journal of Economic Perspectives* 9(4):3-10.

______ . 2001. “The Transmission Mechanism and the Role of Asset Prices in Monetary Policy.” NBER Working Paper Series No.8617.

Moneta, A. 2007. “Mediating between Causes and Probabilities: the Use of Graphical Models in Econometrics.” In J. Williams and F. Russo, eds. *Causality and Probability in the Sciences*, London: College Publications, pp. 109-129.

Moneta, A., and P. Spirtes. 2006. “Graphical Models for the Identification of Causal Structures in Multivariate Time Series Models.” In H.D. Cheng, S.D. Chen, and R.Y. Lin, eds. *Proceedings of 9th Joint Conference on Information Sciences*, pp. 36-68.

Monteforte, L. 2007. “Aggregation Bias in Macro Models: Does it Matter for the Euro Area?” *Economic Modeling* 24:236-261.

Moschini, G., and A. Vissa. 1993. “Flexible Specification of Mixed Demand System.” *American Journal of Agricultural Economics* 75:1-9.

Neves, P. 1987. “Analysis of Consumer Demand in Portugal, 1958-1981.” Memorie de maitrise en sciences economiques, University Catholiqque de Louvrain, Louvain-la-Neuve.

Pearl, J. 1985. “Bayesian networks: A Model of Self-Activated Memory for Evidential Reasoning.” Department of Computer Science, University of California, Los Angeles, Technical Report No. 950021.

______ . 1988. *Probabilistic Reasoning in Intelligent Systems*, San Mateo: Morgan Kaufmann.

______ . 2000. *Causality: Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.

Pearl, J., and T. Verma. 1991. “A Theory Inferred Causation.” In J.A. Allen, R. Fikes, and E. Sandewall, eds. *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pp. 441-452.

Pollak, R.A., and T.J. Wales. 1991. “The Likelihood Dominance Criterion: A New Approach to Model Selection.” *Journal of Econometrics* 47:227-242.

Richardson, T., and P. Spirtes. 1999. “Automated Discovery of Linear Feedback Models.” In C. Glymour and G.F. Cooper, ed. *Computation, Causation and Discovery*, Menlo Park: MIT Press.

Robinson, R.W. 1977. “Counting Unlabeled Acyclic Digraphs.” In C.H.C. Little, eds. *Combinatorial Mathematics*, Springer Lecture Notes in Math 622, Berlin: Springer.

Rodrigues, J. 1999. “Classifying Interdependent Time Series in the Frequency Domain.” ECARES, Universite Libre de Bruxelles, Working Paper.

Rubinfeld, D.L. 2000. “Market Definition with Differentiated Products: The Post/Nabisco Cereal Merger.” *Antitrust Law Journal* 68:163-182.

Saha, A., R Shumway, and H. Talpaz. 1994. “Performance of Likelihood Dominance and Other Nonnested Model Selection Criteria: Some Monte Carlo Results.” Department of Agricultural Economics, Texas A&M University, Working Paper.

Samuelson, P.A. 1965. “Using Full Duality to Show that Simultaneously Additive Direct and Indirect Utilities Implies Unitary Price Elasticity of Demand.” *Econometrica* 33(4):781-796.

Samuelson, P.A., and S. Swamy. 1974. “Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis.” *American Economic Review* 64:566-593.

Sangüesa, R., and U. Cortés. 1997. “Learning Causal Networks from Data: A Survey and a New Algorithm for Recovering Possibilistic Causal Networks.” *AI Communications* 10:31-61.

Schultz, H. 1938. *The Theory and Measurement of Demand*, Illinois: University of Chicago Press.

Schwarz, G. 1978. “Estimating the Dimension of a Model.” *Annals of Statistics* 6:461-464.

Selvanathan, E.A., and D.S. Prasada Rao, 1994. *Index Numbers: A Stochastic Approach*, Ann Arbor: University of Michigan Press.

Shonkwiler, J.S., and T.G. Taylor. 1984. “The Implications of Estimating Market Demand Curves by Least Squares Regression.” *European Review of Agricultural Economics* 11:107-118.

Shumway, C.R., and G.C. Davis. 2001. "Does Consistent Aggregation really Matter?" *Australian Journal of Agricultural and Resource Economics* 45:161-194.

Sims, C.A. 1980. "Macroeconomics and Reality." *Econometrica* 48:1-48.

______ . 1992. "Interpreting the Macroeconomic Time Series Facts: the Effects of Monetary Policy." *European Economic Review* 36:975-1000.

Spirtes, P., and C. Meek. 1995. "Learning Bayesian networks with Discrete Variables form Data." In U.M. Fayyad and R. Uthurumsamy, eds. *Proceedings of 1st international Conference on Knowledge Discovery and Data Mining*, pp. 294-299.

Spirtes, P., C. Glymour, and R. Scheines. 2000. *Causation, Prediction, and Search*, 2nd ed. New York: MIT Press.

Stock, J.H., and M.W. Watson. 2002. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics* 20(2): 147-162.

______ . 2005. "Implications of Dynamic Factor Models for VAR Analysis." NBER Working Paper No.11467.

Stockton, M., O. Capps, Jr., and D.A. Bessler. 2005. "Samuelsons' Full Duality and the Use of Directed Acyclical Graphs: The Birth of Causality Identified Demand Systems," Department of Agricultural Economics, Texas A&M University, Working Paper.

Suppes, P. 1970. *A Probabilistic Theory of Causality*, Amsterdam: North-Holland.

Theil, H. 1954. *Linear Aggregation of Economic Relations*, Amsterdam: North-Holland.

______ . 1971. *Principles of Econometrics*. New York: John Wiley & Sons.

Thurman, W.N. 1986. “Endogeneity Testing in a Supply and Demand Framework.” *Review of Economics and Statistics* 68(4):638-646.

Verma, T., and J. Pearl. 1990. “Equivalence and Synthesis of Causal Models.” Cognitive Systems Laboratory, University of California, Los Angeles, Technical Report R-150.

______ . 1992. “An Algorithm for Deciding if a Set of Observed Independencies has a Causal Explanation.” Cognitive Systems Laboratory, University of California, Los Angeles, Technical Report R-177.

Watson, W.M. 2000. “Macroeconomic Forecasting Using Many Predictors.” In M. Dewatripont, L. Hansen, and S. Turnovsky, eds. *Advances in Economics and Econometrics, Theory and Applications, 8th World Congress of the Econometric Society*, *Vol. III*, pp. 87-115.

Wohlgenant, M.K., and W.F. Hahn. 1982. “Dynamic Adjustment in Monthly Consumer Demands for Meats,” *American Journal of Agricultural Economics* 64:553-557.

Xu, R., and D. Wunsch II. May, 2005. “Survey of Clustering Algorithms.” *IEEE Transactions on Neural Networks* 16(3):645-678.

APPENDIX A
PROPERTIES OF THREE DEMAND SYSTEMS

Direct Demand System

Theoretical implications for direct demand systems can be derived from properties of cost functions as follows:

(a) Homogeneity: the linear homogeneity of cost function in prices implies the zero-degree homogeneity of compensated demand in prices by Hotelling-Shephard lemma $\frac{\partial C(p,u)}{\partial p_n} = q_n^c(p, u)$, which in turn implies that $\sum_{n'=1}^{N} \frac{\partial q_n^c}{\partial p_{n'}} p_{n'} = 0 \cdot q_n^c(p,u) = 0$ by Euler's theorem. By multiplying $\frac{1}{q_n}$ to $\sum_{n'=1}^{N} \frac{\partial q_n^c}{\partial p_{n'}} p_{n'} = 0$, we get $\sum_{n'=1}^{N} \frac{\partial q_n^c}{\partial p_{n'}} \frac{p_{n'}}{q_n} = 0$ or $\sum_{n'=1}^{N} \varepsilon_{n,n'}^c = 0$ or $\sum_{n'=1}^{N} w_n \varepsilon_{n,n'}^c = 0$.

(b) Symmetry: the continuity of cost function implies the symmetry by Young's theorem $\frac{\partial^2 C}{\partial p_n \partial p_{n'}} = \frac{\partial^2 C}{\partial p_{n'} \partial p_n}$, which in turn implies that $\frac{\partial q_n^c}{\partial p_{n'}} = \frac{\partial q_{n'}^c}{\partial p_n}$. By multiplying $\left(\frac{p_n p_{n'}}{y}\right)$ both side of $\frac{\partial q_n^c}{\partial p_{n'}} = \frac{\partial q_{n'}^c}{\partial p_n}$, we get $\left(\frac{p_n q_n}{y}\right)\left(\frac{\partial q_n^c}{\partial p_{n'}} \frac{p_{n'}}{q_n}\right) = \left(\frac{p_{n'} q_{n'}}{y}\right)\left(\frac{\partial q_{n'}^c}{\partial p_n} \frac{p_n}{q_{n'}}\right)$ or $w_n \varepsilon_{n,n'}^c = w_{n'} \varepsilon_{n',n}^c$.

(c) Slutsky equation: By differentiating identity of $q^c(p,u) \equiv q(p,y) \equiv q[p, C(p,u)]$, we get $\frac{\partial q_n^c}{\partial p_{n'}} = \frac{\partial q_n}{\partial p_{n'}} + \frac{\partial q_n}{\partial C} \frac{\partial C}{\partial p_{n'}} = \frac{\partial q_n}{\partial p_{n'}} + \frac{\partial q_n}{\partial y} q_{n'}^c = \frac{\partial q_n}{\partial p_{n'}} + \frac{\partial q_n}{\partial y} q_{n'}$. By multiplying $\left(\frac{p_{n'}}{q_n}\right)$ to both side of $\frac{\partial q_n^c}{\partial p_{n'}} = \frac{\partial q_n}{\partial p_{n'}} + \frac{\partial q_n}{\partial y} q_{n'}$ to get the Slutsky equation in elasticity form, we obtain $\left(\frac{\partial q_n^c}{\partial p_{n'}} \frac{p_{n'}}{q_n}\right) = \left(\frac{\partial q_n}{\partial p_{n'}} \frac{p_{n'}}{q_n}\right) + \left(\frac{\partial q_n}{\partial y} \frac{y}{q_n}\right)\left(\frac{q_n}{y} \frac{p_{n'}}{q_n} q_{n'}\right) = \left(\frac{\partial q_n}{\partial p_{n'}} \frac{p_{n'}}{q_n}\right) + \left(\frac{\partial q_n}{\partial y} \frac{y}{q_n}\right)\left(\frac{p_{n'} q_{n'}}{y}\right)$ or $\varepsilon_{n,n'}^c = \varepsilon_{n,n'} + \varepsilon_n w_{n'}$ or $\varepsilon_{n,n'} = \varepsilon_{n,n'}^c - \varepsilon_n w_{n'}$.

(d) Adding-up: (d1) $\sum_{n=1}^{N} \varepsilon_n w_n = 1$ or (d2) $\sum_{n'=1}^{N} \varepsilon_{n,n'} = -\varepsilon_n$ or (d3) $\sum_{n=1}^{N} w_n \varepsilon_{n,n'} = -w_{n'}$

(d1) by differentiating $y \equiv \sum_{n=1}^{N} p_n \cdot q_n(p,y)$ with respect to (hereafter, w.r.t.) y, we have $\frac{\partial y}{\partial y} = \sum_{n=1}^{N} p_n \cdot \frac{\partial q_n}{\partial y} = \sum_{n=1}^{N} \left(\frac{\partial q_n}{\partial y} \frac{y}{q_n}\right)\left(\frac{q_n}{y} p_n\right)$ or $1 = \sum_{n=1}^{N} \varepsilon_n w_n$.

(d2) using the homogeneity condition $\sum_{n'=1}^{N}\varepsilon_{n,n'}^{c}=0$ in the Slutsky equation $\varepsilon_{n,n'}^{c}=\varepsilon_{n,n'}+\varepsilon_{n}w_{n'}$, we get $\sum_{n'=1}^{N}\varepsilon_{n,n'}^{c}=\sum_{n'=1}^{N}\varepsilon_{n,n'}+\varepsilon_{n}=0$ or $\sum_{n'=1}^{N}\varepsilon_{n,n'}=-\varepsilon_{n}$.

(d3) using the symmetry condition $w_{n}\varepsilon_{n,n'}^{c}=w_{n'}\varepsilon_{n',n}^{c}$ with the Slutsky equation $\varepsilon_{n,n'}^{c}=\varepsilon_{n,n'}+\varepsilon_{n}w_{n'}$, we get $w_{n}\left(\varepsilon_{n,n'}+\varepsilon_{n}w_{n'}\right)=w_{n'}\left(\varepsilon_{n',n}+\varepsilon_{n'}w_{n}\right)$ or $w_{n}\varepsilon_{n,n'}=w_{n'}\left(\varepsilon_{n',n}+\varepsilon_{n'}w_{n}-w_{n}\varepsilon_{n}\right)$. By summing up this relation, we get $\sum_{n=1}^{N}w_{n}\varepsilon_{n,n'}=w_{n'}\left(\sum_{n=1}^{N}\varepsilon_{n',n}+\varepsilon_{n'}\sum_{n=1}^{N}w_{n}-\sum_{n=1}^{N}w_{n}\varepsilon_{n}\right)$. Using above two adding-up conditions, we get $\sum_{n=1}^{N}w_{n}\varepsilon_{n,n'}=w_{n'}\left[\left(\sum_{n=1}^{N}\varepsilon_{n',n}+\varepsilon_{n'}\right)-\left(\sum_{n=1}^{N}w_{n}\varepsilon_{n}\right)\right]=w_{n'}\left[0-1\right]=-w_{n'}$ or $\sum_{n=1}^{N}w_{n}\varepsilon_{n,n'}=-w_{n'}$.

(e) Negativity: the concavity of cost function implies $\frac{\partial^{2}C}{\partial p_{n}^{2}}\leq 0$, which by Hotelling-Shephard lemma in turn implies $\frac{\partial q_{n}^{c}}{\partial p_{n}}\leq 0$. By multiplying $\frac{p_{n}}{q_{n}}$ to both side of $\frac{\partial q_{n}^{c}}{\partial p_{n}}\leq 0$, we get $\frac{\partial q_{n}^{c}}{\partial p_{n}}\frac{p_{n}}{q_{n}}\leq 0$, so $\varepsilon_{n,n}^{c}\leq 0$.

Inverse Demand System

Theoretical implications for inverse demand systems can be derived from properties of distance functions as follows:

(a) Homogeneity: the linear homogeneity of distance function in quantities implies the zero-degree homogeneity of compensated demand in quantities by Shephard-Hanoch lemma $\frac{\partial D\left(q,u^{0}\right)}{\partial q}=\pi^{C}\left(q,u^{0}\right)$, which in turn implies that $\sum_{n'=1}^{N}\frac{\partial\pi_{n}^{c}}{\partial q_{n'}}q_{n'}=0\cdot\pi_{n}^{c}\left(q,u\right)=0$ by Euler's theorem. By multiplying $\frac{1}{\pi_{n}}$ to $\sum_{n'=1}^{N}\frac{\partial\pi_{n}^{c}}{\partial q_{n'}}q_{n'}=0$, we get $\sum_{n'=1}^{N}\frac{\partial\pi_{n}^{c}}{\partial q_{n'}}\frac{q_{n'}}{\pi_{n}}=0$ or $\sum_{n'=1}^{N}f_{n,n'}^{c}=0$ or $\sum_{n'=1}^{N}w_{n}f_{n,n'}^{c}=0$.

(b) Symmetry: the continuity of distance function implies the symmetry by Young's theorem $\frac{\partial^{2}D}{\partial q_{n}\partial q_{n'}}=\frac{\partial^{2}D}{\partial q_{n'}\partial q_{n}}$, which in turn implies that $\frac{\partial\pi_{n}^{c}}{\partial q_{n'}}=\frac{\partial\pi_{n'}^{c}}{\partial q_{n}}$. By multiplying $q_{n}\cdot q_{n'}$ both side of $\frac{\partial\pi_{n}^{c}}{\partial q_{n'}}=\frac{\partial\pi_{n'}^{c}}{\partial q_{n}}$, we get $\left(\pi_{n}q_{n}\right)\left(\frac{\partial\pi_{n}^{c}}{\partial q_{n'}}\frac{q_{n'}}{\pi_{n}}\right)=\left(\pi_{n'}q_{n'}\right)\left(\frac{\partial\pi_{n'}^{c}}{\partial q_{n}}\frac{q_{n}}{\pi_{n'}}\right)$ or $w_{n}f_{n,n'}^{c}=w_{n'}f_{n',n}^{c}$.

(c) Antonelli equation: Using $q=k\cdot q^{*}$ where k is scalar and q^{*} is reference vector, we can define $u=U\left(q\right)=U\left(k\cdot q^{*}\right)=U^{*}\left(k,q^{*}\right)$ and $\pi_{n}=f^{n}\left(q\right)=f^{n}\left(k\cdot q^{*}\right)=g^{n}\left(k,q^{*}\right)$. First, by taking total differentiating to $u = U^*(k, q^*)$, we get $du = \sum_{n'}^{N} \frac{\partial U^*}{\partial kq^*_{n'}} \frac{\partial kq^*_{n'}}{\partial k} dk + \sum_{n'}^{N} \frac{\partial U^*}{\partial kq^*_{n'}} \frac{\partial kq^*_{n'}}{\partial q^*_{n'}} dq^*_{n'}$. Let $du = 0$ to compensate and using $q = k \cdot q^*$, we get $0 = \sum_{n'}^{N} \frac{\partial U}{\partial q_{n'}} \cdot q^*_{n'} \cdot dk + \sum_{n'}^{N} \frac{\partial U}{\partial q_{n'}} \cdot k \cdot dq^*_{n'}$, so $\sum_{n'}^{N} \frac{\partial U}{\partial q_{n'}} \cdot q^*_{n'} \cdot dk = -\left(\sum_{n'}^{N} \frac{\partial U}{\partial q_{n'}} \cdot k \cdot dq^*_{n'} \right)$ which is $\frac{dk}{dq^*_{n'}} = -\left(\sum_{n'}^{N} \frac{\partial U}{\partial q_{n'}} \Big/ \sum_{n'}^{N} \frac{\partial U}{\partial q_{n'}} \cdot q^*_{n'} \right) \cdot k = -\pi_{n'} \cdot k$ by Wold's lemma $\pi_n(q,1) = \frac{\partial U / \partial q_n}{\sum_{n'=1}^{N} (\partial U / \partial q_{n'}) \cdot q_{n'}}$. Thus we get $\left.\frac{dk}{dq^*_{n'}}\right|_{k=1} = -\pi_{n'}$. Second, by taking total differentiating to $\pi_n = g^n(k, q^*)$, we get $d\pi_n = \frac{\partial g^n(k, q^*)}{\partial q_{n'}} dq^*_{n'} + \frac{\partial g^n(k, q^*)}{\partial k} dk$ which is $\frac{d\pi_n}{dq^*_{n'}} = \frac{\partial g^n(k, q^*)}{\partial q_{n'}} + \frac{\partial g^n(k, q^*)}{\partial k} \frac{dk}{dq^*_{n'}}$, which in turn equal to $\frac{d\pi^c_n}{dq^*_{n'}} = \frac{\partial \pi_n}{\partial q_{n'}} - \frac{\partial \pi_n}{\partial k} \pi_{n'}$ by using first result $\left.\frac{dk}{dq^*_{n'}}\right|_{k=1} = -\pi_{n'}$. By multiplying $\frac{q_{n'}}{\pi_n}$ to both side of $\frac{d\pi^c_n}{dq^*_{n'}} = \frac{\partial \pi_n}{\partial q_{n'}} - \frac{\partial \pi_n}{\partial k} \pi_{n'}$, we get $\left(\frac{d\pi^c_n}{dq^*_{n'}} \frac{q_{n'}}{\pi_n} \right) = \left(\frac{\partial \pi_n}{\partial q_{n'}} \frac{q_{n'}}{\pi_n} \right) - \left(\frac{\partial \pi_n}{\partial k} \frac{k}{\pi_n} \right) \left(\frac{\pi_n}{k} \frac{q_{n'}}{\pi_n} \right) \pi_{n'}$ or $\left(\frac{d\pi^c_n}{dq^*_{n'}} \frac{q_{n'}}{\pi_n} \right) = \left(\frac{\partial \pi_n}{\partial q_{n'}} \frac{q_{n'}}{\pi_n} \right) - \left(\frac{\partial \pi_n}{\partial k} \frac{k}{\pi_n} \right) \cdot \pi_{n'} q^*_{n'}$ using $q = k \cdot q^*$. This relation is represented by $f^c_{n,n'} = f_{n,n'} - f_n \cdot w_{n'}$ or $f_{n,n'} = f^c_{n,n'} + f_n \cdot w_{n'}$.

(d) Adding-up: (d1) $-1 = \sum_{n=1}^{N} w_n f_n$ or (d2) $-w_{n'} = \sum_{n=1}^{N} f_{n,n'} w_n$ or (d3) $\sum_{n'=1}^{N} f_{n,n'} = f_n$

(d1) $w_n = \pi_n(q) \cdot q_n$ implies $\sum_{n=1}^{N} w_n = 1 = \sum_{n=1}^{N} \pi_n(q) \cdot q_n$. By differntiating $1 = \sum_{n=1}^{N} \pi_n(q) \cdot q_n$ w.r.t. q_n, we get $0 = \sum_{n=1}^{N} \frac{\partial \pi_n}{\partial q_{n'}} \cdot q_n + \pi_{n'}$. By multiplying $q_{n'}$, we get $0 \cdot q_{n'} = \sum_{n=1}^{N} \left(\frac{\partial \pi_n}{\partial q_{n'}} \cdot \frac{q_{n'}}{\pi_n} \right) (\pi_n q_n) + (\pi_{n'} q_{n'})$, which equal to $0 = \sum_{n=1}^{N} f_{n,n'} w_n + w_{n'}$ or $-w_{n'} = \sum_{n=1}^{N} f_{n,n'} w_n$.

(d2) using the homogeneity condition $\sum_{n'=1}^{N} f^c_{n,n'} = 0$ in the Antonelli equation of $f^c_{n,n'} = f_{n,n'} - f_n \cdot w_{n'}$, we get $\sum_{n'=1}^{N} f^c_{n,n'} = \sum_{n'=1}^{N} f_{n,n'} - f_n \cdot \sum_{n'=1}^{N} w_{n'} = \sum_{n'=1}^{N} f_{n,n'} - f_n = 0$ or $\sum_{n'=1}^{N} f_{n,n'} = f_n$. This relation can also be derived by using $\pi_n = f^n(q) = f^n(k \cdot q^*) = g^n(k, q^*)$. By definition of scale flexibility $f_n \equiv \frac{\partial g^n(k, q^*)}{\partial k} \frac{k}{g^n(k, q^*)}$, and using $\frac{\partial g^n(k, q^*)}{\partial k} = \sum_{n'=1}^{N} \frac{\partial f^n(k \cdot q^*)}{\partial kq^*_{n'}} \frac{\partial kq^*_{n'}}{\partial k}$ and $q = k \cdot q^*$, we get $f_n \equiv \frac{\partial g^n(k,q^*)}{\partial k}\frac{k}{g^n(k,q^*)} = \sum_{n'=1}^{N}\frac{\partial f^n(k\cdot q^*)}{\partial kq^*_{n'}}\frac{\partial kq^*_{n'}}{\partial k}\frac{k}{f^n(k\cdot q^*)} = \sum_{n'=1}^{N}\frac{\partial f^n(q)}{\partial q_{n'}}\cdot q^*_{n'}\cdot\frac{k}{f^n(q)}$ and we further get $f_n = \sum_{n'=1}^{N}\frac{\partial f^n(q)}{\partial q_{n'}}\frac{q_{n'}}{f^n(q)} = \sum_{n'=1}^{N}\frac{\partial \pi_n(q)}{\partial q_{n'}}\frac{q_{n'}}{\pi_n(q)} = \sum_{n'=1}^{N} f_{n,n'}$ or $f_n = \sum_{n'=1}^{N} f_{n,n'}$.

(d3) By summing up above first adding-up condition of $-w_{n'} = \sum_{n=1}^{N} f_{n,n'}w_n$ and using second adding-up condition of $\sum_{n'=1}^{N} f_{n,n'} = f_n$, we get $-\sum_{n'=1}^{N} w_{n'} = \sum_{n'=1}^{N}\sum_{n=1}^{N} f_{n,n'}w_n = \sum_{n=1}^{N} w_n \sum_{n'=1}^{N} f_{n,n'} = \sum_{n=1}^{N} w_n f_n$ or $-1 = \sum_{n=1}^{N} w_n f_n$

(e) Negativity: the concavity of distance function implies $\frac{\partial^2 D}{\partial q_n^2} \le 0$, which by Shephard-Hanoch lemma in turn implies $\frac{\partial \pi_n^c}{\partial q_n} \le 0$. By multiplying $\frac{q_n}{\pi_n}$ to both side of $\frac{\partial \pi_n^c}{\partial q_n} \le 0$, we get $\frac{\partial \pi_n^c}{\partial q_n}\frac{q_n}{\pi_n} \le 0$, so $f_{n,n}^c \le 0$.

Mixed Demand System

Theoretical implications for mixed demand systems can be derived from properties of restricted or rationed cost functions as follows:

(a) Homogeneity: the linear homogeneity of rationed cost function in prices p_A by Samuelson's envelope theorems implies that $q_i^C(p_A, q_B, u) = \frac{\partial C^R}{\partial p_i}$ is zero-homogeneous in price p_A and $p_k^C(p_A, q_B, u) = -\frac{\partial C^R}{\partial q_k}$ is linear homogenous in price p_A . By Euler's theorem, we get $\sum_{j=1}^{m}\frac{\partial q_i^c}{\partial p_j} p_j = 0\cdot q_i^c = 0$ and $\sum_{j=1}^{m}\frac{\partial p_k^c}{\partial p_j} p_j = 1\cdot p_k^c = p_k^c$. By multiplying $\frac{1}{q_i}$ to $\sum_{j=1}^{m}\frac{\partial q_i^c}{\partial p_j} p_j = 0$, we get $\sum_{j=1}^{m}\left(\frac{\partial q_i^c}{\partial p_j}\frac{p_j}{q_i}\right) = 0$ or $\sum_{j=1}^{m}\varepsilon_{i,j}^c = 0$ or $\sum_{j=1}^{m} w_i\cdot\varepsilon_{i,j}^c = 0$ and by using identity $p_k^c \equiv p_k$ for $\sum_{j=1}^{m}\frac{\partial p_k^c}{\partial p_j} p_j = p_k^c$,, we get $\sum_{j=1}^{m}\left(\frac{\partial p_k^c}{\partial p_j}\frac{p_j}{p_k}\right) = 1$ or $\sum_{j=1}^{m} p_{k,j}^c = 1$ or $\sum_{j=1}^{m} w_k\cdot p_{k,j}^c = w_k$

(b) Symmetry: the continuity of rationed cost function implies the symmetry by Young's theorem $\frac{\partial C^R}{\partial p_i \partial p_j} = \frac{\partial C^R}{\partial p_j \partial p_i}$, $\frac{\partial C^R}{\partial q_k \partial q_s} = \frac{\partial C^R}{\partial q_s \partial q_k}$, and $\frac{\partial C^R}{\partial p_i \partial q_k} = \frac{\partial C^R}{\partial q_k \partial p_i}$, which in turn implies that $\frac{\partial q_i^c}{\partial p_j} = \frac{\partial q_j^c}{\partial p_i}$, $\frac{\partial p_k^c}{\partial q_s} = \frac{\partial p_s^c}{\partial q_k}$, and $-\frac{\partial p_k^c}{\partial p_i} = \frac{\partial q_i^c}{\partial q_k}$ respectively. By multiplying $\left(\frac{p_i p_j}{y}\right)$, $\left(\frac{p_k p_s}{y}\right)$, and $\left(\frac{p_i p_k}{y}\right)$, we get $\left(\frac{p_i q_i}{y}\right)\left(\frac{\partial q_i^c}{\partial p_j}\frac{p_j}{q_i}\right) = \left(\frac{p_j q_j}{y}\right)\left(\frac{\partial q_j^c}{\partial p_i}\frac{p_i}{q_j}\right)$, $\left(\frac{p_k q_k}{y}\right)\left(\frac{\partial p_k^c}{\partial q_s}\frac{q_s}{p_k}\right) = \left(\frac{p_s q_s}{y}\right)\left(\frac{\partial p_s^c}{\partial q_k}\frac{q_k}{p_s}\right)$, and $-\left(\frac{p_k q_k}{y}\right)\left(\frac{\partial p_k^c}{\partial p_i}\frac{p_i}{p_k}\right) = \left(\frac{p_i q_i}{y}\right)\left(\frac{\partial q_i^c}{\partial q_k}\frac{q_k}{q_i}\right)$, which are represented by $w_i \cdot \varepsilon_{i,j}^c = w_j \cdot \varepsilon_{j,i}^c$, $w_k \cdot f_{k,s}^c = w_s \cdot f_{s,k}^c$, and $-w_k \cdot p_{k,i}^c = w_i \cdot q_{i,k}^c$ respectively.

(c) Slutsky equation: By differentiate identities of both $q_i^c(p_A, q_B, u) \equiv q_i\left[p_A, q_B, C^M(p_A, q_B, u)\right]$, and $p_k^c(p_A, q_B, u) \equiv p_k\left[p_A, q_B, C^M(p_A, q_B, u)\right]$ w.r.t. both p_j and q_s respectively and using derivative properties of mixed demand functions $\frac{\partial C^M}{\partial p_j} = \frac{\partial C^R}{\partial p_j} + \sum_{r=m+1}^{N} \frac{\partial p_r}{\partial p_j} q_r = q_j^c + \sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial p_j} q_r$ and $\frac{\partial C^M}{\partial q_k} = \frac{\partial C^R}{\partial q_k} + \sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial q_k} q_r + p_k = -p_k + \sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial q_k} q_r + p_k = \sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial q_k} q_r$, we get for group A (c1) $\frac{\partial q_i^c}{\partial p_j} = \frac{\partial q_i}{\partial p_j} + \frac{\partial q_i}{\partial y}\frac{\partial C^M}{\partial p_j} = \frac{\partial q_i}{\partial p_j} + \frac{\partial q_i}{\partial y}\left[q_j^c + \sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial p_j} q_r\right]$ which is $\varepsilon_{i,j}^c = \varepsilon_{i,j} + \varepsilon_i\left\lfloor w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right\rfloor$ through the relation of $\left(\frac{\partial q_i^c}{\partial p_j}\frac{p_j}{q_i}\right) = \left(\frac{\partial q_i}{\partial p_j}\frac{p_j}{q_i}\right) + \left(\frac{\partial q_i}{\partial y}\frac{y}{q_i}\right)\left[\left(\frac{p_j q_j^c}{y}\right) + \sum_{r=m+1}^{N}\left(\frac{p_r q_r}{y}\right)\left(\frac{\partial p_r^c}{\partial p_j}\frac{p_j}{p_r}\right)\right]$, and (c2) $\frac{\partial q_i^c}{\partial q_s} = \frac{\partial q_i}{\partial q_s} + \frac{\partial q_i}{\partial y}\frac{\partial C^M}{\partial q_s} = \frac{\partial q_i}{\partial p_j} + \frac{\partial q_i}{\partial y}\left[\sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial q_s} q_r\right]$ which is $q_{i,s}^c = q_{i,s} + \varepsilon_i \left\lfloor \sum_{r=m+1}^{N} w_r \cdot f_{r,s}^c \right\rfloor$ through the relation of $\left(\frac{\partial q_i^c}{\partial q_s}\frac{q_s}{q_i}\right) = \left(\frac{\partial q_i}{\partial q_s}\frac{q_s}{q_i}\right) + \left(\frac{\partial q_i}{\partial y}\frac{y}{q_i}\right)\left[\sum_{r=m+1}^{N}\left(\frac{p_r q_r}{y}\right)\left(\frac{\partial p_r^c}{\partial q_s}\frac{q_s}{p_r}\right)\right]$, and for group B (c3) $\frac{\partial p_k^c}{\partial p_j} = \frac{\partial p_k}{\partial p_j} + \frac{\partial p_k}{\partial y}\frac{\partial C^M}{\partial p_j} = \frac{\partial p_k}{\partial p_j} + \frac{\partial p_k}{\partial y}\left[q_j^c + \sum_{r=m+1}^{N} \frac{\partial p_r^c}{\partial p_j} q_r\right]$ which is $p_{k,j}^c = p_{k,j} + f_k \left\lfloor w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right\rfloor$ through the relation of $\left(\frac{\partial p_k^c}{\partial p_j}\frac{p_j}{p_k}\right) = \left(\frac{\partial p_k}{\partial p_j}\frac{p_j}{p_k}\right) + \left(\frac{\partial p_k}{\partial y}\frac{y}{p_k}\right)\left[\left(\frac{p_j q_j^c}{y}\right) + \sum_{r=m+1}^{N}\left(\frac{p_r q_r}{y}\right)\left(\frac{\partial p_r^c}{\partial p_j}\frac{p_j}{p_r}\right)\right]$, and (c4) $\frac{\partial p_k^c}{\partial q_s} = \frac{\partial p_k}{\partial q_s} + \frac{\partial p_k}{\partial y}\frac{\partial C^M}{\partial q_s} = \frac{\partial p_k}{\partial q_s} + \frac{\partial p_k}{\partial y}\left[\sum_{r=m+1}^{N}\frac{\partial p_r^c}{\partial q_s}q_r\right]$ which is $f_{k,s}^c = f_{k,s} + f_k\left[\sum_{r=m+1}^{N} w_r \cdot f_{r,s}^c\right]$

through the relation of $\left(\frac{\partial p_k^c}{\partial q_s}\frac{q_s}{p_k}\right) = \left(\frac{\partial p_k}{\partial q_s}\frac{q_s}{p_k}\right) + \left(\frac{\partial p_k}{\partial y}\frac{y}{p_k}\right)\left[\sum_{r=m+1}^{N}\left(\frac{p_r q_r}{y}\right)\left(\frac{\partial p_r^c}{\partial q_s}\frac{q_s}{p_r}\right)\right]$.

(d) Adding-up: (d1) $\sum_{i=1}^{m} w_i\varepsilon_i + \sum_{k=m+1}^{N} w_k f_k = 1$ and (d2) $\sum_{i=1}^{m} w_i\varepsilon_{i,j}^c = 0$ and $\sum_{i=1}^{m} w_i q_{i,s}^c = -w_s$

or (d3) $\varepsilon_i = -\sum_{j=1}^{m}\varepsilon_{i,j}$ and (d4) $f_k = 1 - \sum_{j=1}^{m} p_{k,j}$.

(d1) By differentiating $\sum_{i=1}^{m} p_i q_i(p_A, q_B, y) + \sum_{k=m+1}^{N} p_k(p_A, q_B, y)q_k = y$ w.r.t. y , we get $\sum_{i=1}^{m} p_i \cdot \frac{\partial q_i}{\partial y} + \sum_{k=m+1}^{N}\frac{\partial p_k}{\partial y}\cdot q_k = \frac{\partial y}{\partial y} = 1$, which equal to $\sum_{i=1}^{m}\left(\frac{p_i q_i}{y}\right)\left(\frac{\partial q_i}{\partial y}\frac{y}{q_i}\right) + \sum_{k=m+1}^{N}\left(\frac{p_k q_k}{y}\right)\left(\frac{\partial p_k}{\partial y}\frac{y}{p_k}\right) = 1$

or $\sum_{i=1}^{m} w_i\varepsilon_i + \sum_{k=m+1}^{N} w_k f_k = 1$

(d2) By differentiating $\sum_{i=1}^{m} p_i q_i(p_A, q_B, y) + \sum_{k=m+1}^{N} p_k(p_A, q_B, y)q_k = y$ w.r.t. p_j we get $\sum_{i=1}^{m} p_i \cdot \frac{\partial q_i}{\partial p_j} + q_j + \sum_{k=m+1}^{N}\frac{\partial p_k}{\partial p_j}\cdot q_k = 0$ or $\sum_{i=1}^{m} p_i \cdot \frac{\partial q_i}{\partial p_j}\left(\frac{p_j}{y}\right) + \sum_{k=m+1}^{N}\frac{\partial p_k}{\partial p_j}\cdot q_k\left(\frac{p_j}{y}\right) = -q_j\left(\frac{p_j}{y}\right)$, which is equal to $\sum_{i=1}^{m}\left(\frac{p_i q_i}{y}\right)\left(\frac{\partial q_i}{\partial p_j}\frac{p_j}{q_i}\right) + \sum_{k=m+1}^{N}\left(\frac{p_k q_k}{y}\right)\left(\frac{\partial p_k}{\partial p_j}\frac{p_j}{p_k}\right) = -\left(\frac{p_j q_j}{y}\right)$ or $\sum_{i=1}^{m} w_i\varepsilon_{i,j} + \sum_{k=m+1}^{N} w_k p_{k,j} = -w_j$. And also by differentiating $\sum_{i=1}^{m} p_i q_i(p_A, q_B, y) + \sum_{k=m+1}^{N} p_k(p_A, q_B, y)q_k = y$ w.r.t. q_s we get $\sum_{i=1}^{m} p_i \cdot \frac{\partial q_i}{\partial q_s} + \sum_{k=m+1}^{N}\frac{\partial p_k}{\partial q_s}\cdot q_k + p_s = 0$ or $\sum_{i=1}^{m} p_i \cdot \frac{\partial q_i}{\partial q_s}\left(\frac{q_s}{y}\right) + \sum_{k=m+1}^{N}\frac{\partial p_k}{\partial q_s}\cdot q_k\left(\frac{q_s}{y}\right) = -p_s\left(\frac{q_s}{y}\right)$, which is equal to $\sum_{i=1}^{m}\left(\frac{p_i q_i}{y}\right)\left(\frac{\partial q_i}{\partial q_s}\frac{q_s}{q_i}\right) + \sum_{k=m+1}^{N}\left(\frac{p_k q_k}{y}\right)\left(\frac{\partial p_k}{\partial q_s}\frac{q_s}{p_k}\right) = -\left(\frac{p_s q_s}{y}\right)$ or $\sum_{i=1}^{m} w_i q_{i,s} + \sum_{k=m+1}^{N} w_k f_{k,s} = -w_s$. Adding both results, we get $\sum_{i=1}^{m} w_i\varepsilon_{i,j} + \sum_{i=1}^{m} w_i q_{i,s} + \sum_{k=m+1}^{N} w_k p_{k,j} + \sum_{k=m+1}^{N} w_k f_{k,s} = -w_j - w_s$. Using Slutsky equations, $-w_j - w_s = \sum_{i=1}^{m} w_i\left(\varepsilon_{i,j}^c - \varepsilon_i w_j - \varepsilon_i\sum_{r=m+1}^{N} w_r p_{r,j}^c\right) + \sum_{i=1}^{m} w_i\left(q_{i,s}^c - \varepsilon_i\sum_{r=m+1}^{N} w_r f_{r,s}^c\right) + \sum_{k=m+1}^{N} w_k\left(p_{k,j}^c - f_k w_j - f_k\sum_{r=m+1}^{N} w_r p_{r,j}^c\right)$ $+ \sum_{k=m+1}^{N} w_k\left(f_{k,s}^c - f_k\sum_{r=m+1}^{N} w_r f_{r,s}^c\right)$ or $-w_j - w_s = \sum_{i=1}^{m} w_i\varepsilon_{i,j}^c + \sum_{i=1}^{m} w_i q_{i,s}^c - w_j\left(\sum_{i=1}^{m} w_i\varepsilon_i + \sum_{k=m+1}^{N} w_k f_k\right) + \sum_{r=m+1}^{N} w_r p_{r,j}^c$ $-\left(\sum_{i=1}^{m} w_i\varepsilon_i + \sum_{k=m+1}^{N} w_k f_k\right)\sum_{r=m+1}^{N} w_r p_{r,j}^c + \sum_{r=m+1}^{N} w_r f_{r,s}^c - \sum_{r=m+1}^{N} w_r f_{r,s}^c\left(\sum_{i=1}^{m} w_i\varepsilon_i + \sum_{k=m+1}^{N} w_k f_k\right)$. Using above adding-up condition $\sum_{i=1}^{m} w_i \varepsilon_i + \sum_{k=m+1}^{N} w_k f_k = 1$, we get $-w_j - w_s = \sum_{i=1}^{m} w_i \varepsilon_{i,j}^c + \sum_{i=1}^{m} w_i q_{i,s}^c - w_j$, which implies that $\sum_{i=1}^{m} w_i q_{i,s}^c = -w_s$ since $\sum_{i=1}^{m} w_i \varepsilon_{i,j}^c = \sum_{i=1}^{m} w_j \varepsilon_{j,i}^c = 0$ using symmetry $w_i \cdot \varepsilon_{i,j}^c = w_j \cdot \varepsilon_{j,i}^c$.

(d3) By using Slutsky equation $\varepsilon_{i,j}^c = \varepsilon_{i,j} + \varepsilon_i \left(w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right)$, we can write $\sum_{j=1}^{m} \varepsilon_{i,j}^c = 0$ as $\sum_{j=1}^{m} \left[\varepsilon_{i,j} + \varepsilon_i \left(w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right) \right] = 0$ or $\sum_{j=1}^{m} \varepsilon_{i,j} + \varepsilon_i \left[\sum_{j=1}^{m} w_j + \sum_{r=m+1}^{N} w_r \cdot \sum_{j=1}^{m} p_{r,j}^c \right] = 0$, which , by $\sum_{j=1}^{m} p_{k,j}^c = 1$, is $\varepsilon_i = -\sum_{j=1}^{m} \varepsilon_{i,j}$.

(d4) By using Slutsky equation $p_{k,j}^c = p_{k,j} + f_k \left(w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right)$, we can write $\sum_{j=1}^{m} p_{k,j}^c = 1$ as $\sum_{j=1}^{m} \left[p_{k,j} + f_k \left(w_j + \sum_{r=m+1}^{N} w_r \cdot p_{r,j}^c \right) \right] = 1$ or $\sum_{j=1}^{m} p_{k,j} + f_k \left[\sum_{j=1}^{m} w_j + \sum_{r=m+1}^{N} w_r \cdot \sum_{j=1}^{m} p_{r,j}^c \right] = 1$, which , by $\sum_{j=1}^{m} p_{k,j}^c = 1$, is $f_k = 1 - \sum_{j=1}^{m} p_{k,j}$.

(e) Negativity: the concavity w.r.t. p_A and convexity w.r.t. q_B of the rationed cost function implie $\frac{\partial^2 C^R}{\partial p_i^{\,2}} \le 0$ and $\frac{\partial^2 C^R}{\partial q_k^{\,2}} \ge 0$, which in turn imply that $\frac{\partial q_i^c}{\partial p_i} \le 0$ and $\frac{\partial p_k^c}{\partial q_k} \le 0$ respectively by Samuelson's envelope theorems of $q_i^C(p_A, q_B, u) = \frac{\partial C^R}{\partial p_i}$ and $p_k^C(p_A, q_B, u) = -\frac{\partial C^R}{\partial q_k}$.

APPENDIX B
RELATIONS AMONG THREE DEMAND SYSTEMS

Retrieval of Direct Elasticities from Mixed Elasticities

Direct demand system is related to mixed demand system by using following identities: $q_A^O\left[p_A, p_B^M\left(p_A, q_B, y\right), y\right] \equiv q_A^M\left(p_A, q_B, y\right)$ and $q_B^O\left[p_A, p_B^M\left(p_A, q_B, y\right), y\right] \equiv \overline{q_B^M}$. From identity of $q_A^O\left[p_A, p_B^M\left(p_A, q_B, y\right), y\right] \equiv q_A^M\left(p_A, q_B, y\right)$, (a) by differentiating identity w.r.t. ∇q_B, we get $\frac{\nabla q_A^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla q_B} = \frac{\nabla q_A^M}{\nabla q_B}$ or $\frac{\nabla q_A^O}{\nabla p_B} = \frac{\nabla q_A^M}{\nabla q_B}\left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}$, which can be written as $\left(\frac{\nabla q_A^O}{\nabla p_B}\frac{p_B}{q_A}\right) = \left(\frac{\nabla q_A^M}{\nabla q_B}\frac{q_B}{q_A}\right)\left(\frac{\nabla p_B^M}{\nabla q_B}\frac{q_B}{p_B}\right)^{-1}$ or $E_{AB}^O = Q_{AB}^M \cdot \left(F_{BB}^M\right)^{-1}$, (b) by differentiating w.r.t. ∇p_B, we get $\frac{\nabla q_A^O}{\nabla p_A} + \frac{\nabla q_A^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla p_A} = \frac{\nabla q_A^M}{\nabla p_A}$ or $\frac{\nabla q_A^O}{\nabla p_A} = \frac{\nabla q_A^M}{\nabla p_A} - \frac{\nabla q_A^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla p_A}$ which, using $\frac{\nabla q_A^O}{\nabla p_B} = \frac{\nabla q_A^M}{\nabla q_B}\left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}$, can be written as $\frac{\nabla q_A^O}{\nabla p_A} = \frac{\nabla q_A^M}{\nabla p_A} - \left[\frac{\nabla q_A^M}{\nabla q_B}\left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}\right]\frac{\nabla p_B^M}{\nabla p_A}$ or $E_{AA}^O = E_{AA}^M - Q_{AB}^M \cdot \left(F_{BB}^M\right)^{-1} P_{BA}^M$ through the relation of $\left(\frac{\nabla q_A^O}{\nabla p_A}\frac{p_A}{q_A}\right) = \left(\frac{\nabla q_A^M}{\nabla p_A}\frac{p_A}{q_A}\right) - \left(\frac{\nabla q_A^M}{\nabla q_B}\frac{q_B}{q_A}\right)\left(\frac{\nabla p_B^M}{\nabla q_B}\frac{q_B}{p_B}\right)^{-1}\left(\frac{\nabla p_B^M}{\nabla p_A}\frac{p_A}{p_B}\right)$, and (c) by differentiating w.r.t. ∇y, we also get $\frac{\nabla q_A^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla y} + \frac{\nabla q_A^O}{\nabla y} = \frac{\nabla q_A^M}{\nabla y}$ or $\frac{\nabla q_A^O}{\nabla y} = \frac{\nabla q_A^M}{\nabla y} - \frac{\nabla q_A^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla y}$, which, using $\frac{\nabla q_A^O}{\nabla p_B} = \frac{\nabla q_A^M}{\nabla q_B}\left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}$ again, can be written as $\frac{\nabla q_A^O}{\nabla y} = \frac{\nabla q_A^M}{\nabla y} - \left[\frac{\nabla q_A^M}{\nabla q_B}\left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}\right]\frac{\nabla p_B^M}{\nabla y}$ or $E_A^O = E_A^M - Q_{AB}^M \cdot \left(F_{BB}^M\right)^{-1} F_B^M$ through $\left(\frac{\nabla q_A^O}{\nabla y}\frac{y}{q_A}\right) = \left(\frac{\nabla q_A^M}{\nabla y}\frac{y}{q_A}\right) - \left(\frac{\nabla q_A^M}{\nabla q_B}\frac{q_B}{q_A}\right)\left(\frac{\nabla p_B^M}{\nabla q_B}\frac{q_B}{p_B}\right)^{-1}\left(\frac{\nabla p_B^M}{\nabla y}\frac{y}{p_B}\right)$.

From identities of $q_B^O\left[p_A, p_B^M\left(p_A, q_B, y\right), y\right] \equiv \overline{q_B^M}$, (a) by differentiating w.r.t. ∇q_B, we get $\frac{\nabla q_B^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla q_B} = 1$ or $\frac{\nabla q_B^O}{\nabla p_B} = \left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}$, which equal to $\left(\frac{\nabla q_B^O}{\nabla p_B}\frac{p_B}{q_B}\right) = \left(\frac{\nabla p_B^M}{\nabla q_B}\frac{q_B}{p_B}\right)^{-1}$ or $E_{BB}^O = \left(F_{BB}^M\right)^{-1}$, (b) by differentiating w.r.t. ∇p_A, we get $\frac{\nabla q_B^O}{\nabla p_A} + \frac{\nabla q_B^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla p_A} = 0$ or $\frac{\nabla q_B^O}{\nabla p_A} = -\frac{\nabla q_B^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla p_A}$, which, using

$\frac{\nabla q_B^O}{\nabla p_B} = \left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}$, can be written as $\frac{\nabla q_B^O}{\nabla p_A} = -\left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1} \frac{\nabla p_B^M}{\nabla p_A}$ or $E_{BA}^O = -\left(F_{BB}^M\right)^{-1} P_{BA}^M$ through $\left(\frac{\nabla q_B^O}{\nabla p_A}\frac{p_A}{q_B}\right) = -\left(\frac{\nabla p_B^M}{\nabla q_B}\frac{q_B}{p_B}\right)^{-1}\left(\frac{\nabla p_B^M}{\nabla p_A}\frac{p_A}{p_B}\right)$, and (c) by differentiating w.r.t. ∇y, we get $\frac{\nabla q_B^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla y} + \frac{\nabla q_B^O}{\nabla y} = 0$ or $\frac{\nabla q_B^O}{\nabla y} = -\frac{\nabla q_B^O}{\nabla p_B}\frac{\nabla p_B^M}{\nabla y}$, which, using $\frac{\nabla q_B^O}{\nabla p_B} = \left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}$ again, can be written as $\frac{\nabla q_B^O}{\nabla y} = -\left(\frac{\nabla p_B^M}{\nabla q_B}\right)^{-1}\frac{\nabla p_B^M}{\nabla y}$ or $E_B^O = -\left(F_{BB}^M\right)^{-1} F_B^M$ through the relation of $\left(\frac{\nabla q_B^O}{\nabla y}\frac{y}{q_B}\right) = -\left(\frac{\nabla p_B^M}{\nabla q_B}\frac{q_B}{p_B}\right)^{-1}\left(\frac{\nabla p_B^M}{\nabla y}\frac{y}{p_B}\right)$.

Retrieval of Inverse Flexibilities from Mixed Elasticities

Inverse demand system is related to mixed demand system by using following identities: $p_A^I\left[q_A^M\left(p_A, q_B, y\right), q_B, y\right] \equiv \overline{p_A}$ and $p_B^I\left[q_A^M\left(p_A, q_B, y\right), q_B, y\right] \equiv p_B^M\left(p_A, q_B, y\right)$ which are implied by $\pi_A^I\left[q_A^M\left(\pi_A, q_B, 1\right), q_B, 1\right] \equiv \overline{\pi_A}$ and $\pi_B^I\left[q_A^M\left(\pi_A, q_B, 1\right), q_B, 1\right] \equiv \pi_B^M\left(\pi_A, q_B, 1\right)$ through the relationships of $\pi_A^I\left[q_A^M\left(\pi_A, q_B, 1\right), q_B, 1\right] \cdot y \equiv \overline{\pi_A} \cdot y$ and $\pi_B^I\left[q_A^M\left(\pi_A, q_B, 1\right), q_B, 1\right] \cdot y \equiv \pi_B^M\left(\pi_A, q_B, 1\right) \cdot y$. From identities of $p_A^I\left[q_A^M\left(p_A, q_B, y\right), q_B, y\right] \equiv \overline{p_A}$, (a) by differentiating w.r.t. ∇p_A, we get $\frac{\nabla p_A^I}{\nabla q_A}\frac{\nabla q_A^M}{\nabla p_A} = 1$ or $\frac{\nabla p_A^I}{\nabla q_A} = \left(\frac{\nabla q_A^M}{\nabla p_A}\right)^{-1}$, which equals to $\left(\frac{\nabla p_A^I}{\nabla q_A}\frac{q_A}{p_A}\right) \equiv \left(\frac{\nabla \pi_A^I}{\nabla q_A}\frac{q_A}{\pi_A}\right) = \left(\frac{\nabla q_A^M}{\nabla p_A}\frac{p_A}{q_A}\right)^{-1}$ or $F_{AA}^I = \left(E_{AA}^M\right)^{-1}$, (b) by differentiating w.r.t. ∇q_B, we get $\frac{\nabla p_A^I}{\nabla q_A}\frac{\nabla q_A^M}{\nabla q_B} + \frac{\nabla p_A^I}{\nabla q_B} = 0$ or $\frac{\nabla p_A^I}{\nabla q_B} = -\frac{\nabla p_A^I}{\nabla q_A}\frac{\nabla q_A^M}{\nabla q_B}$, which, using $\frac{\nabla p_A^I}{\nabla q_A} = \left(\frac{\nabla q_A^M}{\nabla p_A}\right)^{-1}$, can be written as $\frac{\nabla p_A^I}{\nabla q_B} = -\left(\frac{\nabla q_A^M}{\nabla p_A}\right)^{-1}\frac{\nabla q_A^M}{\nabla q_B}$ or $\left(\frac{\nabla p_A^I}{\nabla q_B}\frac{q_B}{p_A}\right) \equiv \left(\frac{\nabla \pi_A^I}{\nabla q_B}\frac{q_B}{\pi_A}\right) = -\left(\frac{\nabla q_A^M}{\nabla p_A}\frac{p_A}{q_A}\right)^{-1}\left(\frac{\nabla q_A^M}{\nabla q_B}\frac{q_B}{q_A}\right)$, which in turn equal to $F_{AB}^I = -\left(E_{AA}^M\right)^{-1} Q_{AB}^M$. From identity of $p_B^I\left[q_A^M\left(p_A, q_B, y\right), q_B, y\right] \equiv p_B^M\left(p_A, q_B, y\right)$, (a) by differentiating identity w.r.t. ∇p_A, we get $\frac{\nabla p_B^I}{\nabla q_A}\frac{\nabla q_A^M}{\nabla p_A} = \frac{\nabla p_B^M}{\nabla p_A}$ or $\frac{\nabla p_B^I}{\nabla q_A} = \frac{\nabla p_B^M}{\nabla p_A}\left(\frac{\nabla q_A^M}{\nabla p_A}\right)^{-1}$ which can be written as

$\left(\frac{\nabla p_B^I}{\nabla q_A}\frac{q_A}{p_B}\right) \equiv \left(\frac{\nabla \pi_B^I}{\nabla q_A}\frac{q_A}{\pi_B}\right) = \left(\frac{\nabla p_B^M}{\nabla p_A}\frac{p_A}{p_B}\right)\left(\frac{\nabla q_A^M}{\nabla p_A}\frac{p_A}{q_A}\right)^{-1}$ or $F_{BA}^I = P_{BA}^M \left(E_{AA}^M\right)^{-1}$, (b) by differentiating w.r.t. ∇q_B, we get $\frac{\nabla p_B^I}{\nabla q_A}\frac{\nabla q_A^M}{\nabla q_B} + \frac{\nabla p_B^I}{\nabla q_B} = \frac{\nabla p_B^M}{\nabla q_B}$ or $\frac{\nabla p_B^I}{\nabla q_B} = \frac{\nabla p_B^M}{\nabla q_B} - \frac{\nabla p_B^I}{\nabla q_A}\frac{\nabla q_A^M}{\nabla q_B}$ which, using $\frac{\nabla p_B^I}{\nabla q_A} = \frac{\nabla p_B^M}{\nabla p_A}\left(\frac{\nabla q_A^M}{\nabla p_A}\right)^{-1}$, can be written as $\frac{\nabla p_B^I}{\nabla q_B} = \frac{\nabla p_B^M}{\nabla q_B} - \left[\frac{\nabla p_B^M}{\nabla p_A}\left(\frac{\nabla q_A^M}{\nabla p_A}\right)^{-1}\right]\frac{\nabla q_A^M}{\nabla q_B}$ or $\left(\frac{\nabla p_B^I}{\nabla q_B}\frac{q_B}{p_B}\right) \equiv \left(\frac{\nabla \pi_B^I}{\nabla q_B}\frac{q_B}{\pi_B}\right) = \left(\frac{\nabla p_B^M}{\nabla q_B}\frac{q_B}{p_B}\right) - \left(\frac{\nabla p_B^M}{\nabla p_A}\frac{p_A}{p_B}\right)\left(\frac{\nabla q_A^M}{\nabla p_A}\frac{p_A}{q_A}\right)^{-1}\left(\frac{\nabla q_A^M}{\nabla q_B}\frac{q_B}{q_A}\right)$, which in turn equal to $F_{BB}^I = F_{BB}^M - P_{BA}^M\left(E_{AA}^M\right)^{-1}Q_{AB}^M$. From the relation $f_n = \sum_{n'=1}^{N} f_{n,n'}$ or $F_N^I = RowSum\left(F_{N,N}^I\right)$ of inverse demand function, we get $F_A^I = RowSum\left(F_{AA}^I \vdots F_{AB}^I\right)$ and $F_B^I = RowSum\left(F_{BA}^I \vdots F_{BB}^I\right)$. Using $F_{AA}^I = \left(E_{AA}^M\right)^{-1}$ and $F_{AB}^I = -\left(E_{AA}^M\right)^{-1}Q_{AB}^M$, we can write $F_A^I = RowSum\left[\left(E_{AA}^M\right)^{-1} \vdots -\left(E_{AA}^M\right)^{-1}Q_{AB}^M\right]$. Using $F_{BA}^I = P_{BA}^M\left(E_{AA}^M\right)^{-1}$ and $F_{BB}^I = F_{BB}^M - P_{BA}^M\left(E_{AA}^M\right)^{-1}Q_{AB}^M$, we can write $F_B^I = RowSum\left[P_{BA}^M\left(E_{AA}^M\right)^{-1} \vdots F_{BB}^M - P_{BA}^M\left(E_{AA}^M\right)^{-1}Q_{AB}^M\right]$.

Retrieval of Mixed Elasticities from Direct Elasticities

Theoretical relationships of mixed elasticities to direct elasticities can be derived as follows. From $E_{BB}^O = \left(F_{BB}^M\right)^{-1}$, we get $F_{BB}^M = \left(E_{BB}^O\right)^{-1}$. From $E_{AB}^O = Q_{AB}^M \cdot \left(F_{BB}^M\right)^{-1}$, we get $Q_{AB}^M = E_{AB}^O F_{BB}^M = E_{AB}^O\left(E_{BB}^O\right)^{-1}$ using $F_{BB}^M = \left(E_{BB}^O\right)^{-1}$. From $E_{BA}^O = -\left(F_{BB}^M\right)^{-1}P_{BA}^M$, we get $P_{BA}^M = -F_{BB}^M E_{BA}^O = -\left(E_{BB}^O\right)^{-1}E_{BA}^O$ using $F_{BB}^M = \left(E_{BB}^O\right)^{-1}$. From $E_{AA}^O = E_{AA}^M - Q_{AB}^M \cdot \left(F_{BB}^M\right)^{-1}P_{BA}^M$, we get $E_{AA}^M = E_{AA}^O + Q_{AB}^M \cdot \left(F_{BB}^M\right)^{-1}P_{BA}^M = E_{AA}^O - \left[E_{AB}^O\left(E_{BB}^O\right)^{-1}\right]\cdot\left[E_{BB}^O\right]\cdot\left[\left(E_{BB}^O\right)^{-1}E_{BA}^O\right] = E_{AA}^O - E_{AB}^O\left(E_{BB}^O\right)^{-1}E_{BA}^O$ using $F_{BB}^M = \left(E_{BB}^O\right)^{-1}$, $Q_{AB}^M = E_{AB}^O\left(E_{BB}^O\right)^{-1}$, and $P_{BA}^M = -\left(E_{BB}^O\right)^{-1}E_{BA}^O$. From $E_A^O = E_A^M - Q_{AB}^M \cdot \left(F_{BB}^M\right)^{-1}F_B^M$, we get $E_A^M = E_A^O + Q_{AB}^M \cdot \left(F_{BB}^M\right)^{-1}F_B^M = E_A^O - \left[E_{AB}^O\left(E_{BB}^O\right)^{-1}\right]\cdot\left[E_{BB}^O\right]\cdot\left[\left(E_{BB}^O\right)^{-1}E_B^O\right] = E_A^O - E_{AB}^O\left(E_{BB}^O\right)^{-1}E_B^O$ using $F_{BB}^M = \left(E_{BB}^O\right)^{-1}$, $Q_{AB}^M = E_{AB}^O\left(E_{BB}^O\right)^{-1}$, and $P_{BA}^M = -\left(E_{BB}^O\right)^{-1}E_{BA}^O$. From $E_B^O = -\left(F_{BB}^M\right)^{-1}F_B^M$, we get $F_B^M = -F_{BB}^M E_B^O = -\left(E_{BB}^O\right)^{-1}E_B^O$ using $F_{BB}^M = \left(E_{BB}^O\right)^{-1}$.

Retrieval of Mixed Elasticities from Inverse Flexibilities

Theoretical relationships of mixed elasticities to inverse flexibilities can be derived as follows. From $F_{AA}^{I} = \left(E_{AA}^{M}\right)^{-1}$, we get $E_{AA}^{M} = \left(F_{AA}^{I}\right)^{-1}$. From $F_{BA}^{I} = P_{BA}^{M}\left(E_{AA}^{M}\right)^{-1}$, we get $P_{BA}^{M} = F_{BA}^{I}E_{AA}^{M}$ $= F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$ using $E_{AA}^{M} = \left(F_{AA}^{I}\right)^{-1}$. From $F_{AB}^{I} = -\left(E_{AA}^{M}\right)^{-1}Q_{AB}^{M}$, we get $Q_{AB}^{M} = -E_{AA}^{M}F_{AB}^{I} = -\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$ using $E_{AA}^{M} = \left(F_{AA}^{I}\right)^{-1}$. From $F_{BB}^{I} = F_{BB}^{M} - P_{BA}^{M}\left(E_{AA}^{M}\right)^{-1}Q_{AB}^{M}$, we also get $F_{BB}^{M} = F_{BB}^{I} + P_{BA}^{M}\left(E_{AA}^{M}\right)^{-1}Q_{AB}^{M}$ $= F_{BB}^{I} - \left[F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}\right]\cdot\left[F_{AA}^{I}\right]\cdot\left[\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right] = F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$ using $E_{AA}^{M} = \left(F_{AA}^{I}\right)^{-1}$, $P_{BA}^{M} = F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$, and $Q_{AB}^{M} = -\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$. From the relation $\varepsilon_{i} = -\sum_{j=1}^{m}\varepsilon_{i,j}$ and $f_{k} = 1 - \sum_{j=1}^{m}p_{k,j}$ of mixed demand functions, we get $E_{A}^{M} = -RowSum\left(E_{AA}^{M}\right)$ and $F_{B}^{M} = I - RowSum\left(P_{BA}^{M}\right)$. Using $E_{AA}^{M} = \left(F_{AA}^{I}\right)^{-1}$, we can write $E_{A}^{M} = -RowSum\left[\left(F_{AA}^{I}\right)^{-1}\right]$. Using $P_{BA}^{M} = F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$, we can write $F_{B}^{M} = I - RowSum\left[F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}\right]$.

Retrieval of Direct Elasticities from Inverse Flexibilities

Theoretical relationships of direct elasticities to inverse flexibilities can be derived as follows. From $E_{BB}^{O} = \left(F_{BB}^{M}\right)^{-1}$, we get $E_{BB}^{O} = \left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right]^{-1}$ using $F_{BB}^{M} = F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$. From $E_{AB}^{O} = Q_{AB}^{M}\cdot\left(F_{BB}^{M}\right)^{-1}$, we get $E_{AB}^{O} = -\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\cdot\left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right]^{-1}$ using $Q_{AB}^{M} = -\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$ and $F_{BB}^{M} = F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$. From $E_{BA}^{O} = -\left(F_{BB}^{M}\right)^{-1}P_{BA}^{M}$, we get $E_{BA}^{O} = -\left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right]^{-1}F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$ using $F_{BB}^{M} = F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$ and $P_{BA}^{M} = F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$. From $E_{AA}^{O} = E_{AA}^{M} - Q_{AB}^{M}\cdot\left(F_{BB}^{M}\right)^{-1}P_{BA}^{M}$, we get $E_{AA}^{O} = \left(F_{AA}^{I}\right)^{-1} + \left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\cdot\left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right]^{-1}F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$ using $E_{AA}^{M} = \left(F_{AA}^{I}\right)^{-1}$, $Q_{AB}^{M} = -\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$, $F_{BB}^{M} = F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}$, and $P_{BA}^{M} = F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$. From the relation $\varepsilon_{n} = -\sum_{n'=1}^{N}\varepsilon_{n,n'}$ or $E_{N}^{O} = -RowSum\left(E_{N,N}^{O}\right)$ of direct demand function, we can write $E_{A}^{O} = -RowSum\left(E_{AA}^{O} \vdots E_{AB}^{O}\right)$ and $E_{B}^{O} = -RowSum\left(E_{BA}^{O} \vdots E_{BB}^{O}\right)$. Using above relationships relating elasticity to flexibility $E_{AA}^{O} = \left(F_{AA}^{I}\right)^{-1} + \left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\cdot\left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right]^{-1}F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$ and $E_{AB}^{O} = -\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\cdot\left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right]^{-1}$ and $E_{BA}^{O} = -\left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right]^{-1}F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}$ and $E_{BB}^{O} = \left[F_{BB}^{I} - F_{BA}^{I}\left(F_{AA}^{I}\right)^{-1}F_{AB}^{I}\right]^{-1}$, we can relate expenditure elasticity to flexibilities as follows

$$E_A^O = -RowSum\left\langle \left(F_{AA}^I\right)^{-1} + \left(F_{AA}^I\right)^{-1} F_{AB}^I \cdot \left[F_{BB}^I - F_{BA}^I \left(F_{AA}^I\right)^{-1} F_{AB}^I\right]^{-1} F_{BA}^I \left(F_{AA}^I\right)^{-1} \vdots -\left(F_{AA}^I\right)^{-1} F_{AB}^I \cdot \left[F_{BB}^I - F_{BA}^I \left(F_{AA}^I\right)^{-1} F_{AB}^I\right]^{-1} \right\rangle$$

and $E_B^O = -RowSum\left\langle -\left[F_{BB}^I - F_{BA}^I \left(F_{AA}^I\right)^{-1} F_{AB}^I\right]^{-1} F_{BA}^I \left(F_{AA}^I\right)^{-1} \vdots \left[F_{BB}^I - F_{BA}^I \left(F_{AA}^I\right)^{-1} F_{AB}^I\right]^{-1} \right\rangle$.

Retrieval of Inverse Flexibilities from Direct Elasticities

Theoretical relationships of inverse flexibilities to direct elasticities can be derived as follows. From $F_{AA}^I = \left(E_{AA}^M\right)^{-1}$, we get $F_{AA}^I = \left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1}$ using $E_{AA}^M = E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O$. From $F_{BA}^I = P_{BA}^M \left(E_{AA}^M\right)^{-1}$, we get $F_{AB}^I = -\left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1} E_{AB}^O \left(E_{BB}^O\right)^{-1}$ using $P_{BA}^M = -\left(E_{BB}^O\right)^{-1} E_{BA}^O$ and $E_{AA}^M = E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O$. From $F_{AB}^I = -\left(E_{AA}^M\right)^{-1} Q_{AB}^M$, we get $F_{BA}^I = -\left(E_{BB}^O\right)^{-1} E_{BA}^O \left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1}$ using $E_{AA}^M = E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O$ and $Q_{AB}^M = E_{AB}^O \left(E_{BB}^O\right)^{-1}$. From $F_{BB}^I = F_{BB}^M - P_{BA}^M \left(E_{AA}^M\right)^{-1} Q_{AB}^M$, we get $F_{BB}^I = \left(E_{BB}^O\right)^{-1} + \left(E_{BB}^O\right)^{-1} E_{BA}^O \left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1} E_{AB}^O \left(E_{BB}^O\right)^{-1}$ using $F_{BB}^M = \left(E_{BB}^O\right)^{-1}$, $P_{BA}^M = -\left(E_{BB}^O\right)^{-1} E_{BA}^O$, $E_{AA}^M = E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O$, and $Q_{AB}^M = E_{AB}^O \left(E_{BB}^O\right)^{-1}$. From the relation $f_n = \sum_{n'=1}^{N} f_{n,n'}$ or $F_N^I = RowSum\left(F_{N,N}^I\right)$ of inverse demand function, we can write $F_A^I = RowSum\left(F_{AA}^I \vdots F_{AB}^I\right)$ and $F_B^I = RowSum\left(F_{BA}^I \vdots F_{BB}^I\right)$. Using above relations of elasticity to flexibility $F_{AA}^I = \left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1}$, $F_{AB}^I = -\left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1} E_{AB}^O \left(E_{BB}^O\right)^{-1}$, $F_{BA}^I = -\left(E_{BB}^O\right)^{-1} E_{BA}^O \left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1}$ and $F_{BB}^I = \left(E_{BB}^O\right)^{-1} + \left(E_{BB}^O\right)^{-1} E_{BA}^O \left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1} E_{AB}^O \left(E_{BB}^O\right)^{-1}$, we can relate scale flexibility of inverse demand to elasticities of direct demand as follows $F_A^I = RowSum\left[\left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1} \vdots -\left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1} E_{AB}^O \left(E_{BB}^O\right)^{-1}\right]$ and $F_B^I = RowSum\left[-\left(E_{BB}^O\right)^{-1} E_{BA}^O \left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1} \vdots \left(E_{BB}^O\right)^{-1} + \left(E_{BB}^O\right)^{-1} E_{BA}^O \left[E_{AA}^O - E_{AB}^O \left(E_{BB}^O\right)^{-1} E_{BA}^O\right]^{-1} E_{AB}^O \left(E_{BB}^O\right)^{-1}\right]$.

APPENDIX C
DIFFERENTIAL FAMILY OF THREE DEMAND SYSTEMS

For specifications of differential family of demand system, the log differential property of $dz = z \cdot d\ln z$ or $d\ln z = dz/z$ is frequently used for any variable z. For example, by taking total differentiate of identity $y \equiv \sum_{n=1}^{N} p_n q_n$, we get $dy \equiv \sum_{n=1}^{N} p_n dq_n + \sum_{n=1}^{N} q_n dp_n$, which can be written as $\left(yd\ln y\right) \equiv \sum_{n=1}^{N} p_n\left(q_n d\ln q_n\right) + \sum_{n=1}^{N} q_n\left(p_n d\ln p_n\right)$ by log differential property and represented as $d\ln y \equiv \sum_{n=1}^{N}\left(\frac{p_n q_n}{y}\right) d\ln q_n + \sum_{n=1}^{N}\left(\frac{p_n q_n}{y}\right) d\ln p_n$ or $d\ln y \equiv \sum_{n=1}^{N} w_n d\ln q_n + \sum_{n=1}^{N} w_n d\ln p_n \equiv d\ln Q + d\ln P$. Similarly by taking total differentiate of identity $y \equiv \sum_{i=1}^{m} p_i q_i + \sum_{k=m+1}^{N} p_k q_k$, we get $dy \equiv \sum_{i=1}^{m} p_i dq_i + \sum_{i=1}^{m} q_i dp_i + \sum_{k=m+1}^{N} p_k dq_k + \sum_{k=m+1}^{N} q_k dp_k$, which, using $dz = z \cdot d\ln z$, can be written as $\left(yd\ln y\right) \equiv \sum_{i=1}^{m} p_i\left(q_i d\ln q_i\right) + \sum_{i=1}^{m} q_i\left(p_i d\ln p_i\right) + \sum_{k=m+1}^{N} p_k\left(q_k d\ln q_k\right) + \sum_{k=m+1}^{N} q_k\left(p_k d\ln p_k\right)$ and represented as $d\ln y \equiv \sum_{i=1}^{m}\left(\frac{p_i q_i}{y}\right) d\ln q_i + \sum_{i=1}^{m}\left(\frac{p_i q_i}{y}\right) d\ln p_i + \sum_{k=m+1}^{N}\left(\frac{p_k q_k}{y}\right) d\ln q_k + \sum_{k=m+1}^{N}\left(\frac{p_k q_k}{y}\right) d\ln p_k$ by multiplying $\frac{1}{y}$ or $d\ln y \equiv \sum_{i=1}^{m} w_i d\ln q_i + \sum_{i=1}^{m} w_i d\ln p_i + \sum_{k=m+1}^{N} w_k d\ln q_k + \sum_{k=m+1}^{N} w_k d\ln p_k$ by budget share definition or $d\ln y \equiv d\ln Q_A + d\ln P_A + d\ln Q_B + d\ln P_B = \left[d\ln Q_A + d\ln Q_B + d\ln P_B\right] + d\ln P_A = d\ln \bar{y} + d\ln P_A$. For another example, by taking total differentiate of identity $w_n \equiv \frac{p_n q_n}{y}$, we get $dw_n = \frac{p_n}{y} dq_n + \frac{q_n}{y} dp_n - \frac{p_n q_n}{y^2} dy$, which, using the log differential property $dz = z \cdot d\ln z$, can be written as $dw_n = \left(\frac{p_n q_n}{y}\right) d\ln q_n + \left(\frac{q_n p_n}{y}\right) d\ln p_n - \left(\frac{p_n q_n}{y^2} \cdot y\right) d\ln y$ and, by budget share definition, can be also represented as $dw_n = w_n d\ln q_n + w_n d\ln p_n - w_n d\ln y$, which can be either $dw_n = w_n d\ln q_n + w_n d\ln \pi_n$ or $dw_n = w_n d\ln q_n + w_n d\ln p_n - w_n\left[d\ln Q + d\ln P\right]$. The Kronecker delta $\delta_{n,n'} = 1$ for $n = n'$ and $\delta_{n,n'} = 0$ for $n \neq n'$ is also frequently used. For example, $\sum_{n'=1}^{N} w_{n'} z_{n'} - z_n$ can be written as $\sum_{n'=1}^{N} w_{n'} z_{n'} - \sum_{n'=1}^{N} \delta_{n,n'} z_{n'} = \sum_{n'=1}^{N}\left(w_{n'} - \delta_{n,n'}\right) \cdot z_{n'}$.

Rotterdam Functional Form

Specification of the Rotterdam direct demand systems can be derived as follows. By taking total differentiation of the uncompensated direct demand $q_n = q_n(y, p_1, \cdots, p_N)$, we get $dq_n = \frac{\partial q_n}{\partial y} dy + \sum_{n'=1}^{N} \frac{\partial q_n}{\partial p_{n'}} dp_{n'}$. By using log differential property $dz = z \cdot d \ln z$, This equation can be written as $q_n d \ln q_n = \frac{\partial q_n}{\partial y} y d \ln y + \sum_{n'=1}^{N} \frac{\partial q_n}{\partial p_{n'}} p_{n'} d \ln p_{n'}$, which, by multiplying $\frac{p_n}{y}$, can be written as $\left(\frac{p_n q_n}{y}\right) d \ln q_n = \left(\frac{p_n q_n}{y}\right)\left(\frac{\partial q_n}{\partial y}\frac{y}{q_n}\right) d \ln y + \sum_{n'=1}^{N}\left(\frac{p_n q_n}{y}\right)\left(\frac{\partial q_n}{\partial p_{n'}}\frac{p_{n'}}{q_n}\right) d \ln p_{n'}$ or, by budget share definition, $w_n d \ln q_n = w_n \varepsilon_n d \ln y + \sum_{n'=1}^{N} w_n \varepsilon_{n,n'} d \ln p_{n'}$. By using Slutsky relation of $w_n \varepsilon_{n,n'} = w_n \varepsilon^c_{n,n'} - w_n \varepsilon_n w_{n'}$ into this equation, we have $w_n d \ln q_n = w_n \varepsilon_n d \ln y + \sum_{n'=1}^{N}\left[w_n \varepsilon^c_{n,n'} - w_n \varepsilon_n w_{n'}\right] d \ln p_{n'}$, which can be written as $w_n d \ln q_n = w_n \varepsilon_n \left[d \ln y - \sum_{n'=1}^{N} w_{n'} d \ln p_{n'}\right] + \sum_{n'=1}^{N} w_n \varepsilon^c_{n,n'} d \ln p_{n'}$ or, by using identity $d \ln y \equiv d \ln Q + d \ln P$, $w_n d \ln q_n = w_n \varepsilon_n \cdot d \ln Q + \sum_{n'=1}^{N} w_n \varepsilon^c_{n,n'} \cdot d \ln p_{n'}$.

Specification of the Rotterdam inverse demand systems can be derived as follows. By taking total differentiation of the compensated inverse demand $\pi_n = \pi_n(u, q_1, \cdots, q_N)$, we get $d\pi_n = \frac{\partial \pi_n}{\partial u} du + \sum_{n'=1}^{N} \frac{\partial \pi_n}{\partial q_{n'}} dq_{n'}$. By using $\frac{\partial \pi_n}{\partial u} du = \pi_n \left(\frac{\partial \ln \pi_n}{\partial \ln k}\right) \sum_{n'=1}^{N} w_{n'} d \ln q_{n'}$ derived below, we also get $d\pi_n = \pi_n \left(\frac{\partial \ln \pi_n}{\partial \ln k}\right) \sum_{n'=1}^{N} w_{n'} d \ln q_{n'} + \sum_{n'=1}^{N} \frac{\partial \pi_n}{\partial q_{n'}} dq_{n'}$. By log differential property, this equation can be written as $\pi_n d \ln \pi_n = \pi_n \left(\frac{\partial \ln \pi_n}{\partial \ln k}\right) d \ln Q + \sum_{n'=1}^{N} \frac{\partial \pi_n}{\partial q_{n'}} q_{n'} d \ln q_{n'}$, which, by multiplying q_n to both sides, can be represented by $(\pi_n q_n) d \ln \pi_n = (\pi_n q_n)\left(\frac{\partial \ln \pi_n}{\partial \ln k}\right) d \ln Q + \sum_{n'=1}^{N}\left(\frac{\pi_n}{q_{n'}} q_{n'} q_n\right)\left(\frac{\partial \pi_n}{\partial q_{n'}}\frac{q_{n'}}{\pi_n}\right) d \ln q_{n'}$ or $w_n d \ln \pi_n = w_n f_n d \ln Q + \sum_{n'=1}^{N} w_n f^c_{n,n'} d \ln q_{n'}$. The relation of $\frac{\partial \pi_n}{\partial u} du = \pi_n \left(\frac{\partial \ln \pi_n}{\partial \ln k}\right) \sum_{n'=1}^{N} w_{n'} d \ln q_{n'}$ can be derived as follows. First, by differentiating $u = U(q)$, we get $du = \sum_{n'=1}^{N} \frac{\partial u}{\partial q_{n'}} dq_{n'}$, which can be written as $u d \ln u = \sum_{n'=1}^{N} \frac{\partial u}{\partial q_{n'}} q_{n'} d \ln q_{n'}$ by log differential property and represented by

$d\ln u = \sum_{n'=1}^{N}\left(\frac{\partial u}{\partial q_{n'}}\frac{q_{n'}}{u}\right)d\ln q_{n'}$ or $d\ln u = \sum_{n'=1}^{N}\left(\frac{\partial \ln u}{\partial \ln q_{n'}}\right)d\ln q_{n'}$. Second, by differentiating $u = U(q) = U(k\cdot q^*)$ w.r.t. k, we get $du = \sum_{n=1}^{N}\frac{\partial u}{\partial kq_n^*}\frac{\partial kq_n^*}{\partial k}dk = \sum_{n=1}^{N}\frac{\partial u}{\partial q_n}q_n^* dk$, which can be written as $ud\ln u = \sum_{n=1}^{N}\frac{\partial u}{\partial q_n}q_n^* kd\ln k = \sum_{n=1}^{N}\frac{\partial u}{\partial q_n}q_n d\ln k$ by log differential property and represented by $d\ln u = \sum_{n=1}^{N}\left(\frac{\partial u}{\partial q_n}\frac{q_n}{u}\right)d\ln k = \sum_{n=1}^{N}\left(\frac{\partial \ln u}{\partial \ln q_n}\right)d\ln k$ or $\frac{d\ln u}{d\ln k} = \sum_{n=1}^{N}\left(\frac{\partial \ln u}{\partial \ln q_n}\right)$. Third, $\frac{\partial \pi_n}{\partial u}du$ can be written as $\pi_n\left(\frac{\partial \ln \pi_n}{\partial \ln k}\right)\left[\frac{d\ln u}{\partial \ln u / \partial \ln k}\right]$ by log differential property and represented by $\frac{\partial \pi_n}{\partial u}du = \pi_n\left(\frac{\partial \ln \pi_n}{\partial \ln k}\right)\sum_{n'=1}^{N}\left[\frac{\left(\partial \ln u / d\ln q_{n'}\right)}{\sum_{n=1}^{N}\partial \ln u / \partial \ln q_n}\right]d\ln q_{n'}$ using above two results. Finally, by using following relation of $w_{n'} \equiv [p_{n'}]\frac{q_{n'}}{y} = [\pi_{n'}\cdot y]\frac{q_{n'}}{y} = \left[\frac{\frac{\partial U}{\partial q_{n'}}}{\sum_{n=1}^{N}\frac{\partial U}{\partial q_n}q_n}\cdot y\right]\cdot\frac{q_{n'}}{y} = \left[\frac{\frac{\partial U}{\partial q_{n'}}\frac{q_{n'}}{u}}{\sum_{n=1}^{N}\frac{\partial U}{\partial q_n}\frac{q_n}{u}}\right] = \left[\frac{\frac{\partial \ln U}{\partial \ln q_{n'}}}{\sum_{n=1}^{N}\frac{\partial \ln U}{\partial \ln q_n}}\right]$ based on Wold theorem $\pi_{n'} = \left[\frac{\frac{\partial U}{\partial q_{n'}}}{\sum_{n=1}^{N}\frac{\partial U}{\partial q_n}q_n}\right]$, we get $\frac{\partial \pi_n}{\partial u}du = \pi_n\left(\frac{\partial \ln \pi_n}{\partial \ln k}\right)\sum_{n'=1}^{N}w_{n'}d\ln q_{n'}$.

Specification of the Rotterdam mixed demand systems can be derived as follows. By taking total differentiation of uncompensated mixed demand $q_i = q_i(y, p_1, \cdots, p_m, q_{m+1}, \cdots, q_N)$, we get $dq_i = \frac{\partial q_i}{\partial y}dy + \sum_{j=1}^{m}\frac{\partial q_i}{\partial p_j}dp_j + \sum_{s=m+1}^{N}\frac{\partial q_i}{\partial q_s}dq_s$. By log differential property, this equation can be written as $q_i d\ln q_i = \frac{\partial q_i}{\partial y}yd\ln y + \sum_{j=1}^{m}\frac{\partial q_i}{\partial p_j}p_j d\ln p_j + \sum_{s=m+1}^{N}\frac{\partial q_i}{\partial q_s}q_s d\ln q_s$, which, by multiplying $\frac{p_i}{y}$, is equal to $\left(\frac{p_i q_i}{y}\right)d\ln q_i = \left(\frac{p_i q_i}{y}\right)\left(\frac{\partial q_i}{\partial y}\frac{y}{q_i}\right)d\ln y + \left(\frac{p_i q_i}{y}\right)\sum_{j=1}^{m}\left(\frac{\partial q_i}{\partial p_j}\frac{p_j}{q_i}\right)d\ln p_j + \left(\frac{p_i q_i}{y}\right)\sum_{s=m+1}^{N}\left(\frac{\partial q_i}{\partial q_s}\frac{q_s}{q_i}\right)d\ln q_s$ or $w_i d\ln q_i = w_i\varepsilon_i d\ln y + w_i\sum_{j=1}^{m}\varepsilon_{i,j}d\ln p_j + w_i\sum_{s=m+1}^{N}q_{i,s}d\ln q_s$. By using Slutsky decomposition relations of $\varepsilon_{i,j} = \varepsilon^c_{i,j} - \varepsilon_i \left(w_j + \sum_{k=m+1}^{N} w_k \cdot p^c_{k,j} \right)$ and $q_{i,s} = q^c_{i,s} - \varepsilon_i \left(\sum_{r=m+1}^{N} w_r \cdot f^c_{r,s} \right)$ into this equation, we can get $w_i d \ln q_i = w_i \varepsilon_i d \ln y + w_i \sum_{j=1}^{m} \left[\varepsilon^c_{i,j} - \varepsilon_i \left(w_j + \sum_{k=m+1}^{N} w_k \cdot p^c_{k,j} \right) \right] d \ln p_j + w_i \sum_{s=m+1}^{N} \left[q^c_{i,s} - \varepsilon_i \left(\sum_{r=m+1}^{N} w_r \cdot f^c_{r,s} \right) \right] d \ln q_s$ or $w_i d \ln q_i = w_i \varepsilon_i \left[d \ln y - \sum_{j=1}^{m} w_j d \ln p_j \right] + \sum_{j=1}^{m} \left[w_i \varepsilon^c_{i,j} - w_i \varepsilon_i \left(\sum_{k=m+1}^{N} w_k \cdot p^c_{k,j} \right) \right] \cdot d \ln p_j + \sum_{s=m+1}^{N} \left[w_i q^c_{i,s} - w_i \varepsilon_i \left(\sum_{r=m+1}^{N} w_r f^c_{r,s} \right) \right] \cdot d \ln q_s$ or $w_i d \ln q_i = (w_i \varepsilon_i) \cdot d \ln \bar{y} + \sum_{j=1}^{m} \left[w_i \varepsilon^c_{i,j} - w_i \varepsilon_i \left(\sum_{k=m+1}^{N} w_k \cdot p^c_{k,j} \right) \right] \cdot d \ln p_j + \sum_{s=m+1}^{N} \left[w_i q^c_{i,s} - w_i \varepsilon_i \left(\sum_{r=m+1}^{N} w_r f^c_{r,s} \right) \right] \cdot d \ln q_s$. By taking total differentiation of uncompensated mixed demand $p_k = p_k(y, p_1, \cdots, p_m, q_{m+1}, \cdots, q_N)$, we get $dp_k = \frac{\partial p_k}{\partial y} dy + \sum_{j=1}^{m} \frac{\partial p_k}{\partial p_j} dp_j + \sum_{s=m+1}^{N} \frac{\partial p_k}{\partial q_s} dq_s$. By log differential property, this equation becomes

$p_k d \ln p_k = \frac{\partial p_k}{\partial y} y d \ln y + \sum_{j=1}^{m} \frac{\partial p_k}{\partial p_j} p_j d \ln p_j + \sum_{s=m+1}^{N} \frac{\partial p_k}{\partial q_s} q_s d \ln q_s$, which, by multiplying $\frac{p_i}{y}$, is equal to

$$\left(\frac{p_k q_k}{y} \right) d \ln p_i = \left(\frac{p_k q_k}{y} \right) \left(\frac{\partial p_k}{\partial y} \frac{y}{p_k} \right) d \ln y + \left(\frac{p_k q_k}{y} \right) \sum_{j=1}^{m} \left(\frac{\partial p_k}{\partial p_j} \frac{p_j}{p_k} \right) d \ln p_j + \left(\frac{p_k q_k}{y} \right) \sum_{s=m+1}^{N} \left(\frac{\partial p_k}{\partial q_s} \frac{q_s}{p_k} \right) d \ln q_s$$

or $w_k d \ln p_k = w_k f_k d \ln y + w_k \sum_{j=1}^{m} p_{k,j} d \ln p_j + w_k \sum_{s=m+1}^{N} f_{k,s} d \ln q_s$. By using Slutsky decomposition relations of $p_{k,j} = p^c_{k,j} - f_k \left(w_j + \sum_{r=m+1}^{N} w_r \cdot p^c_{r,j} \right)$ and $f_{k,s} = f^c_{k,s} - f_k \left(\sum_{r=m+1}^{N} w_r \cdot f^c_{r,s} \right)$ into this equation, we get $w_k d \ln p_k = w_k f_k d \ln y + w_k \sum_{j=1}^{m} \left[p^c_{k,j} - f_k \left(w_j + \sum_{r=m+1}^{N} w_r \cdot p^c_{r,j} \right) \right] \cdot d \ln p_j + w_k \sum_{s=m+1}^{N} \left[f^c_{k,s} - f_k \left(\sum_{r=m+1}^{N} w_r \cdot f^c_{r,s} \right) \right] \cdot d \ln q_s$ or $w_k d \ln p_k = w_k f_k \left[d \ln y - \sum_{j=1}^{m} w_j d \ln p_j \right] + \sum_{j=1}^{m} \left[w_k p^c_{k,j} - w_k f_k \cdot \left(\sum_{r=m+1}^{N} w_r p^c_{r,j} \right) \right] \cdot d \ln p_j + \sum_{s=m+1}^{N} \left[w_k f^c_{k,s} - w_k f_k \cdot \left(\sum_{r=m+1}^{N} w_r f^c_{r,s} \right) \right] \cdot d \ln q_s$ or $w_k d \ln p_k = (w_k f_k) \cdot d \ln \bar{y} + \sum_{j=1}^{m} \left[w_k p^c_{k,j} - w_k f_k \cdot \left(\sum_{r=m+1}^{N} w_r p^c_{r,j} \right) \right] \cdot d \ln p_j + \sum_{s=m+1}^{N} \left[w_k f^c_{k,s} - w_k f_k \cdot \left(\sum_{r=m+1}^{N} w_r f^c_{r,s} \right) \right] \cdot d \ln q_s$.

Differential LA/AIDS Functional Form

Originally the LA/AIDS direct demand systems can be derived by using following specification of cost function: $y \equiv C(u, P) = \exp[a(P) + ub(P)]$ or $\ln y \equiv \ln C(u, P) = a(P) + ub(P)$ where $a(p) = \alpha_0 + \sum_{n=1}^{N} \alpha_n \ln p_n + \frac{1}{2} \sum_{n=1}^{N} \sum_{n'=1}^{N} \gamma_{n,n'} \ln p_n \ln p_{n'}$ and $b(p) = \beta_0 \prod_{n=1}^{N} p_n^{\beta_n}$. By taking differentiation, we get $\frac{\partial \ln C}{\partial \ln p_n} = \alpha_n + \sum_{n'=1}^{N} \gamma_{n,n'} \ln p_{n'} + \beta_n \cdot ub(P)$, which, by using both log differential property and Hotelling-Shephard lemma $\partial C / \partial p_n = q_n$, can be written as $\left(\frac{\partial C}{\partial p_n}\frac{p_n}{C}\right)=\left(\frac{q_n p_n}{y}\right)=\alpha_n+\sum_{n'=1}^{N}\gamma_{n,n'}\ln p_{n'}+\beta_n\ln Q$, where $ub(P)=\ln y-a(P)\approx\ln y-\ln P\equiv\ln Q$ is derived by using a linear approximation of $a(p)\approx\sum_{n'=1}^{N}w_{n'}\ln p_{n'}\equiv\ln P$. This level version of LA/AIDS $w_n=\alpha_n+\beta_n\ln Q+\sum_{n'=1}^{N}\gamma_{n,n'}\ln p_{n'}$ can be written differential version of LA/AIDS $dw_n=\beta_n d\ln Q+\sum_{n'=1}^{N}\gamma_{n,n'}d\ln p_{n'}$.

Based on the similar logical procedure, the LA/AIDS inverse demand systems can be derived by using following specification of distance function: $1\equiv D(u,q)=\exp[a(q)+ub(q)]$ or $\ln 1=0\equiv\ln D(u,q)=a(q)+ub(q)$ where $a(q)=\alpha_0+\sum_{n=1}^{N}\alpha_n\ln q_n+\frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\gamma_{n,n'}\ln q_n\ln q_{n'}$ and $b(q)=\beta_0\prod_{n=1}^{N}q_n^{\beta_n}$. By taking differentiation, we get $\frac{\partial\ln D}{\partial\ln q_n}=\alpha_n+\sum_{n'=1}^{N}\gamma_{n,n'}\ln q_{n'}+\beta_n\cdot ub(q)$, which, by log differential property and Shephard-Hanoch lemma $\partial D/\partial q_n=\pi_n$, can be written as $\left(\frac{\partial D}{\partial q_n}\frac{q_n}{D}\right)=\left(\frac{\pi_n q_n}{1}\right)=\alpha_n+\sum_{n'=1}^{N}\gamma_{n,n'}\ln q_{n'}+\beta_n\ln Q$, where $ub(q)=a(q)\approx\ln Q$ is derived by using a linear approximation of $a(q)\approx\sum_{n'=1}^{N}w_{n'}\ln q_{n'}\equiv\ln Q$. This level version of LA/AIDS $w_n=\alpha_n+\beta_n\ln Q+\sum_{n'=1}^{N}\gamma_{n,n'}\ln q_{n'}$ can be written differential version of LA/AIDS $dw_n=\beta_n d\ln Q+\sum_{n'=1}^{N}\gamma_{n,n'}d\ln q_{n'}$.

CBS Functional Form

Originally the CBS direct demand systems can be derived by subtracting $w_n d\ln Q$ from both side of Rotterdam models to introduce variational expenditure elasticity into Rotterdam specification as follows: $w_n d\ln q_n-w_n d\ln Q=w_n\varepsilon_n d\ln Q+\sum_{n'=1}^{N}w_n\varepsilon_{n,n'}^{c}d\ln p_{n'}-w_n d\ln Q$, which can be represented by $w_n d\ln\left(\frac{q_n}{Q}\right)=[w_n\varepsilon_n-w_n]d\ln Q+\sum_{n'=1}^{N}w_n\varepsilon_{n,n'}^{c}d\ln p_{n'}$.

Similarly the CBS inverse demand systems can be derived by adding $w_n d\ln Q$ to both side of Rotterdam models to introduce variational scale flexibility as follows: $w_n d\ln\pi_n+w_n d\ln Q=w_n f_n d\ln Q+\sum_{n'=1}^{N}w_n f_{n,n'}^{c}d\ln q_{n'}+w_n d\ln Q$ which can be represented by

$\left\{ w_n d \ln\left(\frac{p_n}{P}\right) \right\} = [w_n f_n + w_n] d \ln Q + \sum_{n'=1}^{N} [w_n f^c_{n,n'}] d \ln q_{n'}$, by using the relation of $d \ln \pi_n + d \ln Q$ $= [d \ln p_n - d \ln Y] + d \ln Q \;= d \ln p_n - [d \ln Y - d \ln Q] = d \ln p_n - d \ln P = d \ln\left(\frac{p_n}{P}\right)$.

NBR Functional Form

Originally the NBR direct demand systems can be derived by adding $w_n d \ln Q$ to both side of LA/AIDS models to introduce constant expenditure elasticity into LA/AIDS specification as follows: $dw_n + w_n d \ln Q = [w_n \varepsilon_n - w_n] d \ln Q + \sum_{n'=1}^{N} [w_n \varepsilon^c_{n,n'} - w_n (w_{n'} - \delta_{n,n'})] d \ln p_{n'} + w_n d \ln Q$ or $\{ dw_n + w_n d \ln Q \} = [w_n \varepsilon_n] d \ln Q + \sum_{n'=1}^{N} [w_n \varepsilon^c_{n,n'} - w_n (w_{n'} - \delta_{n,n'})] d \ln p_{n'}$.

Similarly the NBR inverse demand systems can be derived by subtracting $w_n d \ln Q$ from both side of LA/AIDS models to introduce constant scale flexibility into LA/AIDS specification as follows: $dw_n - w_n d \ln Q = [w_n f_n + w_n] d \ln Q + \sum_{n'=1}^{N} [w_n f^c_{n,n'} - w_n (w_{n'} - \delta_{n,n'})] d \ln q_{n'} - w_n d \ln Q$ or $\{ dw_n - w_n d \ln Q \} = [w_n f_n] d \ln Q + \sum_{n'=1}^{N} [w_n f^c_{n,n'} - w_n (w_{n'} - \delta_{n,n'})] d \ln q_{n'}$.

Relation among Four Functional Forms

Since mathematical equivalences between Rotterdam and CBS and between LA/AIDS and NBR are obvious, it is enough to show relationships between Rotterdam and differential version of LA/AIDS to connect all four differential family functional forms.

In direct demand functions, using $dw_n = \langle w_n d \ln q_n \rangle + w_n d \ln p_n - w_n [d \ln Q + d \ln P]$, Rotterdam can be written as differential version of LA/AIDS $dw_n = \beta_n d \ln Q + \sum_{n'=1}^{N} \gamma_{n,n'} d \ln p_{n'}$ through $dw_n = \left\langle w_n \varepsilon_n d \ln Q + \sum_{n'=1}^{N} w_n \varepsilon^c_{n,n'} d \ln p_{n'} \right\rangle + w_n \left[\sum_{n'=1}^{N} \delta_{n,n'} d \ln p_{n'} \right] - w_n d \ln Q - w_n \left[\sum_{n'=1}^{N} w_{n'} d \ln p_{n'} \right]$ by using parameterization of $\beta_n = [w_n \varepsilon_n - w_n]$ and $\gamma_{n,n'} = [w_n \varepsilon^c_{n,n'} - w_n (w_{n'} - \delta_{n,n'})]$ and by using $w_n d \ln q_n = \langle dw_n \rangle - w_n d \ln p_n + w_n [d \ln Q + d \ln P]$, LA/AIDS can be written as Rotterdam via $w_n d \ln q_n = \left\langle [w_n \varepsilon_n - w_n] d \ln Q + \sum_{n'=1}^{N} [w_n \varepsilon^c_{n,n'} - w_n (w_{n'} - \delta_{n,n'})] d \ln p_{n'} \right\rangle - w_n \left[\sum_{n'=1}^{N} \delta_{n,n'} d \ln p_{n'} \right] + w_n d \ln Q + w_n \sum_{n'=1}^{N} w_{n'} d \ln p_{n'}$.

In inverse demand functions, using $dw_n = \langle w_n d \ln \pi_n \rangle + w_n d \ln q_n + w_n d \ln Q - w_n d \ln Q$, Rotterdam can be written as differential version of LA/AIDS $dw_n = \beta_n d \ln Q + \sum_{n'=1}^{N} \gamma_{n,n'} d \ln q_{n'}$ through $dw_n = \left\langle \left[w_n f_n \right] d \ln Q + \sum_{n'=1}^{N} w_n f_{n,n'}^c d \ln q_{n'} \right\rangle + w_n \left[\sum_{n'=1}^{N} \delta_{n,n'} d \ln q_{n'} \right] + w_n d \ln Q - w_n \left[\sum_{n'=1}^{N} w_{n'} d \ln q_{n'} \right]$ by putting $\beta_n = \left[w_n f_n + w_n \right]$ and $\gamma_{n,n'} = \left[w_n f_{n,n'}^c - w_n \left(w_{n'} - \delta_{n,n'} \right) \right]$ and using $w_n d \ln \pi_n = \langle dw_n \rangle - w_n d \ln q_n - w_n d \ln Q + w_n d \ln Q$, LA/AIDS can be written as Rotterdam via $w_n d \ln \pi_n = \left\langle \left[w_n f_n + w_n \right] d \ln Q + \sum_{n'=1}^{N} \left[w_n f_{n,n'}^c - w_n \left(w_{n'} - \delta_{n,n'} \right) \right] d \ln q_{n'} \right\rangle - w_n \left[\sum_{n'=1}^{N} \delta_{n,n'} d \ln q_{n'} \right] - w_n d \ln Q + w_n \sum_{n'=1}^{N} w_{n'} d \ln q_{n'}$.

APPENDIX D
DATA DESCRIPTION*

Var. #	Description of Variables	Brand Category	UPC Code
001	SUNKIST STRAWBERRY	SUNKIST	4640010041
002	SUNKIST ORANGE	SUNKIST	4640014021
003	CANADA DRY GINGER ALE	CANADA DRY	1690000013
004	CANADA DRY GINGER ALE	CANADA DRY	1690000083
005	SPRITE	SPRITE	4900000132
006	COCA-COLA CLASSIC	COKE	4900000634
007	COKE DIET	COKE	4900000658
008	COKE DIET CAFFEINE FREE	COKE	4900000929
009	PEPSI-COLA	PEPSI	1200000013
010	PEPSI-DIET	PEPSI	1200000050
011	DIET PEPSI CAFFEINE FREE	PEPSI	1200000494
012	CAFFEINE FREE PEPSI	PEPSI	1200000490
013	MOUNTAIN DEW	MOUNTAIN DEW	1200000085
014	SEVEN-UP	SEVEN-UP	7800000038
015	SEVEN-UP DIET	SEVEN-UP	7800000079
016	DR PEPPER SUGAR FREE	DR PEPPER	5490000030
017	DR PEPPER	DR PEPPER	5490000029
018	A & W DIET ROOT BEER	A & W	7020200006
019	A & W ROOT BEER	A & W	7020200005
020	DIET RITE COLA	RITE COLA	2950005254
021	DIET RITE RED RASPBERRY	RITE COLA	2950085254
022	LIPTON BRISK ICED TEA	LIPTON	4100000814
023	LIPTON DIET BRISK TEA	LIPTON	4100010728

* All the products are size of 6/12 oz.
* The classification and ordering of variables are based on the result of empirical analysis

APPENDIX E

DATA DESCRIPTION*

Var. #	Variable Name	Descriptions	T_Code	Slow
Exchange Rate Variable Group (ExRate)				
001	EX rate: Canada	FOREIGN EXCHANGE RATE: CANADA (CANADIAN $ PER U.S.$)	5	0
002	Ex rate: UK	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)	5	0
003	Ex rate: Switz	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U.S.$)	5	0
004	Ex rate: avg	UNITED STATES;EFFECTIVE EXCHANGE RATE(MERM)(INDEX NO.)	5	0
005	Ex rate: Japan	FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.$)	5	0
Stock Market Variable Group (Stock)				
006	Consumer expect	U. OF MICH. INDEX OF CONSUMER EXPECTATIONS(BCD-83)	2	0
007	S&P PE ratio	S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO (%,NSA)	5	0
008	S&P: indust	S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS (1941-43=10)	5	0
009	S&P 500	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)	5	0
Money Aggregate Variable Group (Money)				
010	M2 (real)	MONEY SUPPLY - M2 IN 1996 DOLLARS (BCI)	5	0
011	M2	MONEY STOCK:M2(M1+O'NITE RPS,EURO$,G/P&B/D MMMFS&SAV&SM TIME DEP(BIL$,	5	0
012	M3	MONEY STOCK: M3(M2+LG TIME DEP,TERM RP'S&INST ONLY MMMFS)(BIL$,SA)	5	0
013	M1	MONEY STOCK: M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL$,SA)	5	0
014	MB	MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES(MIL$,SA)	5	0
015	Reserves tot	DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS(MIL$,SA)	5	0
016	Reserves nonbor	DEPOSITORY INST RESERVES:NONBORROWED,ADJ RES REQ CHGS(MIL$,SA)	5	0
Price Variable Group (Price)				
017	CPI-U: ex shelter	CPI-U: ALL ITEMS LESS SHELTER (82-84=100,SA)	6	1
018	CPI-U: comm.	CPI-U: COMMODITIES (82-84=100,SA)	6	1
019	CPI-U: ex med	CPI-U: ALL ITEMS LESS MEDICAL CARE (82-84=100,SA)	6	1
020	CPI-U: all	CPI-U: ALL ITEMS (82-84=100,SA)	6	1
021	CPI-U: transp	CPI-U: TRANSPORTATION (82-84=100,SA)	6	1
022	CPI-U: ex food	CPI-U: ALL ITEMS LESS FOOD (82-84=100,SA)	6	1

* In the transformation code (T-code), the following numbers are used for each transformation: 1: no transformation. 2: first difference, 4: logarithm, 5:first difference of logarithm, and 6: second difference of logarithm.
* In the block recursive assumption (Slow), the number of 1 denotes the assumed slow-moving variables.

Var. #	Variable Name	Descriptions	T_Code	Slow
023	PPI: int mat'ls	PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES & COMPONENTS(82=100,SA)	6	0
024	PPI: cons gds	PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS (82=100,SA)	6	0
025	PPI: fin gds	PRODUCER PRICE INDEX: FINISHED GOODS (82=100,SA)	6	0
026	PPI: crude mat'ls	PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)	6	0
027	Commod: spot price	SPOT MARKET PRICE INDEX:BLS & CRB: ALL COMMODITIES(1967=100)	6	0
028	Sens mat'ls price	INDEX OF SENSITIVE MATERIALS PRICES (1990=100)(BCI-99A)	6	0

Interest Rate Variable Group (Interest)

029	Baa bond	BOND YIELD: MOODY'S BAA CORPORATE (% PER ANNUM)	2	0
030	Aaabond	BOND YIELD: MOODY'S AAA CORPORATE (% PER ANNUM)	2	0
031	10 yr T-bond	INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(% PER ANN,NSA)	2	0
032	5 yr T-bond	INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.(% PER ANN,NSA)	2	0
033	1 yr T-bond	INTEREST RATE: U.S.TREASURY CONST MATURITIES,1-YR.(% PER ANN,NSA)	2	0
034	6 mo T-bill	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,6-MO.(% PER ANN,NSA)	2	0
035	3 mo T-bill	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,3-MO.(% PER ANN,NSA)	2	0
036	Commpaper	Cmmercial Paper Rate (AC)	2	0

Spread Variable Group (Spread)

037	CP-FF spread	cp90-fyff	1	0
038	3 mo-FF spread	fygm3-fyff	1	0
039	6 mo-FF spread	fygm6-fyff	1	0
040	1 yr-FF spread	fygt1-fyff	1	0
041	5 yr-FFspread	fygt5-fyff	1	0
042	10yr-FF spread	fygt10-fyff	1	0
043	Aaa-FF spread	fyaaac-fyff	1	0
044	Baa-FF spread	fybaac-fyff	1	0

Var. #	Variable Name	Descriptions	T_Code	Slow

Housing Market Variable Group (House)

045	HStarts: NE	HOUSING STARTS:NORTHEAST (THOUS.U.)S.A.	4	0
046	BP: NE	HOUSES AUTHORIZED BY BUILD. PERMITS:NORTHEAST(THOU.U.)S.A	4	0
047	HStarts: MW	HOUSING STARTS:MIDWEST(THOUS.U.)S.A.	4	0
048	BP: MW	HOUSES AUTHORIZED BY BUILD. PERMITS:MIDWEST(THOU.U.)S.A.	4	0
049	BP: West	HOUSES AUTHORIZED BY BUILD. PERMITS:WEST(THOU.U.)S.A.	4	0
050	HStarts: West	HOUSING STARTS:WEST (THOUS.U.)S.A.	4	0
051	HStarts: Total	HOUSING STARTS:NONFARM(1947-58);TOTAL FARM&NONFARM(1959-)(THOUS.,SA	4	0
052	BP: total	HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING UNITS (THOUS.,SAAR)	4	0
053	HStarts: South	HOUSING STARTS:SOUTH (THOUS.U.)S.A.	4	0
054	BP: South	HOUSES AUTHORIZED BY BUILD. PERMITS:SOUTH(THOU.U.)S.A.	4	0

NAPM Variable Group (NAPM)

055	NAPM com price	NAPM COMMODITY PRICES INDEX (PERCENT)	1	0
056	NAPM Invent	NAPM INVENTORIES INDEX (PERCENT)	1	0
057	NAPM vendor del	NAPM VENDOR DELIVERIES INDEX (PERCENT)	1	0
058	NAPM empl	NAPM EMPLOYMENT INDEX (PERCENT)	1	1
059	PMI	PURCHASING MANAGERS' INDEX (SA)	1	0
060	NAPM prodn	NAPM PRODUCTION INDEX (PERCENT)	1	1
061	NAPM new ordrs	NAPM NEW ORDERS INDEX (PERCENT)	1	0

Employment Variable Group (Emp)

062	Emp CPS total	CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)	5	1
063	Emp CPS nonag	CIVILIAN LABOR FORCE: EMPLOYED, NONAGRIC.INDUSTRIES (THOUS.,SA)	5	1
064	Emp-hrs nonag	Employee hours in nonag. establishments (AR, bil. hours)	5	1
065	Emp: const	EMPLOYEES ON NONFARM PAYROLLS - CONSTRUCTION	5	1
066	Emp: retail	EMPLOYEES ON NONFARM PAYROLLS - RETAIL TRADE	5	1
067	Emp: TTU	EMPLOYEES ON NONFARM PAYROLLS - TRADE, TRANSPORTATION, AND UTILITIES	5	1
068	Emp: services	EMPLOYEES ON NONFARM PAYROLLS - SERVICE-PROVIDING	5	1
069	Emp: total	EMPLOYEES ON NONFARM PAYROLLS - TOTAL PRIVATE	5	1
070	Emp: gds prod	EMPLOYEES ON NONFARM PAYROLLS - GOODS-PRODUCING	5	1
071	Emp: mfg	EMPLOYEES ON NONFARM PAYROLLS - MANUFACTURING	5	1
072	Emp: dble gds	EMPLOYEES ON NONFARM PAYROLLS - DURABLE GOODS	5	1
073	Emp: nondbles	EMPLOYEES ON NONFARM PAYROLLS - NONDURABLE GOODS	5	1
074	Emp: wholesale	EMPLOYEES ON NONFARM PAYROLLS - WHOLESALE TRADE	5	1
075	Emp: FIRE	EMPLOYEES ON NONFARM PAYROLLS - FINANCIAL ACTIVITIES	5	1

Var. #	Variable Name	Descriptions	T_Code	Slow

Output Variable Group (Output)

076	IP:nondble mats	INDUSTRIAL PRODUCTION INDEX - NONDURABLE GOODS MATERIALS	5	1
077	IP:bus eqpt	INDUSTRIAL PRODUCTION INDEX - BUSINESS EQUIPMENT	5	1
078	IP: dble mats	INDUSTRIAL PRODUCTION INDEX - DURABLE GOODS MATERIALS	5	1
079	IP: matls	INDUSTRIAL PRODUCTION INDEX - MATERIALS	5	1
080	IP: total	INDUSTRIAL PRODUCTION INDEX - TOTAL INDEX	5	1
081	IP: mfg	INDUSTRIAL PRODUCTION INDEX - MANUFACTURING (SIC)	5	1
082	Cap util	Capacity Utilization (Mfg)	2	1
083	IP: products	INDUSTRIAL PRODUCTION INDEX - PRODUCTS, TOTAL	5	1
084	IP: final prod	INDUSTRIAL PRODUCTION INDEX - FINAL PRODUCTS	5	1
085	IP: cons gds	INDUSTRIAL PRODUCTION INDEX - CONSUMER GOODS	5	1
086	IP: cons dble	INDUSTRIAL PRODUCTION INDEX - DURABLE CONSUMER GOODS	5	1
087	IP:cons nondble	INDUSTRIAL PRODUCTION INDEX - NONDURABLE CONSUMER GOODS	5	1
088	PI	Personal income (AR, bil. chain 2000 $)	5	1
089	PI less transfers	Personal income less transfer payments (AR, bil. chain 2000 $)	5	1

Consumption/Investment Variable Group (Cons/Inv)

090	Orders: cap gds	Mfrs' new orders, nondefense capital goods (mil. chain 1982 $)	5	0
091	Orders: dble gds	Mfrs' new orders, durable goods industries (bil. chain 2000 $)	5	0
092	Orders: cons gds	Mfrs' new orders, consumer goods and materials (bil. chain 1982 $)	5	0
093	M&T sales	Manufacturing and trade sales (mil. Chain 1996 $)	5	1
094	M&T invent/sales	Ratio, mfg. and trade inventories to sales (based on chain 2000 $)	2	0
095	Retail sales	Sales of retail stores (mil. Chain 2000 $)	5	1
096	Consumption	Real Consumption (AC) A0m224/gmdc	5	1

Unemployment Variable Group (Unemp)

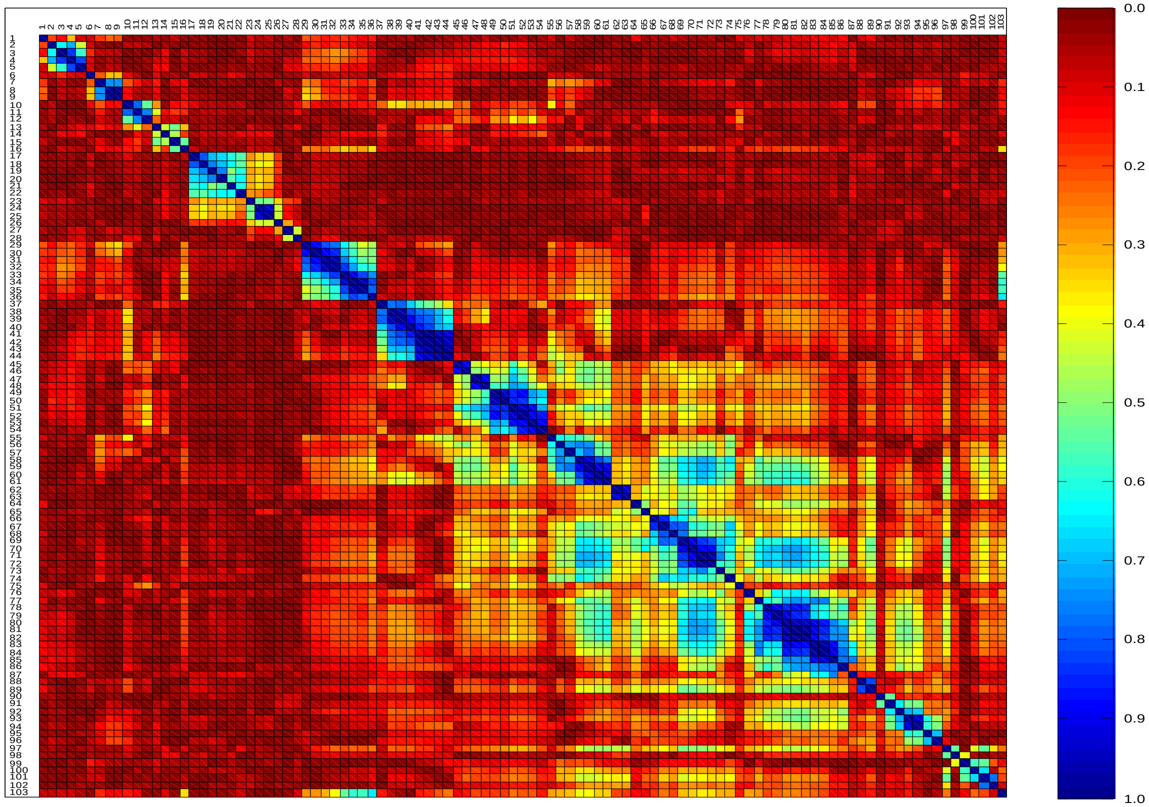
097	U: all	UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER (%,SA)	2	1
098	U < 5 wks	UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5 WKS (THOUS.,SA)	5	1
099	U: mean duration	UNEMPLOY.BY DURATION: AVERAGE(MEAN)DURATION IN WEEKS (SA)	2	1
100	U 27+ wks	UNEMPLOY.BY DURATION: PERSONS UNEMPL.27 WKS + (THOUS,SA)	5	1
101	U 15+ wks	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS + (THOUS.,SA)	5	1
102	U 15-26 wks	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS (THOUS.,SA)	5	1

Federal Funds Rate Variable (FFR)

103	FedFunds	INTEREST RATE: FEDERAL FUNDS (EFFECTIVE) (% PER ANNUM,NSA)	2	0

APPENDIX F

STANDARD STATIC CORRELATION MATRIX



* See Appendix E for the description of variables, where variables are in the same order.

VITA

Name: DAE-HEUM KWON

Address: Department of Agricultural Economics
Texas A&M University 2124 TAMU
College Station, Texas 77843-2124

E-mail Address: heum_us@tamu.edu

Education: Ph.D. Agricultural Economics
Texas A&M University, 2007

M.S. Agricultural Economics
Korea University, 1998

B.S. Agricultural Economics
Korea University, 1996