Essays on necessary and sufficient conditions for global and local identification in linear and nonlinear models

Xin Liang

Department of Economics

McGill University
Montreal, Quebec

February 2015

A thesis submitted to McGill University in partial fulfillment of
the requirements of the degree of Doctor of Philosophy

©Xin Liang 2015

DEDICATION

This thesis is dedicated to my wife Nan and my son Daniel.

ACKNOWLEDGEMENT

As a Ph.D. student, I have chosen the most ideal location in Canada and one of the best in the world – McGill University – to study my interested field in theoretical and applied econometrics due to its strong and vigorous research environment for econometrics.

Foremost, I want to express my sincere gratitude to my primary advisor Professor Jean-Marie Dufour, one of the world-leading econometricians, for his excellent guidance on my Ph.D. study. I am always amazed by his tremendous and thorough knowledge that covers almost every major field in econometrics and inspired by his perdurable enthusiasm and energy for conducting rigorous research. I am also deeply affected by his scrupulous working spirit that will not allow for the the least bit negligence, which has lays a solid foundation for my future academic career. It is my great honour to study under his direction and have him as my mentor. It would not have been possible to complete this doctoral thesis without his insightful comments, great patience, continuous encouragement and full support.

I would also like to thank my secondary supervisor Professor Victoria Zinde-Walsh, who not only teaches me knowledge of econometrics but also provides invaluable suggestions about everyday life and family which benefit me a lot. To me, Vicky is both an outstanding advisor and close friend from whom I can always seek advice and comfort. I am very lucky to know her.

My special thanks go to Professor Russell Davidson who is willing to be part of my thesis defense committee and Professor John W. Galbraith who is always willing to help and give his best suggestions.

I am also indebted to a group of wonderful faculty members in the Department of Economics including but not limited to Professor Jagdish Handa, Professor Francisco Alvarez-Cuadrado, Professor Markus Poschke, Professor Ngo Van Long, Professor Fabian Lange, Professor Hassan Benchekroun and Professor Rohan Dutta, who either teach me graduate courses or provide academic comments on my Ph.D. research. My thanks are also extended to the faculty members at the Department of Mathematics and Statistics, inclusive of Professor Christian Genest, Professor David Stephens, Professor Dmitry Jakobson, Professor Bruce Shepherd and Professor William Anderson, for their permission to let me sit in their classes and patience to answer my questions. I would like to take this opportunity to record my sincere thanks to Elaine Garnham, Judy Dear, Jackie Gregory and Lisa Stevenson for their professionalism and extraordinary administrative assistance.

I would like to acknowledge the financial supports from McGill Fellowship, SSHRC and funding sources of Professor Jean-Marie Dufour, which allow me to concentrate on my Ph.D. research and finally accomplish the thesis. I thank McGill University for offering

me academic and technical support and I am grateful to the University of Montreal and Concordia University for providing their library facilities.

Last, but by no means least, I am thankful for understanding and support of my dear family members, especially my parents who always stand by me through all these years.

CONTRIBUTION OF AUTHORS

The second and fourth chapters of this thesis are co-authored with Professor Jean-Marie Dufour. I'm responsible for the completion of the two papers. I selected the research topics, composed literature reviews and conducted both theoretical and empirical analysis. Professor Jean-Marie Dufour guided me through some of the major theoretical derivations and provided advice on empirical applications. The third chapter is my independent research paper.

ABSTRACT

The first essay considers a class of generalized linear models (deemed "partially linear models") where parameters of interest determine the distribution of the data through multiplication by a known matrix, such as a matrix of explanatory variables whose rank may be deficient. This setup not only covers linear regression models with collinearity (such as cases where the number of explanatory variables is potentially very large or the number observations is inferior to the number of variables) and a general error covariance matrix, but a wide spectrum of other models used in econometrics, such as linear median regressions and quantile regressions, generalized linear mixed models, probit and Tobit models, multinomial logit models and other discrete choice models, exponential models, index models, etc. In this essay, we first provide a general necessary and sufficient condition for the global identification of a general transformation of model parameters (when the full parameter vector is not typically identified) based on a new separability condition. This general global identification condition is no way limited to partially linear models, but its discovery was motivated by such models. The general result is then applied to partially linear models, and the class of identifiable parameters is fully characterized for such models. Even though none of the original individual parameters of the model may be identified, it is possible to describe the class of linear transforms which can be identified. In view of deriving illuminating and usable conditions, several different (though equivalent) characterizations are derived through abstract linear algebra and generalized inverse methods. The effect of adding restrictions is also considered, and the corresponding identification conditions are supplied. Despite their fundamental nature, these results are new and should be useful in many areas of statistics and econometrics.

The second essay reconsiders the problem of characterizing identifiable parameters in linear IV regressions and simultaneous equations models (SEMs), using methods similar to the ones developed in the first essay. The recent econometric literature on weak instruments mainly deals with this basic setup, and the distributional theory as well as the appropriate statistical methods depend crucially on whether the parameters of interest are identifiable. Surprisingly, a general characterization of such parameters is not available. Further, the case where the instrument matrix does not have full rank [e.g., when the number of instruments exceeds the number of observations] has not been apparently considered. In this work, we extend the approaches and results given in the first essay to this general setup. More precisely, we study the general case where some model parameters are not identifiable, without any restriction on the rank of the instrument matrix, and we characterize which linear transformations of the structural parameters are identifiable. An important observation is that identifiable parameters may depend on the instrument matrix (in addition

to the parameters of the reduced form), and a number of alternative characterizations are provided. These results are also applicable to partially linear IV-type models where the linear IV structure is embedded in a nonlinear structure, such as a quantile specification or a discrete choice model (as in the first essay).

The third essay takes up the challenging problem of characterizing the identification of nonlinear functions of parameters in nonlinear models. The setup is fundamentally semi-parametric, and the basic assumption is that structural parameters of interest determine a number of identifiable parameters through a nonlinear equation (such as a conditional or unconditional moment equation). Such models are quite common in econometrics, and include for example nonlinear models typically estimated by GMM, and dynamic stochastic general equilibrium (DSGE) models (used in macroeconomics and finance). Again, we consider the general case where not all model parameters are identifiable, with the purpose of characterizing nonlinear parameter transformations which are identifiable. The literature on this problem is very thin, and only deals with the identification of the full parameter vector in the equation of interest. In view of the fact global identification is extremely difficult to achieve, this paper looks at the problem from a local identification viewpoint. Both sufficient conditions, as well as necessary and sufficient conditions are derived under assumptions of differentiability of the relevant moment equations and parameter transformations. On the way, a notion of identification intermediate between global identification and the standard notion of local identification (at a point), namely identification around a point, is introduced and plays an important role in the proposed theory. Some classical results on identification in likelihood models are also derived and extended. Finally, the results are applied to identification problems in DSGE models.

RÉSUMÉ

Cette thèse comprend trois d'essais sur l'identification des modèles économétriques. La littérature sur les problèmes d'identification s'est beaucoup développée au cours des dernières années, particulièrement dans le but de proposer des méthodes d'inférence statistique fiables lorsque certains paramètres ne sont pas identifiables (ou faiblement identifiables). Ici, nous nous concentrons sur l'établissement de conditions nécessaires et suffisantes pour l'identification de modèles linéaires et non-linéaires, dans le but de développer des méthodes d'inférence statistique fiables sur les paramètres identifiables.

Le premier essai considère une classe générale de modèles linéaires généralisés (appelés "modèles partiellement linéaires") où les paramètres d'intérêt déterminent la distribution des données via une multiplication par une matrice connue, telle une matrice de variables explicatives dont le rang peut être déficient. Ce cadre comprend non seulement les modèles de régression linéaire avec collinéarité (ce qui se produit, par exemple, lorsque le nombre de variables explicatives est supérieur au nombre d'observations) et une matrice de covariance générale, mais aussi une gamme d'autres modèles économétriques, tels la régression en médiane et la régression-quantile, les modèles linéaires généralisés mixtes, les modèles probit et Tobit, les modèles logit multilinéaires ainsi que divers modèles de choix discrets, etc. Dans cet essai, nous donnons d'abord une condition nécessaire et suffisante générale pour l'identification globale d'une transformation générale des paramètres du modèle (lorsque le vecteur complet des paramètres du modèle n'est pas identifiable) sous une hypothèse de séparabilité. Cette condition globale n'est en rien limitée aux modèles partiellement linéaires, mais sa formulation est motivée par ces modèles. Nous appliquons ensuite ce résultat aux modèles partiellement linéaires, et nous caractérisons complètement la classe des paramètres linéaires identifiables. Même si aucun des paramètres originaux du modèle n'est identifiable, nous montrons qu'il est possible de décrire la classe des transformations linéaires de ces paramètres qui sont identifiables. Afin d'obtenir des conditions faciles à interpréter et à utiliser, nous dérivons plusieurs formulations différentes en exploitant divers résultats d'algèbre linéaires notamment sur les inverses généralisées. Nous traitons aussi le cas où des restrictions sont ajoutées au modèle, et nous donnons les conditions d'identification applicables dans ce cas. En dépit de leur caractère fondamental, ces résultats sont nouveaux et devraient être utiles dans plusieurs domaines de la statistique et de l'économétrie.

Le second essai reconsidère le problème qui consiste à caractériser les paramètres identifiables dans les modèles de régression IV et les modèles à équations simultanées linéaires, en utilisant des méthodes semblables à celles mises en oeuvre dans le premier essai. La littérature économétrique récente sur les instruments faibles considère ce cadre, et la théorie

distributionnelle ainsi les méthodes statistiques appropriées dépendent des paramètres qui sont identifiables. Étonnamment, une théorie générale ne semble pas disponible. En particulier, le cas où la matrice des instruments n'est pas de plein rang [e.g., lorsque le nombre d'instruments excède le nombre d'observations] n'a pas été étudié. Dans cet essai, nous étendons les approches et les résultats du premier essai à ce cadre général. Plus précisément, nous étudions des modèles où certains paramètres ne sont pas identifiables, sans restriction sur le rang de la matrice des instruments, et nous établissons quelles transformations linéaires des paramètres du modèle sont identifiables. Nous soulignons que l'identifiabilité des paramètres dépend de la matrice des instruments (en sus des paramètres de la forme réduite), et nous donnons plusieurs caractérisations. Ces résultats sont aussi applicables à des modèles partiellement linéaires de type IV, où la structure linéaire IV est plongée dans une structure non-linéaire, tels qu'une spécification basées sur des quantiles ou un modèle à choix discrets.

Le troisième essai étudie le problème qui consiste à caractériser l'identification de paramètres non-linéaires dans des modèles non-linéaires. Le cadre considéré est semi-paramétrique. Nous supposons que les paramètres structurels d'intérêt déterminent les paramètres identifiables via une équation non-linéaire, telle une équation de moment conditionnelle ou inconditionnelle. De tels modèles sont communs en économétrie, et comprennent par exemple des modèles non-linéaires estimés par la méthode généralisée des moments (GMM) et divers modèles dynamiques d'équilibre général (DGSE) utilisés en macroéconomie et en finance. À nouveau, nous considérons le cas où tous les paramètres peuvent ne pas être identifiables, dans le but de caractériser les transformations non-linéaires de paramètres qui sont identifiables. La littérature sur ce sujet est très mince, et les quelques résultats disponibles considèrent l'identification du vecteur complet des paramètres du modèle. Étant donné que l'identification globale est difficile à caractériser dans les modèles non-linéaires, nous étudions l'identification d'un point de vue local. Nous donnons des conditions suffisantes d'identification, ainsi que des conditions nécessaires et suffisantes sous des hypothèses de différentiabilité des conditions de moments et des transformations des paramètres. Nous proposons aussi une notion d'identification intermédiaire entre l'identification globale et l'identification en un point: l'identification autour d'un point. Celle-ci joue une rôle important dans la théorie proposée. Nous généralisons certains résultats classiques d'identification dans les modèles de vraisemblance. Finalement, nous appliquons nos résultats à l'identification dans les modèles DSGE.

Contents

List of Definitions, Assumptions, Propositions and Theorems

Chapter 1

Introduction

The earliest study on identification can be traced back to Lenoir (1913); see Fox (1968) and Christ (1985). However, identification has not become a heated topic until the seminal work by Koopmans and Reiersøl (1950) who look at identification issue through reformulation of specification problems. Statistical inference on linear regression models has generally focused on the estimability of parameters and hypothesis testing on estimable parameters. In comparison, identification has been relatively less studied in the literature. Nevertheless, it is now widely accepted that identification should be treated independently of statistical inference and it is always advisable to check identifiability of parameters before we conduct any statistical inference; see Manski (1995). Without identification of parameters, distributional theories and statistical methods could become unreliable and even misleading; see Dufour and Hsiao (2008). Even though researchers realize the potential threats imposed on by identification failures, it is usually difficult to verify identification directly from empirical models especially when they involve high dimensional and highly nonlinear functions of parameters. Thus the literature usually assume identification without serious check on it.

However, due to the fact that there exists a relationship between identifiability and estimability [see Chipman (1964), Reiersøl (1963), Rothenberg (1971), Bunke and Bunke (1974) and Seely (1977)], i.e., estimability implies identifiability but the converse is not generally true, it is interesting to ask whether it is possible to obtain identification conditions from well-established results regarding estimability in the sense of Bose (1944) in linear models. The first essay not only provides the above question with an affirmative answer, but also proposes both necessary and sufficient conditions for identification in the much wider class of partially linear models where the classical conditions for estimability do not hold since $X\beta$ in partially linear models is no longer the mean of the response variable but rather a monotone nonlinear function of the mean and the maximum like-

lihood estimators obtained through numerical methods are generally biased. A partially linear model includes linear regression models with collinearity as well as a wide spectrum of other models used in econometrics, such as linear median regressions and quantile regressions, generalized linear mixed models, probit and Tobit models, multinomial logit models, discrete choice models, exponential models, index models, etc. This paper focuses on identification of an arbitrary linear transformation $Q\beta$ given that $X\beta$ is identifiable, where the design matrix X may not have a full-column rank. We propose a necessary and sufficient condition for identification of a transformation of model parameters, which is applied to partially linear models where the class of identifiable parameters can be fully characterized. Equivalent closed-form representations are also provided to easily check identifiability through properties of linear subspaces and generalized inverse techniques. We also consider the effect of restrictions on identification for partially linear models.

The general identification conditions for the partially linear models can be either local or global based on the specific setups of the models. Since recent research on weak instruments mainly focuses on linear IV regressions and SEMs, it is worthwhile investigating parametric function identification of the SEMs in the most general framework. The second essay studies global identification conditions in a linear system, i.e., both the structural equations and the prior restrictions are linear in structural parameters. It aims to answer the following questions. Under what circumstances can the arbitrary linear combinations of parameters Q(X)A be identified although A itself is not identifiable? What are the necessary and sufficient conditions for identifiability when nuisance parameters are involved? How can we check identification of the SEMs in a more straightforward way compared to the classical rank and order conditions? This paper provides global identification conditions in the most general way by relaxing a series of classical assumptions. To begin with, since the full rank matrix of the exogenous variables and the nonsingular coefficient matrix of the endogenous variables are only conducive to identifying the reduced form parameters and not directly related to the identification of structural parameters, identification of the deep parameters can be achieved even if the reduced form equation does not have a unique expression. Besides, the SEMs setup is extended to include instrumental variables X. Therefore, we can study identification through conditional expectation on instruments, which means the parameter identification depends on the realized value of instruments. The relaxation of exogenous or fixed X provides flexibility of studying identification in that it helps deal with some essential issues in econometrics, such as many instruments, weak instruments and missing instruments. In addition, our SEMs setup allows for the presence of nuisance parameters and also the distribution of the structural shocks can have a mean-variance structure that is not separable. The proposed general rank condition is applied to

both univariate structural equation and multivariate structural equations. It is shown that the influential global identification results by Rothenberg (1971) and Richmond (1974) can be easily included as special cases of our general rank condition.

Global identification of parameters of interest θ could be achieved for linear regression models under strong assumptions and usually researchers can only obtain local identification for nonlinear regression models; see Rothenberg (1971). The third essay looks into identification conditions of a general nonlinear function $\beta(\theta)$ in terms of another identifiable nonlinear function $\gamma(\theta)$ in nonlinear models and we name it "parametric function identification". In comparison to the first and second essays, we assume both $\beta(\theta)$ and $\gamma(\theta)$ are nonlinear functions of θ. Hence, the general conditions for global identification based on the ranks of X and Q in linear models has to be replaced by the conditions for local identification through the relationship between the ranks of the Jacobian matrix of $\beta(\theta)$ and the Jacobian matrix of $\gamma(\theta)$. This is a major breakthrough in identification research in that we do not impose any restriction on these Jacobian matrices and it is still possible to identify some function of θ despite the failure to fully identify θ. More specifically, despite the situation where none of the structural parameters may be identified, it is still possible to characterize some nonlinear transformation which is identifiable. Nevertheless, literature on this topic is very scarce and focuses on full identification of the entire parameter vector. Furthermore, we are interested in point identification of parametric functions which is an extension to the partial identification in the sense of Phillips (1989) and Bekker and Wansbeek (2001) but is different from set-valued identification by Manski (2003) and Tamer (2010). Moreover, since we assume that the probability distribution of observed variables depends not only on parameters of interest but also nuisance parameters which are allowed to be infinite in number, the setup of this paper is fundamentally semiparametric. Besides, parameters of interest can determine a number of parameters through a nonlinear equation, for instance, a nonlinear conditional or unconditional moment equation. Thus our framework covers moment conditions as a special case. On the other hand, from a statistician or an econometrician point of view, it is inadequate to just establish local identification conditions at a specific parameter value θ_0 as the current literature does. Rather it makes more sense to study local identification within a neighborhood of θ_0 so that we can justify an estimator as locally identifiable, which lays a solid foundation for reliable statistical inference. This paper provides both necessary and sufficient conditions for local identification under the assumption that the relevant moments equations or nonlinear transformations are differentiable. On top of that, we take into account the restrictions on parameters of interest which represent hypotheses about parameters. Therefore the proposed necessary and sufficient conditions are very useful for statistical inference purpose. Last but not least,

this paper is conducive to empirical research for identification. We demonstrate that the proposed identification conditions of nonlinear parametric functions can generalize a number of well-known classical rank conditions as special cases, whether these conclusions are reached for SEMs by Fisher (1966), Rothenberg (1971), Bowden (1973), Bekker, Merckens and Wansbeek (1994) and Chen, Chernozhukov, Lee and Newey (2011) or for the DSGE models by Iskrev (2010) and Komunjer and Ng (2011).

The thesis is organized as follows. The second chapter proposes a general global identification condition for a general function of parameters based on separability assumption. Chapter Three focuses on global identification for linear SEMs and IV regressions. The fourth chapter deals with local identification of nonlinear parametric functions in nonlinear models. Chapter Five concludes.

Chapter 2

Parameter separability, canonical representations, estimability and global partial identification of linear parameters

Jean-Marie Dufour and Xin Liang

Abstract In this paper, we give necessary and sufficient conditions for global pointwise identification of certain parameters when the complete parameter vector is not identified. We first derive a widely applicable condition for the identification of a parameter transformation (linear or nonlinear) in a general statistical model. The proposed identification condition relies on parameter sufficiency and a general form of (potentially nonlinear) separability. These results are then applied to study the identifiability of linear parameters in partially linear models. A model is partially linear for the parameter vector β if the conditional distribution of the data given X depends on β through $X\beta$, where X is a known matrix. We focus here on situations where X may not have full-column rank, and $X\beta$ can be interpreted as an identifiable parameter (such as the mean of the observations). Besides linear regressions with possibly non-scalar error covariance matrix (allowing for heteroskedasticity and serial dependence), partially linear models include several widely used statistical models: generalized linear models and linear mixed models, median regression, quantile regressions, various discrete choice models (such as probit and Tobit models), single index models, etc. The relationship between estimability and identifiability of linear parameters in partially linear models is studied. We observe that usual condi-

tions for parameter estimability in linear regressions – *a fortiori* in partially linear models – are not necessary for identification, so estimability is not equivalent to identifiability. We characterize the identification of an arbitrary vector $Q\beta$ in a partially linear model, where Q is a fixed matrix. Several equivalent identifiability conditions are provided, and closed-form representations are provided for the corresponding "identification sets". The proposed identifiability conditions include a number of easily interpretable conditions not previously supplied in the literature on estimability. Among several possible alternative representations, we propose two representations which are both unique in all cases and intuitively appealing: (1) the QR representation and (2) the Hermite representation. In particular, the Hermite representation clearly separates the components of β which are individually identifiable from those which are not, and provides parsimonious interpretable representations for the other identifiable linear parameters. Several applications of the proposed results are finally discussed.

2.1. Introduction

A central feature which determines inference on a statistical model is parameter identification. By that, we mean whether parameter values are uniquely determined by the distribution of the observations. This concept has a long history in statistics and econometrics; see the review in Dufour and Hsiao (2008) and the references therein.

Identification theory most often focuses on formulating conditions for identifying the full parameter vector associated with a model. Identification theory and the relevant inference procedures are much more difficult when full identification fails. Models, however, may easily not be fully identifiable, even though some parameters (or parameter transformations) may be identifiable (*partial identification*). Similarly, results on *global identification* are much scarcer than partial identification ones (which characterize identification in the neighborhood of a point); see Komunjer (2012). In this paper, we focus on global necessary and sufficient identification conditions which allow for identification failure for the model as a whole. We make two types of contributions to this topic.

First, we consider a general statistical model and define a flexible notion of identifiability as a relation between two parametric functions. In this comprehensive framework, we derive a widely applicable condition for global identification of certain transformations (linear or nonlinear) of model parameters. The proposed identification condition relies on parameter sufficiency [Barankin (1960)] and a general form of (potentially nonlinear) parameter separability. This result is applicable in parametric, semiparametric, and even nonparametric models (in the sense that finite and infinite-dimensional parameters may

be considered). The "nature" of the parameters is weakly specified: the parameters and parametric functions considered may be points, sets, functions, etc.

Second, we reconsider one of the most basic identification problems, namely the identification of "linear parameters". More precisely, if β is a $k \times 1$ real vector, we consider the problem of identifying $Q\beta$ when $X\beta$ is identifiable (although β may not be identifiable), where Q and X are $q \times k$ and $n \times k$ real matrices. A standard setup of this type is the linear regression model

$$y = X\beta + u \tag{2.1.1}$$

where $\mathbb{E}(y) = X\beta$. If X does not have full-column rank, then β is typically not identifiable.

Linear models where the regressor matrix X has deficient rank show up easily in the presence of dummy explanatory variables or structural shifts [see Dufour (1982)]. Allowing for regressor matrices with deficient rank provides one with great flexibility in selecting regressors. In particular, no arbitrary regressor selection or reparameterization is needed, so the model may be studied using a coordinate-free approach [Kruskal (1961, 1968), Dufour (1982)]. As a result, many standard treatments of linear regression models allow for regressor matrices with deficient rank; see, for example, Scheffé (1959), Searle (1971), Rao (1973), Magnus and Neudecker (1998) and Christensen (2011). Situations where the number of regressors exceeds the number of observations can be considered, such as the estimation of reduced form equations when the number of instruments is larger than the number of equations and when studying structural change [Dufour (1982), Cantrell, Burrows and Vuong (1991)]. Interestingly, the fact that the number of explanatory variables exceeds the number of observations does not preclude identification of certain parameters. We provide here conditions for deciding which linear parameters remain identifiable in such situations.

Other models involving linear parameters include: (1) median regression, where $X\beta$ represents the medians of the observations in y [see Coudin and Dufour (2009)]; (2) quantile regression, where $X\beta$ contains the q-quantiles of the components of y [Koenker (2005)]; (3) various models based on distributions in exponential families [Lehmann (2005, Chapter 2)]; (4) many discrete choice models, including probit, Tobit, and multinomial logit models [Gouriéroux (2000), Lewbel (2000), Collett (2003*a*, 2003*b*)]; (5) generalized linear models [Nelder and Wedderburn (1972), McCullagh and Nelder (1989), Diggle, Heagerty, Liang and Zeger (2002), Jiang (2007), Dobson and Barnett (2008)]; (6) index models [Horowitz (2009)]; etc. When the distribution of the data involves a parameter of the form $X\beta$, we will say that the model is *partially linear* with respect to β. In such models, $X\beta$ may not represent the mean of y, but another feature of the distribution. The fact X has full column rank is typically taken for granted in this literature. In general partially linear models where X

can have arbitrary rank, the problem of identifying parameter transformations such as $Q\beta$ has been barely studied, if at all. It turns out that the generic identification characterization discussed above will be useful for that purpose.

Discussions of statistical inference on linear regression models have generally focused on characterizing, estimating and testing so-called *estimable* parameters. Following the introduction of this notion by Bose (1944), a parameter $Q\beta$ is "estimable" in the context of a linear regression if it is possible to find an unbiased linear estimator of $Q\beta$. The standard condition for the estimability of $Q\beta$ is the existence of a matrix B such that $Q' = X'B$; see Rao (1945, 1973) and Scheffé (1959). Various estimability conditions have been proposed, and the literature on this topic is now considerable; see for instance Tuncer (1985), Seely (1970), Alalouf and Styan (1979*a*), Alalouf and Styan (1979*b*), Marsaglia and Styan (1974), Milliken (1971), Baksalary and Kala (1976), Eubank and Webster (1985). In more general partially linear models, $Q\beta$ may not be estimable even if $Q' = X'B$ holds.

The relationship between identifiability and estimability has been occasionally discussed in the literature; see Chipman (1964), Reiersøl (1963), Rothenberg (1971), Bunke and Bunke (1974), Seely (1977), Paulino and de Bragança Pereira (1994), Dufour and Hsiao (2008) and Christensen (2011, Ch. 2). It is easy to see that an estimable parameter must be identifiable (estimability entails identifiability); see Rothenberg (1971, Theorem 4). But the converse is not generally true: even if $Q\beta$ is not identifiable through $X\beta$, it could be identifiable through other features of the distribution. In section 2.5 of this paper, we provide a simple counter-example. There are special cases where indeed identifiability and estimability are equivalent [such as the classical linear model discussed by Reiersøl (1963), Bunke and Bunke (1974), Seely (1977)], but this setup is quite restrictive. It would be useful to have more general conditions for ensuring such an equivalence. In this paper, we study the relation between identifiability and estimability in partially linear models along the following lines.

First, on observing that standard conditions for estimability may not be necessary for identifying a linear parameter $Q\beta$, we give a general condition under which conditions for estimability are indeed necessary and sufficient. This result has independent interest which goes beyond the framework of linear or partially linear models. Indeed, no linear structure is required and the conditions are applicable even if the model contains unidentified nuisance parameters.

Second, we show that various conditions proposed for checking estimability in linear regression can in fact be applied to check identifiability in the much wider class of partially linear models. For the sake of generality, we also allow for the presence of linear restrictions on regression coefficients [like Baksalary and Kala (1976), Seely (1977), and Alalouf

and Styan (1979*b*)] as well as a general covariance matrix (characterizing heteroskedasticity and serial dependence) which may depend on unknown parameters.

Third, out of looking at estimability from the viewpoint of identifiability, we derive alternative conditions for both identifiability and estimability which are intuitive and easy to apply. Namely, the condition $Q' = X'B$ states that the coefficient vector that defines each component of $Q\beta$ is a linear combination of the observations in X (which correspond to the rows of X). The required conditions on the explanatory variables (the columns of X) of the model are not apparent. Since empirical investigators usually think in terms of the variables of a model, such a condition may not be easy to understand and apply in practice. We propose new estimability conditions based on the required relationships between model explanatory variables.

Fourth, we discuss how the complete set of identifiable parameters may be represented in a convenient way through a canonical representation which is both unique and interpretable. We observe that, among several representation, two representations are especially appealing: (1) the QR representation and (2) the Hermite normal representation.

Fifth, the results presented are applied to linear regression models with possibly non-scalar covariance matrix – allowing for heteroskedasticity and error dependence which may depend on unknown parameters – as well as several partially linear models.

In Section 2.2, we recall and complete some basic definitions relevant to the concept of identification considered in this paper, in particular identification of a general parametric function in terms of another parametric function. In Section 2.3, we give general necessary and conditions for identification of a (possibly nonlinear) parametric vector function. In Section 2.4, we review alternative conditions for the existence of functional relations between subspaces of Euclidean vector space. In Section 2.5, we give general necessary and sufficient conditions for partial global identification of linear parameters in linear regression with possibly non-scalar covariance matrix which allows for heteroskedasticity and error dependence. Canonical identifiable representations are proposed in Section 2.6. Applications to (non-regression) partially linear models are discussed in Section 2.7. A number of more specific examples are discussed in Section 2.8. We conclude in Section 2.9. The proofs are presented in Appendix.

2.2. Identification

In this section, we recall and complete some basic definitions relevant to the concept of identification. We consider a general parameter $\psi(\theta)$ in models parameterized by θ.

Let $(\Omega, \mathscr{A}, \mathscr{P})$ be a statistical model, where Ω is a sample space, $\mathscr{A}$ is a σ-algebra

of subsets of Ω, $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ is a family of probability measure on $(\Omega, \mathscr{A})$, and Θ is a parameter space. Let also Θ_0 be a subset of Θ ($\Theta_0 \subseteq \Theta$). We consider a general parametric function $\psi : \Theta_0 \to \Psi$. At this point, no restriction is imposed on the dimensions of Θ, Θ_0 or Ψ, so both parametric and nonparametric models are allowed. In particular, there is no restriction on the structure of the space Ψ, which may consist of points in a Euclidean vector space, subsets of another space (e.g., equivalence classes [1]), functions, etc. Reference to subsets Θ_0 of Θ will allow us to consider functions ψ which may not be defined everywhere, as well as restrictions on the parameter space.

Definition 2.2.1 IDENTIFICATION OF PARAMETRIC FUNCTIONS. *Let $\Theta_0 \subseteq \Theta$. The parametric function $\psi : \Theta_0 \to \Psi$ is* identifiable *on Θ_0 if and only if*

$$\big(\psi(\theta_1) \neq \psi(\theta_2)\big) \Rightarrow \big(\mathsf{P}_{\theta_1} \neq \mathsf{P}_{\theta_2}\big), \quad \forall \theta_1, \theta_2 \in \Theta_0. \tag{2.2.1}$$

This definition emphasizes the fact that different values of $\psi(\theta)$ can be distinguished empirically by looking at the distribution of the data. Condition (2.2.1) is equivalent to

$$\big(\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2}\big) \Rightarrow \big(\psi(\theta_1) = \psi(\theta_2)\big), \quad \forall \theta_1, \theta_2 \in \Theta_0. \tag{2.2.2}$$

When Θ_0 is not empty ($\Theta_0 \neq \varnothing$),the parameter $\psi(\theta)$ is identifiable if it can be represented as a function which depends on θ only through P_θ :

$$\psi(\theta) = h_\psi(\mathsf{P}_\theta) \text{ for } \theta \in \Theta_0 \tag{2.2.3}$$

where $h_\psi : \mathscr{P}(\Theta_0) \to \Psi$ and $\mathscr{P}(\Theta_0) = \{\mathsf{P}_\theta : \theta \in \Theta_0\}$. Here, identifiability is defined on some given subset Θ_0 of Θ, which includes $\Theta_0 = \Theta$ as a special case. Even though we wish to focus on cases where $\Theta_0 = \Theta$, it will be useful to have the above general definition. Because the property (2.2.1) holds for every pair of points in Θ_0 [as opposed to neighborhoods of individual points], the identification is "global" for Θ_0.

When $\Theta_0$is empty ($\Theta_0 = \varnothing$), the implication (2.2.1) is trivially satisfied, so we will consider that any parameter $\psi(\theta)$ is identifiable on the empty set. Even though this degenerate case has little practical interest, this convention is convenient and will allow us to simplify exposition.

[1] Equivalence class determined by an element x of a set A, is defined as a certain subset E of A such that

$$E = \{y \,|y \sim x\},$$

where the symbol "$\sim$" is called equivalence relation on the set A which has three properties: reflexivity, symmetry and transitivity. For more details on equivalence relations, please refer to Munkres (2000).

To make Definition **2.2.1** easier to apply, we shall focus on the case where two parametric functions $\delta(\theta)$ and $\gamma(\theta)$ are considered, and one of them [say, $\gamma(\theta)$] is identifiable. If

$$\big(\delta(\theta_1) \neq \delta(\theta_2)\big) \Rightarrow \big(\gamma(\theta_1) \neq \gamma(\theta_2)\big), \quad \forall \theta_1, \theta_2 \in \Theta_0, \tag{2.2.4}$$

or, equivalently, if

$$\delta(\theta) = \bar{\delta}\big(\gamma(\theta)\big), \quad \forall \theta \in \Theta_0, \tag{2.2.5}$$

for some function $\bar{\delta} : \Gamma \to B$, it is straightforward to see that $\delta(\theta)$ is identifiable on Θ_0. In such cases, we will say that $\delta(\theta)$ is *identifiable by* $\gamma(\theta)$ or $\gamma(\theta)$*-identifiable* on Θ_0. For future reference, we state formally this definition.

Definition 2.2.2 IDENTIFICATION IN TERMS OF ANOTHER PARAMETRIC FUNCTION. *Let $\delta : \Theta_0 \to B$ and $\gamma : \Theta_0 \to \Gamma$ be two parametric functions on the statistical model $(\Omega, \mathscr{A}, \mathscr{P})$ where $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$, where $\Theta_0 \subseteq \Theta$. If $\gamma(\theta)$ is identifiable on Θ_0 and condition* (2.2.4) *holds, we say that $\delta(\theta)$ is $\gamma(\theta)$-identifiable on Θ_0.*

Constraints are often imposed on model parameters, such as

$$\rho(\theta) = \rho_0, \quad \forall \theta \in \Theta_0, \tag{2.2.6}$$

where $\rho : \Theta_0 \to C$ is another parametric function and $\rho_0 \in C$. In such cases, $\rho(\theta)$ can also be viewed as a "parameter". Following Definition **2.2.2**, $\rho(\theta)$ is identifiable on Θ_0 under the restriction (2.2.6), because we have

$$\rho(\theta_1) = \rho(\theta_2), \quad \forall \theta_1, \theta_2 \in \Theta_0, \tag{2.2.7}$$

and the implication

$$\big(\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2}\big) \Rightarrow \big(\rho(\theta_1) = \rho(\theta_2)\big), \quad \forall \theta_1, \theta_2 \in \Theta_0,$$

immediately holds. In other words, imposing a parametric restriction is equivalent to the introduction of an identifiable parameter [in addition to $\gamma(\theta)$].

2.3. Necessary and sufficient conditions for global partial identification

A condition of the form (2.2.4) is sufficient for identification, but it is not necessary. In many cases, it is also useful to have conditions which are both necessary and sufficient. We

will now give general conditions of this type, which can be applied in many setups.

Following Barankin (1960), we consider the case where the parametric function $\gamma(\theta)$ is "sufficient" on Θ_0 according to the following definition.

Definition 2.3.1 PARAMETRIC FUNCTION SUFFICIENCY. *Let $\gamma : \Theta \to \Gamma$ a parametric function on the statistical model $(\Omega, \mathscr{A}, \mathscr{P})$ with $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$, and $\varnothing \neq \Theta_0 \subseteq \Theta$. The parametric function $\gamma(\theta)$ is sufficient for $\mathscr{P}$ on Θ_0 if and only if*

$$\big(\gamma(\theta_1) = \gamma(\theta_2)\big) \Rightarrow \big(\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2}\big), \quad \forall \theta_1, \theta_2 \in \Theta_0 . \tag{2.3.1}$$

The above sufficiency property means that P_θ can be parameterized in terms of $\gamma(\theta)$ at least for $\theta \in \Theta_0$, *i.e.* we can find a family of probability measures $\bar{\mathscr{P}}(\Theta_0) = \{\bar{\mathsf{P}}_\theta : \theta \in \Theta_0\}$ on $(\Omega, \mathscr{A})$ such that $\mathsf{P}_\theta = \bar{\mathsf{P}}_{\gamma(\theta)}$ for $\theta \in \Theta_0$.[2] Note $\gamma(\theta)$ *need not be identifiable*. Following Paulino and de Bragança Pereira (1994) and Dasgupta, Self and Gupta (2007), we will say that $\gamma(\theta)$ is "identifying" on Θ_0 if $\gamma(\theta)$ is both sufficient and identifiable on Θ_0. When this is the case, it is easy to see that condition (2.2.4) is both necessary and sufficient (as opposed to only sufficient) for identifying another parametric function $\delta(\theta)$. For completeness sake, we state this basic observation in the following proposition.

Proposition 2.3.2 NECESSARY AND SUFFICIENT CONDITION FOR IDENTIFICATION WITHOUT NUISANCE PARAMETERS. *Let $\delta : \Theta_0 \to B$ and $\gamma : \Theta_0 \to \Gamma$ be two parametric functions on the statistical model $(\Omega, \mathscr{A}, \mathscr{P})$ with $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$, and $\varnothing \neq \Theta_0 \subseteq \Theta$. Suppose $\gamma(\theta)$ is both sufficient for $\mathscr{P}$ on Θ_0 and identifiable on Θ_0. Then, $\delta(\theta)$ is identifiable on Θ_0 if and only if*

$$\big(\delta(\theta_1) \neq \delta(\theta_2)\big) \Rightarrow \big(\gamma(\theta_1) \neq \gamma(\theta_2)\big), \quad \forall \theta_1, \theta_2 \in \Theta_0 . \tag{2.3.2}$$

The assumption that $\gamma(\theta)$ is identifying is however quite restrictive. The typical situation is one where $\delta(\theta)$ depends on $\gamma(\theta)$, while P_θ also depends on nuisance parameters. We will now consider a general setup which allows for nuisance parameters. In such cases, we will need restrictions on the family of probability distributions in order to obtain necessary and sufficient conditions for identification of $\delta(\theta)$.

Besides $\delta : \Theta_0 \to B$, we consider two other functions $\gamma_1 : \Theta_0 \to \Gamma_1$ and $\gamma_2 : \Theta_0 \to \Gamma_2$, one of which $[\gamma_2(\theta)]$ may be interpreted as a "nuisance parameter". We now make the following generic assumption.

[2] For further discussion of the concept of parameter sufficiency, the reader may consult: Florens, Mouchart and Rolin (1985), Florens, Mouchart and Rolin (1990), Paulino and de Bragança Pereira (1994), and Oulhaj and Mouchart (2003).

Assumption 2.3.3 NUISANCE PARAMETER SEPARABILITY. *For any $\theta \in \Theta_0$ and $\bar{\gamma}_2 \in \gamma_2(\Theta_0)$, we can find $\theta^* \in \Theta_0$ such that*

$$\delta(\theta^*) = \delta(\theta)\,, \quad \gamma_1(\theta^*) = \gamma_1(\theta) \text{ and } \gamma_2(\theta^*) = \bar{\gamma}_2\,. \tag{2.3.3}$$

The latter condition means that $\gamma_2(\theta)$ is not restricted by setting the values of $\delta(\theta)$ and $\gamma_1(\theta)$, so (in this sense) the parametric function $\gamma_2(\theta)$ is "free" with respect to $\delta(\theta)$ and $\gamma_1(\theta)$. We can then state the following necessary and sufficient condition for the identification of $\delta(\theta)$.

Theorem 2.3.4 NECESSARY AND SUFFICIENT CONDITION FOR IDENTIFICATION WITH NUISANCE PARAMETERS. *Let $\delta : \Theta_0 \to B$, $\gamma_1 : \Theta_0 \to \Gamma_1$ and $\gamma_2 : \Theta_0 \to \Gamma_2$ be three parametric functions on the statistical model $(\Omega, \mathscr{A}, \mathscr{P})$ with $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$, $\gamma(\theta) = \big(\gamma_1(\theta),\ \gamma_2(\theta)\big) \in \Gamma_1 \times \Gamma_2$, where $\varnothing \neq \Theta_0 \subseteq \Theta$. Suppose the following three conditions hold:*

(1) *Assumption* **2.3.3** *(nuisance parameter separability) is satisfied;*

(2) *$\gamma(\theta)$ is sufficient for $\mathscr{P}$ on Θ_0;*

(3) *$\gamma_1(\theta)$ is identifiable on Θ_0.*

Then, $\delta(\theta)$ is identifiable on Θ_0 if and only if

$$\big(\delta(\theta_1) \neq \delta(\theta_2)\big) \Rightarrow \big(\gamma_1(\theta_1) \neq \gamma_1(\theta_2)\big), \quad \forall \theta_1, \theta_2 \in \Theta_0\,. \tag{2.3.4}$$

In the latter theorem, there is no identifiability assumption on $\gamma_2(\theta)$ or $\gamma(\theta)$. The parameter sufficiency condition for $\gamma(\theta)$ means the distribution P_θ depends on θ only through $\gamma_1(\theta)$ and $\gamma_2(\theta)$. Then, when $\gamma_1(\theta)$ is identifiable and $\gamma_2(\theta)$ is "free" as described by Assumption **2.3.3**, condition (2.3.4) is necessary and sufficient for $\delta(\theta)$ to be identifiable (on Θ_0). Clearly, (2.3.4) is equivalent to

$$\big(\gamma_1(\theta_1) = \gamma_1(\theta_2)\big) \Rightarrow \big(\delta(\theta_1) = \delta(\theta_2)\big), \quad \forall \theta_1, \theta_2 \in \Theta\,, \tag{2.3.5}$$

which means there is a function $\bar{\delta}(\cdot)$ such that

$$\delta\big(\theta\big) = \bar{\delta}\big(\gamma_1(\theta)\big)\,, \quad \forall \theta\,. \tag{2.3.6}$$

We now give alternative conditions which may be easier to check in practice.

Assumption 2.3.5 REDUCIBILITY. *There is a function $\phi : \Theta_0 \to \Delta$, such that*

$$\big(\phi(\theta_1) = \phi(\theta_2)\big) \Rightarrow \big(\delta(\theta_1) = \delta(\theta_2) \text{ and } \gamma_1(\theta_1) = \gamma_1(\theta_2)\big), \quad \forall \theta_1, \theta_2 \in \Theta_0 . \quad (2.3.7)$$

Assumption 2.3.6 SEPARABILITY. *The function $\phi : \Theta_0 \to \Delta$ is separable from $\gamma_2 : \Theta_0 \to \Gamma_2$ in the following sense: for any $\theta \in \Theta_0$ and $\bar{\gamma}_2 \in \gamma_2(\Theta_0)$, we can find $\theta^* \in \Theta_0$ such that*

$$\phi(\theta^*) = \phi(\theta) \text{ and } \gamma_2(\theta^*) = \bar{\gamma}_2 . \quad (2.3.8)$$

Assumption **2.3.5** means that both $\gamma_1(\theta)$ and $\delta(\theta)$ can be expressed as functions of $\phi(\theta)$:

$$\delta(\theta) = g[\phi(\theta)] \text{ and } \gamma_1(\theta) = h_1[\phi(\theta)] , \quad \forall \theta \in \Theta_0 , \quad (2.3.9)$$

for some functions $g : \Delta \to B$ and $h_1 : \Delta \to \Gamma_1$.

Proposition 2.3.7 REDUCIBILITY CHARACTERIZATION OF NUISANCE PARAMETER SEPARABILITY. *Assumption* **2.3.3** *is equivalent to the conjunction of Assumptions* **2.3.5** *and* **2.3.6**.

It is also interesting to observe that the separability property in Assumption **2.3.6** can be reformulated as an assumption on the image sets of the functions $\phi : \Theta_0 \to \Delta$ and $\gamma_2 : \Theta_0 \to \Gamma_2$ as follows.

Proposition 2.3.8 IMAGE SET CHARACTERIZATION OF NUISANCE PARAMETER SEPARABILITY. *Let $\nu : \Theta_0 \to \Delta \times \Gamma_2$ be the function defined by $\nu(\theta) = \big(\phi(\theta), \gamma_2(\theta)\big)$. Then the separability Assumption* **2.3.6** *is equivalent to*

$$\nu(\Theta_0) = \phi(\Theta_0) \times \gamma_2(\Theta_0) . \quad (2.3.10)$$

In the context of the latter formulation, neither $\phi(\theta)$ nor $\gamma_2(\theta)$ need be identifiable. Only $\gamma_1(\theta)$ is taken to be identifiable. An important special case of conditions **2.3.5** - **2.3.6** is the one where $\phi(\theta)$ is a subvector of θ; for example, we may have:

$$\theta = \begin{bmatrix} \theta^{(1)} \\ \theta^{(2)} \end{bmatrix}, \ \phi(\theta) = \theta^{(1)} . \quad (2.3.11)$$

In this case, Assumption **2.3.6** reduces to the case where $\theta^{(2)}$ can move freely of $\theta^{(1)}$. Again, neither $\theta^{(1)}$ nor $\theta^{(2)}$ needs to be identifiable.

Theorem **2.3.4** is quite general. Both linear and nonlinear models are covered, and the

identifiable parametric function $\gamma_1(\theta)$ is not restricted to be mean of the observations. We will give an example to illustrate this argument in section 2.5.

2.4. Functional dependence between linear subspaces

In view of studying the identification of linear parameters, we will now review and extend some useful linear algebra results with the purpose of explicitly formulating functional dependence properties between subspaces of a Euclidean space. Even though these properties follow directly from general linear algebra, the formulation given here – in particular, the interpretation in terms of functional dependence – is not available elsewhere. It is especially well adapted to study identification. Unless otherwise stated, we make no rank assumption on the matrices considered.

For any real $m \times n$ matrix A, we denote the column (or image) space of A by $\mathrm{Im}(A) \equiv \{y \in \mathbb{R}^m : Ax = y \text{ for some } x \in \mathbb{R}^n\}$, and the corresponding kernel space by $\ker(A) \equiv \{x \in \mathbb{R}^n : Ax = 0\}$. Further, any $n \times m$ matrix A^- such that $AA^-A = A$ is called a *generalized inverse* (or *g-inverse*) of A; see Rao and Mitra (1971). We denote the i-th row of A by $A_{i\cdot}$ and its j-th column by $A_{\cdot j}$, so $A_{i\cdot}$ is a $1 \times n$ matrix and $A_{\cdot j}$ an $m \times 1$ matrix. Further, $A_{(i\cdot)}$ is the $(m-1) \times n$ matrix obtained by suppressing the i-th row of A, and $A_{(\cdot j)}$ the $m \times (n-1)$ matrix obtained by suppressing the j-th column of A.

For future reference, we first state a basic lemma linking the rank of a matrix to the ranks of linear transformations of its submatrices. We could not find a source where this lemma is stated in appropriate form.

Lemma 2.4.1 CONDITIONS FOR SUBMATRIX RANK ADDITIVITY. *Let Z, Q, V_1 and V_2 be $p \times k$, $q \times k$, $q \times p$ and $p \times q$ real matrices. Then,*

$$\mathrm{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \mathrm{rank}\begin{bmatrix} Z \\ Q + V_1 Z \end{bmatrix} = \mathrm{rank}\begin{bmatrix} Z + V_2 Q \\ Q \end{bmatrix} \tag{2.4.1}$$

and the following rank additivity properties hold:

(*a*) *if*

$$\mathrm{Im}(Z') \cap \mathrm{Im}(Q' + Z'V_1{}') = \{0\} \tag{2.4.2}$$

then

$$\mathrm{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \mathrm{rank}(Z) + \mathrm{rank}(Q + V_1 Z)\,; \tag{2.4.3}$$

(*b*) *if*

$$\mathrm{Im}(Q') \cap \mathrm{Im}(Z' + Q'V_2') = \{0\} \tag{2.4.4}$$

then

$$\mathrm{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \mathrm{rank}(Z + V_2 Q) + \mathrm{rank}\,(Q)\,; \tag{2.4.5}$$

(*c*) *if* $V_1 = -QZ^-$ *for some g-inverse* Z^- *of Z, then* (2.4.2) *and* (2.4.3) *hold;*

(*d*) *if* $V_2 = -ZQ^-$ *for some g-inverse* Q^- *of Q, then* (2.4.4) *and* (2.4.5) *hold.*

We will now give several characterizations of functional dependence between two linear subspaces.

Proposition 2.4.2 CONDITIONS FOR FUNCTIONAL DEPENDENCE BETWEEN LINEAR SUBSPACES. *Let Z and Q be* $p \times k$ *and* $q \times k$ *real matrices. Then the following statements are equivalent:*

$$\textit{there is a function } g : \mathrm{Im}(Z) \mapsto \mathrm{Im}(Q) \textit{ such that } Q\beta = g(Z\beta),\ \forall \beta \in \mathbb{R}^k\,; \tag{2.4.6}$$

$$[(Z\beta_1 = Z\beta_2) \Rightarrow (Q\beta_1 = Q\beta_2)]\ (\forall \beta_1,\ \beta_2 \in \mathbb{R}^k)\,; \tag{2.4.7}$$

$$\ker(Z) \subseteq \ker(Q)\,; \tag{2.4.8}$$

$$\mathrm{Im}(Q') \subseteq \mathrm{Im}(Z')\,; \tag{2.4.9}$$

$$Q = BZ, \textit{ for some matrix } B\,; \tag{2.4.10}$$

$$\mathrm{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \mathrm{rank}(Z)\,; \tag{2.4.11}$$

$$\mathrm{rank}\begin{bmatrix} Z \\ Q + V_1 Z \end{bmatrix} = \mathrm{rank}(Z), \textit{ for any } q \times p \textit{ matrix } V_1\,; \tag{2.4.12}$$

$$\mathrm{rank}\begin{bmatrix} Z + V_2 Q \\ Q \end{bmatrix} = \mathrm{rank}(Z), \textit{ for any } p \times q \textit{ matrix } V_2\,; \tag{2.4.13}$$

$$\mathrm{rank}\begin{bmatrix} Z \\ SQ \end{bmatrix} = \mathrm{rank}(Z), \textit{ for any matrix } S \textit{ such that } \mathrm{rank}(SQ) = \mathrm{rank}(Q)\,; \tag{2.4.14}$$

$$Q = QZ^-Z, \textit{ for some g-inverse } Z^-\,; \tag{2.4.15}$$

$$\mathrm{rank}[Z(I_k - Q^-Q)] = \mathrm{rank}(Z) - \mathrm{rank}(Q), \textit{ for some g-inverse } Q^-\,. \tag{2.4.16}$$

The above proposition gives conditions for checking the property that $Q\beta$ is determined by $Z\beta$, *i.e.* $Q\beta$ is a function of $Z\beta$ [as stated by (2.4.6)]. Interestingly, no additional *a priori* assumption on the form of the function $g(\cdot)$ is needed, and no restriction is imposed on the ranks of Z and Q. In particular, the rows of Q may be linearly dependent, which allows one to consider situations with "redundant parameters" in a transparent way. The equivalence with (2.4.7) simply means the functional dependence property is an injective relation between $Z\beta$ to $Q\beta$. (2.4.8) and (2.4.9) show that functional dependence can be interpreted as the nesting of subspaces: either between the column spaces of Z and Q, or between the kernel spaces of Q' and Z'. (2.4.10) expresses the fact that the rows of Q must be linear combinations of the rows of Z. (2.4.11) - (2.4.14) allow one to check functional dependence by looking at the ranks of extended matrices built from linear transformations of Z and Q. Finally, (2.4.15) and (2.4.16) show that functional dependence can be handily expressed through projection operators. This follows on observing that $(Z^-Z)'$ is a projection matrix on the space $\mathrm{Im}(Z')$, while equation (2.4.15) can be rewritten as

$$(Z^-Z)'Q' = Q', \tag{2.4.17}$$

i.e. any column of Q' [or any row of Q] belongs to the space spanned by the columns of $(Z^-Z)'$ [the rows of (Z^-Z)]. Similarly $(I_k - Q^-Q)' = I_k - (Q^-Q)'$ is a projection matrix on $\mathrm{Im}(Q')^\perp = \ker(Q)$, the subspace orthogonal to $\mathrm{Im}(Q')$: each column of the matrix $[I_k - (Q^-Q)']Z'$ is the projection of the corresponding column of Z' on $\ker(Q)$; see Baksalary and Kala (1976) for further discussion of this type of condition in the context of studying estimability.

Clearly, the conditions of Proposition **2.4.2** are satisfied when $Q = 0$, for in this case $Q\beta$ is a constant function ($Q\beta = 0$). An important case to which Proposition **2.4.2** applies is the one where $Q\beta$ is a scalar, *i.e.* $Q = l'$ where l is a $k \times 1$ vector. Any one of the conditions (2.4.7) - (2.4.16) can be used to check the functional dependence of $l'\beta$ on $Z\beta$. In particular, (2.4.9) means that l must belong to the column space of Z', *i.e.*

$$l \in \mathrm{Im}(Z') \tag{2.4.18}$$

while (2.4.11) takes the form

$$\mathrm{rank}\begin{bmatrix} Z \\ l' \end{bmatrix} = \mathrm{rank}(Z)\,. \tag{2.4.19}$$

For example, if $l = (1, 0, \ldots, 0)'$ and $\beta = (\beta_1, \ldots, \beta_k)'$, we have $l'\beta = \beta_1$ the first com-

ponent of β. On writing $Z = \begin{bmatrix} z_1 & Z_2 \end{bmatrix}$ where z_1 is the first column of Z, we get the condition:

$$\operatorname{rank}[z_1 \quad Z_2] = \operatorname{rank}\begin{bmatrix} z_1 & Z_2 \\ 1 & 0' \end{bmatrix} = 1 + \operatorname{rank}(Z_2)\,. \tag{2.4.20}$$

In other words, β_1 is uniquely determined by $Z\beta$ if and only if z_1 is linearly independent of the other columns of Z :

$$z_1 \notin \operatorname{Im}(Z_2) \text{ and } z_1 \neq 0\,. \tag{2.4.21}$$

This holds *irrespective of the rank of* Z_2. More generally, it is easy to see that the k-th component of β is uniquely determined by $Z\beta$ if and only if the k-th column of Z is linearly independent of the other columns of Z.

Identification is often achieved by imposing restrictions on model parameters. To take this into account, we will now extend Proposition **2.4.2** to allow for linear restrictions. There are two distinct ways of doing this.

The first one consists in imposing linear implicit restrictions of the form

$$R\beta = c_0 \tag{2.4.22}$$

where R is an $m \times k$ matrix (with $m \geq 1$). We shall not make any assumption on the rank of R. However, using this type of restriction requires one to make a consistency assumption, *i.e.* $c_0 \in \operatorname{Im}(R)$; without it, the equation (2.4.22) has no solution. Because $R\beta$ is a linear function, we call such restrictions *implicit linear restrictions*.

The second approach consists in imposing explicit restrictions of the form:

$$\beta = b_0 + Ce \text{ for some } e \in \mathbb{R}^m \tag{2.4.23}$$

where b_0 is a $k \times 1$ vector, and C is a $k \times m$ real matrix. Here, b_0 and C are fixed, while e is a free variable, so β belongs to an affine subset of $\mathbb{R}^k$. No condition is imposed on the value of m or the rank of C, so we may have $m \leq k$ or $m > k$. Further, no additional consistency assumption is needed here, because the set $\{\beta : \beta = b_0 + Cd,\ d \in \mathbb{R}^m\} \subseteq \mathbb{R}^k$ is never empty. Because Cd is linear function, we call such restrictions *explicit linear restrictions*, even though the set of values β which satisfy (2.4.23) is an affine subset of $\mathbb{R}^k$ (not a linear subspace).

Implicit linear restrictions can be reexpressed in explicit form, and vice-versa. For future reference, we state precisely this basic property in the following lemma.

Lemma 2.4.3 EQUIVALENCE BETWEEN IMPLICIT AND EXPLICIT LINEAR RESTRICTIONS. *Let R and C be $m \times k$ and $k \times m$ real matrices, respectively, c_0 an $m \times 1$ real*

vector, and b_0 a $k \times 1$ real vector. Then, the following equivalences hold: (a) if $c_0 \in \mathrm{Im}(R)$, then, for any g-inverse R^- of R,

$$\left[R\beta = c_0\right] \Leftrightarrow \left[\beta = R^- c_0 + (I_k - R^- R)d\,, \quad \textit{for some } d \in \mathbb{R}^k\right]; \tag{2.4.24}$$

(b) for any g-inverse C^- of C,

$$\left[\beta = b_0 + Ce \textit{ for some } e \in \mathbb{R}^m\right] \Leftrightarrow \left[(I_k - CC^-)\beta = (I_k - CC^-)b_0\right]. \tag{2.4.25}$$

We underscore again the matrices R and C need not have full rank, so the ranks of R and C can be smaller than m, and the choice of g-inverses R^- and C^- is arbitrary. While (2.4.24) is a standard result of linear algebra, we did not find a source where the sufficiency of the condition $(I_k - CC^-)\beta = (I_k - CC^-)b_0$ in (2.4.25) is stated. The main interest of this condition consists in characterizing the existence of solutions for the representation $\beta = b_0 + Ce$, when C is not invertible. Indeed, when $m = k$ and C is invertible, the condition holds trivially, for then $C^- = C^{-1}$, $I_k - CC^- = 0$, and $\beta = b_0 + Ce$ has a unique solution for e.

Despite the formal equivalence between implicit and explicit formulations, these two approaches yield different types of functional dependence conditions, each of which may be useful in different circumstances. In the following proposition, we consider the case of implicit linear restrictions.

Proposition 2.4.4 CONDITIONS FOR FUNCTIONAL DEPENDENCE BETWEEN SUBSPACES WITH IMPLICIT LINEAR RESTRICTIONS. *Let Z, Q and R be $p \times k$, $q \times k$ and $m \times k$ real matrices, respectively, $c_0 \in \mathrm{Im}(R)$, $\mathscr{L}_I(R, c_0) = \{x \in \mathbb{R}^k : Rx = c_0\}$, and*

$$\bar{Z} = \begin{bmatrix} Z \\ R \end{bmatrix}. \tag{2.4.26}$$

Then the following statements are equivalent:

$$\begin{gathered} \textit{for any } c_0 \in \mathrm{Im}(R), \textit{ there is a function } \; g_{c_0} : \mathrm{Im}(Z) \mapsto \mathrm{Im}(Q) \\ \textit{such that } Q\beta = g_{c_0}(Z\beta),\ \forall \beta \in \mathscr{L}_I(R, c_0); \end{gathered} \tag{2.4.27}$$

$$\textit{there is a function } \; \bar{g} : \mathrm{Im}(\bar{Z}) \mapsto \mathrm{Im}(Q) \textit{ such that } Q\beta = \bar{g}(\bar{Z}\beta),\ \forall \beta \in \mathbb{R}^k; \tag{2.4.28}$$

$$[(\bar{Z}\beta_1 = \bar{Z}\beta_2) \Rightarrow (Q\beta_1 = Q\beta_2)]\ (\forall \beta_1, \beta_2 \in \mathbb{R}^k); \tag{2.4.29}$$

$$\ker(\bar{Z}) \subseteq \ker(Q); \tag{2.4.30}$$

$$\mathrm{Im}(Q') \subseteq \mathrm{Im}(\bar{Z}')\,; \tag{2.4.31}$$

$$Q = B\bar{Z}, \textit{ for some matrix } B\,; \tag{2.4.32}$$

$$\mathrm{rank}\begin{bmatrix} \bar{Z} \\ Q \end{bmatrix} = \mathrm{rank}(\bar{Z})\,; \tag{2.4.33}$$

$$\mathrm{rank}\begin{bmatrix} \bar{Z} \\ Q + V_1\bar{Z} \end{bmatrix} = \mathrm{rank}(\bar{Z}), \textit{ for any } q \times (p+m) \textit{ matrix } V_1\,; \tag{2.4.34}$$

$$\mathrm{rank}\begin{bmatrix} \bar{Z} + V_2 Q \\ Q \end{bmatrix} = \mathrm{rank}(\bar{Z}), \textit{ for any } (p+m) \times q \textit{ matrix } V_2\,; \tag{2.4.35}$$

$$\mathrm{rank}\begin{bmatrix} \bar{Z} \\ SQ \end{bmatrix} = \mathrm{rank}(\bar{Z}), \textit{ for any matrix } S \textit{ such that } \mathrm{rank}(SQ) = \mathrm{rank}(Q)\,; \tag{2.4.36}$$

$$Q = Q\bar{Z}^{-}\bar{Z}, \textit{ for some g-inverse } \bar{Z}^{-}\,; \tag{2.4.37}$$

$$\mathrm{rank}[\bar{Z}(I_k - Q^{-}Q)\,] = \mathrm{rank}(\bar{Z}) - \mathrm{rank}(Q), \textit{ for some g-inverse } Q^{-}\,. \tag{2.4.38}$$

The property (2.4.27), for which characterizations are given above, means that for $c_0 \in \mathrm{Im}(R)$, $Q\beta$ is uniquely determined by $Z\beta$ when $\beta \in \mathscr{L}_I(R, c_0)$. In (2.4.28), this is formulated as the existence of a function of the type:

$$Q\beta = \bar{g}(\bar{Z}\beta) = g(Z\beta, R\beta)\,, \quad \beta \in \mathbb{R}^k. \tag{2.4.39}$$

In other words, once the value of $R\beta$ is set, $Q\beta$ is uniquely by $Z\beta$. Then, the other characterizations given in Proposition **2.4.2** all apply provided Z is replaced by the enlarged matrix $\bar{Z}$. In other words, adding linear (consistent) restrictions can be viewed as equivalent to adding rows to to the Z matrix, which in turn can increase the rank of Z, hence the chance that $Z\beta$ uniquely determines $Q\beta$.

In particular, condition (2.4.33) can be rewritten as:

$$\mathrm{rank}\begin{bmatrix} Z \\ R \\ Q \end{bmatrix} = \mathrm{rank}\begin{bmatrix} Z \\ R \end{bmatrix}. \tag{2.4.40}$$

Clearly, if condition (2.4.11) holds (no restriction), then (2.4.40) must also hold irrespective of the matrix R. The condition

$$\mathrm{rank}\begin{bmatrix} R \\ Q \end{bmatrix} = \mathrm{rank}(R) \tag{2.4.41}$$

also entails (2.4.33), so it is sufficient (though not necessary) for the functional dependence property to hold under restrictions of the type (2.4.22). In this case, $Q\beta$ is completely determined by the restrictions, and the Z matrix is irrelevant.

If $R = 0$, there is no restriction on β, and we have:

$$\operatorname{rank}\begin{bmatrix} Z \\ R \\ Q \end{bmatrix} = \operatorname{rank}\begin{bmatrix} Z \\ Q \end{bmatrix}, \quad \operatorname{rank}\begin{bmatrix} Z \\ R \end{bmatrix} = \operatorname{rank}(Z), \tag{2.4.42}$$

so condition (2.4.33) entails (2.4.11). If $Q = (1, 0, \ldots, 0)'$ and $R = \begin{bmatrix} r_1 & R_2 \end{bmatrix}$ where r_1 is the first column of R, we get the condition:

$$\operatorname{rank}\begin{bmatrix} z_1 & Z_2 \\ r_1 & R_2 \\ 1 & 0' \end{bmatrix} = \operatorname{rank}\begin{bmatrix} z_1 & Z_2 \\ r_1 & R_2 \end{bmatrix}. \tag{2.4.43}$$

So, if $r_1 \neq 0$ and $R_2 = 0$, this condition is satisfied.

We will now give characterizations based on explicit linear restrictions.

Proposition 2.4.5 CONDITIONS FOR FUNCTIONAL DEPENDENCE BETWEEN SUBSPACES WITH EXPLICIT LINEAR RESTRICTIONS. *Let Z and Q be $p \times k$ and $q \times k$ real matrices. Suppose β satisfies the restriction* (2.4.23) *and let $\mathscr{L}_E(C, b_0) = \{x \in \mathbb{R}^k : x = b_0 + Cy,\ y \in \mathbb{R}^m\}$. Then the following statements are equivalent:*

for any $b_0 \in \mathbb{R}^k$, there is a function $g_{b_0} : \operatorname{Im}(Z) \mapsto \operatorname{Im}(Q)$ such that $Q\beta = g_{b_0}(Z\beta),\ \forall \beta \in \mathscr{L}_E(C, b_0)$; (2.4.44)

$$[(ZCd_1 = ZCd_2) \Rightarrow (QCd_1 = QCd_2)](\forall d_1,\ d_2 \in \mathbb{R}^m); \tag{2.4.45}$$

$$\ker(ZC) \subseteq \ker(QC); \tag{2.4.46}$$

$$\operatorname{Im}(C'Q') \subseteq \operatorname{Im}(C'Z'); \tag{2.4.47}$$

$$QC = BZC, \text{ for some matrix } B; \tag{2.4.48}$$

$$\operatorname{rank}\begin{bmatrix} ZC \\ QC \end{bmatrix} = \operatorname{rank}(ZC); \tag{2.4.49}$$

$$\operatorname{rank}\begin{bmatrix} ZC \\ QC + V_1 ZC \end{bmatrix} = \operatorname{rank}(ZC), \text{ for any } q \times p \text{ matrix } V_1; \tag{2.4.50}$$

$$\operatorname{rank}\begin{bmatrix} ZC+V_2QC \\ QC \end{bmatrix} = \operatorname{rank}(ZC), \textit{ for any } p\times q \textit{ matrix } V_2\,; \tag{2.4.51}$$

$$\operatorname{rank}\begin{bmatrix} ZC \\ SQC \end{bmatrix} = \operatorname{rank}(ZC), \textit{ for any matrix } S \textit{ such that } \operatorname{rank}(SQC)=\operatorname{rank}(QC)\,; \tag{2.4.52}$$

$$QC = QC(ZC)^{-}ZC, \textit{ for some g-inverse } (ZC)^{-}\,; \tag{2.4.53}$$

$$\operatorname{rank}(\,ZC[I_k-(QC)^{-}QC]) = \operatorname{rank}(ZC)-\operatorname{rank}(QC), \textit{ for some g-inverse } (QC)^{-}\,. \tag{2.4.54}$$

As opposed to restrictions in implicit form, the above characterizations are "multiplicative" for they involve multiplying Z and Q by C, rather than extending Z. By taking $m=k$ and $C=I_k$, we see that Proposition **2.4.2** is entailed by Proposition **2.4.5**; the value of b_0 is then irrelevant. In the opposite case where $C=0$, we have $\beta=b_0$, so β is completely restricted and $g(\cdot)$ is a constant function [$g(\beta)=b_0$]. It is then easy to directly check any one of the equivalent conditions of Proposition **2.4.5**. More generally, on applying (2.4.46), we see that the functional dependence property holds whenever $QC=0$, *i.e.* if each column of C is orthogonal to the rows of Q [the spaces $\operatorname{Im}(C)$ and $\operatorname{Im}(Q')$ are orthogonal]; the structure of the Z matrix is then irrelevant.

Let us now look at the case where $Q\beta$ is scalar ($Q=l'$). The extension of the linear independence condition (2.4.21) to situations where general linear restrictions are imposed on β leads to a more complex result. In the following proposition, we provide such an extension which is applicable to $l'\beta$ under such restrictions. Here, $C_{\cdot j}$ denotes the j-th column of C (a $k\times 1$ matrix), and $C_{(\cdot j)}$ the $k\times(m-1)$ matrix obtained by suppressing the j-th column of C. Similarly, $C_{j\cdot}$ denotes the j-th row of C (a $1\times m$ matrix), and $C_{(j\cdot)}$ the $(k-1)\times m$ matrix obtained by suppressing the j-th row of C. Since functional dependence holds immediately when $C'l=0$, we focus on the case where $C'l\neq 0$, so there is at least one column $C_{\cdot j}$ such that $C'_{\cdot j}l\neq 0$.

Proposition 2.4.6 CONDITIONS FOR FUNCTIONAL DEPENDENCE OF A LINEAR SCALAR PARAMETER WITH EXPLICIT LINEAR RESTRICTIONS. *Let $l=(l_1,\ldots,l_k)'$ be a $k\times 1$ real vector. Suppose β satisfies the restriction* (2.4.23)*, with $C'l\neq 0$. Let $L=\{x: x=b_0+Cy,\ y\in\mathbb{R}^m\}$, $C_{\cdot j}$ any column of C such that $C'_{\cdot j}l\neq 0$, where $1\leq j\leq m$,*

$$N_j = I_k - \frac{1}{C'_{\cdot j}l}C_{\cdot j}l' \tag{2.4.55}$$

and $l_{(j)}$ the $(k-1)\times 1$ vector obtained by dropping the j-th component l_j from l. Then the

following conditions are equivalent:

$$\textit{there is a function } g : \mathscr{L} \mapsto \mathrm{Im}(Q) \textit{ such that } l'\beta = g(Z\beta),\ \forall \beta \in \mathscr{L}\,; \tag{2.4.56}$$

$$C'l = C'_{j\cdot} l_j + C'_{(j\cdot)} l_{(j)} \in \mathrm{Im}\big(C'Z'\big); \tag{2.4.57}$$

$$\mathrm{rank}\,[ZC] = 1 + \mathrm{rank}\left[ZN_jC_{(\cdot j)}\right]\,; \tag{2.4.58}$$

$$ZC_{\cdot j} \notin \mathrm{Im}\left[ZN_jC_{(\cdot j)}\right]. \tag{2.4.59}$$

Note the matrix N_j is an idempotent matrix such that $N_jC_{\cdot j} = 0$, *i.e.* a projector onto the space orthogonal to $C_{\cdot j}$. For $C = I_k$ (no restriction) and $l = (1, 0, \ldots, 0)'$, conditions (2.4.20) and (2.4.21) follow from the above proposition: simply set $j = 1$, so

$$C_{\cdot 1} = (1, 0, \ldots, 0)',\ C_{(\cdot 1)} = \begin{bmatrix} 0' \\ I_{k-1} \end{bmatrix},\ N_1 = I_k - C_{\cdot 1}l' = \begin{bmatrix} 0 & 0' \\ 0 & I_{k-1} \end{bmatrix}, \tag{2.4.60}$$

$$ZC_{\cdot 1} = [z_1 \quad Z_2]\,C_{\cdot 1} = z_1, \quad ZN_1C_{(\cdot 1)} = [z_1 \quad Z_2] \begin{bmatrix} 0' \\ I_{k-1} \end{bmatrix} = Z_2\,. \tag{2.4.61}$$

When $C_{\cdot 1} = (1, 0, \ldots, 0)'$ and $C_{(\cdot 1)} = 0$, all the components of β are fixed except β_1, and condition (2.4.59) means that $z_1 \neq 0$, without further conditions on Z. Of course, Proposition **2.4.6** covers a wide array of more complex cases.

As mentioned earlier, implicit linear restrictions can be put in explicit form, and vice-versa. Consequently, in view of (2.4.24), the conditions of Proposition **2.4.5** can be applied to implicit linear restrictions by taking $C = I_k - R^-R$. Similarly, in view of (2.4.25) the conditions of Proposition **2.4.4** can be applied to explicit restrictions with $R = I_k - CC^-$.

For example, for implicit linear restrictions, condition (2.4.49) yields:

$$\mathrm{rank} \begin{bmatrix} Z(I_k - R^-R) \\ Q(I_k - R^-R) \end{bmatrix} = \mathrm{rank}[Z(I_k - R^-R)]\,. \tag{2.4.62}$$

Similarly, for explicit linear restrictions, condition (2.4.49) gives:

$$\mathrm{rank} \begin{bmatrix} Z \\ I_k - CC^- \\ Q \end{bmatrix} = \mathrm{rank} \begin{bmatrix} Z \\ I_k - CC^- \end{bmatrix}. \tag{2.4.63}$$

2.5. Identification of linear parameters in linear regressions with non-scalar covariance

In this section, we study the identification of linear parameters in the context of a linear regression model. More precisely, we consider the model

$$y = X\beta + u \tag{2.5.1}$$

where y is an $n \times 1$ vector of observations, X is an $n \times k$ matrix of rank r with $0 \leq r \leq k$, u is an $n \times 1$ vector of random disturbances whose distribution depends on X, β and a (possibly infinite dimensional) parameter $\lambda \in \Lambda$, where

$$\theta \in \Theta_0 \subseteq \mathbb{R}^k \times \Lambda \quad \text{where } \theta := \begin{pmatrix} \beta \\ \lambda \end{pmatrix}. \tag{2.5.2}$$

Further, the distribution of u – which we denote $\mathsf{P}_u(\theta, X)$ – has finite second moments (for X given) with

$$\mathsf{E}_\theta(u\,|\,X) = 0\,, \quad \mathsf{E}_\theta(uu'|\,X) = \Sigma(\theta\,|\,X)\,, \tag{2.5.3}$$

where $\mathsf{E}_\theta(\cdot\,|\,X)$ and $\mathsf{V}_\theta(\cdot\,|\,X)$ represent the expected value and variance operator, under the distribution $\mathsf{P}_u(\theta, X)$. Thus,

$$\mathsf{E}_\theta(y\,|\,X) = X\beta\,, \quad \mathsf{V}_\theta(y|\,X) = \Sigma(\theta\,|\,X)\,. \tag{2.5.4}$$

Assumption (2.5.3) allows for general forms of heteroskedasticity and dependence between model disturbances, as well as dependence of $\mathsf{P}_u(\theta, X)$ on β and X. Further, neither β nor λ need be identifiable, but $X\beta$ and $\Sigma(\theta\,|\,X)$ are identifiable because they are moments of observed variables. Note also the parameters β and λ may be restricted in arbitrary ways, through the condition $\theta \in \Theta_0$ in (2.5.2), and $\Sigma(\theta\,|\,X)$ can be singular. A standard special case of the above model is the one where

$$u \sim \mathrm{N}[0,\, \sigma^2 I_n] \tag{2.5.5}$$

i.e. $\lambda = \sigma^2$ and $\Lambda = [0, \infty)$.

We wish to study the identification of $Q\beta$ in the context of such models. For future reference, we define the functions $\beta(\theta)$ and $\lambda(\theta)$ as the corresponding components of θ,

with $\beta(\Theta_0)$ and $\lambda(\Theta_0)$ the images of Θ_0 induced by these functions:

$$\beta(\theta) = \beta\,,\ \lambda(\theta) = \lambda\,, \tag{2.5.6}$$

$$\beta(\Theta_0) = \{x \in \mathbb{R}^k : x = \beta(\theta)\,,\ \theta \in \Theta_0\}, \tag{2.5.7}$$

$$\lambda(\Theta_0) = \{z \in \Lambda : z = \lambda(\theta)\,,\ \theta \in \Theta_0\}\,. \tag{2.5.8}$$

Note $\mathsf{P}_u(\theta, X)$ may be viewed as a parameter (a probability measure determined by θ and X) in a space of functions. If we set $\lambda(\theta) = \mathsf{P}_u(\theta, X)$, the error distribution depends on θ only through $\lambda(\theta)$. In such a case, $\lambda(\theta)$ may be viewed a parameter which takes its values in space of functions (the probability measures which represent the allowed error distributions).

2.5.1. Identification of linear parameters in linear regression

The usual estimability condition for $Q\beta$ is

$$\mathrm{Im}(Q') \subseteq \mathrm{Im}(X')\,. \tag{2.5.9}$$

Consequently, each one of the conditions (2.4.8) - (2.4.16) with $Z = X$ [in Proposition **2.4.2**] is sufficient to ensure the estimability of $Q\beta$. This also entails the same conditions are sufficient for identifiability of $Q\beta$.

It is easy to see that these conditions are not necessary for identifiability. This comes from the fact that $\mathsf{V}(y|X)$ may depend on β, and thus provide information on $Q\beta$ even when $\mathsf{E}(y|X) = X\beta$ does not. The following example provides a simple example of this possibility.

Example 2.5.1 Consider a linear model with a univariate regressor

$$y_i = x_i\beta + u_i,\ i = 1, 2, \ldots, n, \tag{2.5.10}$$

where x_i is a scalar fixed variable, n is an even integer, and $u_1,\ u_2, \ldots,\ u_n$ are random disturbances such that

$$u_{2i} = u_{2i-1}\beta + \varepsilon_{2i},\ i = 1, 2, \ldots, n/2. \tag{2.5.11}$$

and $\varepsilon_1,\ \varepsilon_2, \ldots,\ \varepsilon_n$ are i.i.d $N[0, \sigma^2]$, with $\sigma^2 > 0$. If $x_i = 0$, for $i = 1, 2, \ldots, n$, we have

$X\beta = 0$, so β is not identifiable from $X\beta$. However, when $X\beta = 0$, we also have:

$$y_{2i} = y_{2i-1}\beta + \varepsilon_{2i},\ i = 1, 2, \ldots, n/2. \tag{2.5.12}$$

Using the regression (2.5.12), the ordinary least squares estimator of β is unbiased, and β is thus identifiable.

We will give a general condition under which restrictions on X and Q are both necessary and sufficient for $Q\beta$ to be $X\beta$-identifiable. For this, we consider the following separability assumption.

Assumption 2.5.2 NUISANCE PARAMETER SEPARABILITY. *Let Q be an $m \times k$ fixed matrix. In the context of the model described by* (2.5.1) - (2.5.3)*, we have the following property: for any $\theta \in \Theta_0$, we can find $\theta^* \in \Theta_0$ such that*

$$Q\beta(\theta^*) = Q\beta(\theta), \quad X\beta(\theta^*) = X\beta(\theta) \quad \textit{and} \quad \mathsf{P}_u(\theta^*, X) = \mathsf{P}_u(\theta, X). \tag{2.5.13}$$

It is easy to see that the above assumption holds when β is not restricted by λ and the distribution of u does not depend on β, *i.e.* if we have the following two properties:

$$[\beta(\theta_1) = \beta(\theta_2)] \Rightarrow [\mathsf{E}_{\theta_1}(y|X) = \mathsf{E}_{\theta_2}(y|X)],\ \forall \theta_1, \theta_2 \in \Theta_0, \tag{2.5.14}$$

$$[\lambda(\theta_1) = \lambda(\theta_2)] \Rightarrow [\mathsf{P}_u(\theta_1, X) = \mathsf{P}_u(\theta_2, X)],\ \forall \theta_1, \theta_2 \in \Theta_0. \tag{2.5.15}$$

These conditions may be interpreted in terms of "parameter sufficiency" (on Θ_0): according to (2.5.14), β is "sufficient" for $\mathsf{E}_\theta(y|X)$, while (2.5.15) means $\lambda(\theta)$ is sufficient for $\mathsf{P}_u(\theta, X)$. Of course, these conditions imply Assumption **2.5.2**, but they are not necessary for it. The results presented below do not require (2.5.14) - (2.5.15). It is also of interest to note that Assumption **2.5.2** allows for both linear and nonlinear restrictions on β [through the set Θ_0].

We first state a generic necessary and sufficient condition for the identification of $Q\beta$.

Theorem 2.5.3 IDENTIFIABILITY OF LINEAR REGRESSION COEFFICIENTS UNDER ERROR DISTRIBUTION SEPARABILITY. *Let Q be an $m \times k$ fixed matrix. In the context of the model described by* (2.5.1) - (2.5.3)*, suppose Assumption* **2.5.2** *holds. Then, $Q\beta$ is $(X\beta)$-identifiable on Θ_0 if and only if*

$$(X\beta_1 = X\beta_2) \Rightarrow (Q\beta_1 = Q\beta_2),\ \forall \beta_1, \beta_2 \in \beta(\Theta_0). \tag{2.5.16}$$

This theorem allows one to focus on the relationship between $Q\beta$ and $X\beta$ in order to study the identification of $Q\beta$, namely to state characterizations of $(X\beta)$-identifiability. Of course, the sufficiency of condition (2.5.16) follows directly from Definition **2.2.2**. The main point consists in observing that condition (2.5.16) is also necessary in the linear model (2.5.1) - (2.5.3) when parameter separability holds.

We can now give necessary and sufficient conditions for identifiability of linear parameters in linear regression models with deficient rank and general error structures.

Theorem 2.5.4 CONDITIONS FOR IDENTIFICATION OF LINEAR PARAMETERS IN LINEAR REGRESSIONS. *Let Q be an $m \times k$ fixed real matrix. In the context of the model described by* $(2.5.1)$ - $(2.5.3)$*, each one of the following equivalent conditions entails that $Q\beta$ is $(X\beta)$-identifiable on Θ_0:*

$$\ker(X) \subseteq \ker(Q)\,; \tag{2.5.17}$$

$$\mathrm{Im}\left(Q'\right) \subseteq \mathrm{Im}\left(X'\right)\,; \tag{2.5.18}$$

$$Q = AX\,, \text{ for some matrix } A\,; \tag{2.5.19}$$

$$\mathrm{rank}\begin{bmatrix} X \\ Q \end{bmatrix} = \mathrm{rank}(X)\,; \tag{2.5.20}$$

$$\mathrm{rank}\begin{bmatrix} X \\ Q + V_1 X \end{bmatrix} = \mathrm{rank}(X), \text{ for any } q \times n \text{ matrix } V_1\,; \tag{2.5.21}$$

$$\mathrm{rank}\begin{bmatrix} X + V_2 Q \\ Q \end{bmatrix} = \mathrm{rank}(X), \text{ for any } n \times q \text{ matrix } V_2\,; \tag{2.5.22}$$

$$\mathrm{rank}\begin{bmatrix} X \\ SQ \end{bmatrix} = \mathrm{rank}(X)\,, \text{ for any matrix } S \text{ such that } \mathrm{rank}(SQ) = \mathrm{rank}(Q)\,; \tag{2.5.23}$$

$$\mathrm{rank}\left\{X\left(I_k - Q^- Q\right)\right\} = \mathrm{rank}(X) - \mathrm{rank}(Q)\,, \text{ for some g-inverse } Q^-\,; \tag{2.5.24}$$

$$Q = QX^- X\,, \text{ for some g-inverse } X^-\,. \tag{2.5.25}$$

If furthermore Assumption **2.5.2** *holds and $\beta(\Theta_0) = \mathbb{R}^k$, each one of the conditions* $(2.5.17)$ - $(2.5.25)$ *is necessary for $Q\beta$ to be $(X\beta)$-identifiable on Θ_0.*

Let us now consider the case where linear restrictions are imposed on β. We first consider implicit linear restrictions:

$$R\beta = c_0 \tag{2.5.26}$$

where R is an $m \times k$ matrix (with $m \geq 1$) and $c_0 \in \mathrm{Im}(R)$. We make no rank assumption on R. Using the results of sections 2.3 and 2.4, we get the following theorem on the identification

of $Q\beta$ when implicit linear restrictions are imposed.

Theorem 2.5.5 CONDITIONS FOR IDENTIFICATION OF LINEAR PARAMETERS IN LINEAR REGRESSION WITH IMPLICIT LINEAR RESTRICTIONS. *Let Q and R be $q \times k$ and $m \times k$ real matrices, respectively, $c_0 \in \mathrm{Im}(R)$, and*

$$\bar{X} = \begin{bmatrix} X \\ R \end{bmatrix}. \tag{2.5.27}$$

In the context of the model described by (2.5.1) - (2.5.3)*, suppose the implicit linear restriction* (2.5.26) *holds, and let $\bar{\Theta}_{0I}(R, c_0) = \{\theta \in \Theta_0 : R\beta(\theta) = c_0\}$, $\bar{\beta}_I(R, c_0) = \{x \in \mathbb{R}^k : x = \beta(\theta),\ \theta \in \bar{\Theta}_{0I}(R, c_0)\}$. Then, each one of the following equivalent conditions entails that $Q\beta$ is $(X\beta)$-identifiable on $\bar{\Theta}_{0I}(R, c_0)$, for any $c_0 \in \mathrm{Im}(R)$:*

$$\ker(\bar{X}) \subseteq \ker(Q); \tag{2.5.28}$$

$$\mathrm{Im}(Q') \subseteq \mathrm{Im}(\bar{X}'); \tag{2.5.29}$$

$$Q = B\bar{X}, \text{ for some matrix } B; \tag{2.5.30}$$

$$\mathrm{rank}\begin{bmatrix} \bar{X} \\ Q \end{bmatrix} = \mathrm{rank}(\bar{X}); \tag{2.5.31}$$

$$\mathrm{rank}\begin{bmatrix} \bar{X} \\ Q + V_1\bar{X} \end{bmatrix} = \mathrm{rank}(\bar{X}), \text{ for any } q \times (p+m) \text{ matrix } V_1; \tag{2.5.32}$$

$$\mathrm{rank}\begin{bmatrix} \bar{X} + V_2 Q \\ Q \end{bmatrix} = \mathrm{rank}(\bar{X}), \text{ for any } (p+m) \times q \text{ matrix } V_2; \tag{2.5.33}$$

$$\mathrm{rank}\begin{bmatrix} \bar{X} \\ SQ \end{bmatrix} = \mathrm{rank}(\bar{X}), \text{ for any matrix } S \text{ such that } \mathrm{rank}(SQ) = \mathrm{rank}(Q); \tag{2.5.34}$$

$$Q = Q\bar{X}^-\bar{X}, \text{ for some g-inverse } \bar{X}^-; \tag{2.5.35}$$

$$\mathrm{rank}[(I_k - Q^-Q)] = \mathrm{rank}(\bar{X}) - \mathrm{rank}(Q), \text{ for some g-inverse } Q^-. \tag{2.5.36}$$

If furthermore Assumption **2.5.2** *holds and $\bar{\beta}_I(R, c_0) = \{\beta \in \mathbb{R}^k : R\beta = c_0\}$, each one of the conditions* (2.5.28) - (2.5.36) *is necessary for $Q\beta$ to be $(X\beta)$-identifiable on $\bar{\Theta}_{0I}(R, c_0)$.*

In certain cases, linear restrictions in explicit form may be more convenient:

$$\beta = b_0 + Ce \text{ for some } e \in \mathbb{R}^m \tag{2.5.37}$$

where b_0 is some $k \times 1$ vector and C is some $k \times m$ real matrix. This yields the following characterizations of identifiability. for $Q\beta$.

Theorem 2.5.6 CONDITIONS FOR IDENTIFICATION OF LINEAR PARAMETERS IN LINEAR REGRESSION WITH EXPLICIT LINEAR RESTRICTIONS. *Let Q be a $q \times k$ real matrices. In the context of the model described by* (2.5.1) - (2.5.3)*, suppose the explicit linear restriction* (2.5.37) *holds, and let $\bar{\Theta}_{0E}(C, b_0) = \{\theta \in \Theta_0 : \beta(\theta) = b_0 + Ce$ for some $e \in \mathbb{R}^m\}$, $\bar{\beta}_E(C, b_0) = \{\beta \in \mathbb{R}^k : \beta = \beta(\theta),\ \theta \in \bar{\Theta}_{0E}(C, b_0)\}$. Then, each one of the following equivalent conditions entails that $Q\beta$ is $(X\beta)$-identifiable on $\bar{\Theta}_{0E}(C, b_0)$, for any $b_0 \in \mathbb{R}^k$:*

$$\ker(XC) \subseteq \ker(QC)\,; \tag{2.5.38}$$

$$\mathrm{Im}(C'Q') \subseteq \mathrm{Im}(C'X')\,; \tag{2.5.39}$$

$$QC = BXC, \textit{ for some matrix } B\,; \tag{2.5.40}$$

$$\mathrm{rank}\begin{bmatrix} XC \\ QC \end{bmatrix} = \mathrm{rank}(XC)\,; \tag{2.5.41}$$

$$\mathrm{rank}\begin{bmatrix} XC \\ QC + V_1XC \end{bmatrix} = \mathrm{rank}(XC), \textit{ for any } q \times p \textit{ matrix } V_1\,; \tag{2.5.42}$$

$$\mathrm{rank}\begin{bmatrix} XC + V_2QC \\ QC \end{bmatrix} = \mathrm{rank}(XC), \textit{ for any } p \times q \textit{ matrix } V_2\,; \tag{2.5.43}$$

$$\mathrm{rank}\begin{bmatrix} XC \\ SQC \end{bmatrix} = \mathrm{rank}(XC), \textit{ for any matrix } S \textit{ such that } \mathrm{rank}(SQC) = \mathrm{rank}(QC)\,; \tag{2.5.44}$$

$$QC = QC(XC)^-XC, \textit{ for some g-inverse } (XC)^-\,; \tag{2.5.45}$$

$$\mathrm{rank}(XC[I_k - (QC)^-QC]) = \mathrm{rank}(XC) - \mathrm{rank}(QC), \textit{ for some g-inverse } (QC)^-\,. \tag{2.5.46}$$

If furthermore Assumption **2.5.2** *holds and $\bar{\beta}_E(C, c_0) = \{\beta \in \mathbb{R}^k : \beta = b_0 + Ce$ for some $e \in \mathbb{R}^m\}$, each one of the conditions* (2.5.38) - (2.5.46) *is necessary for $Q\beta$ to be $(X\beta)$-identifiable on $\bar{\Theta}_{0E}(C, b_0)$, for any $b_0 \in \mathbb{R}^k$.*

As mentioned earlier, implicit linear restrictions can be put in explicit form, and vice-versa. Consequently, in view of (2.4.24), the conditions of Theorem **2.5.6** can be applied to implicit linear restrictions on taking $C = I_k - R^-R$. Similarly, in view of (2.4.25) the conditions of Theorem **2.5.5** can be applied to explicit restrictions with $R = I_k - CC^-$.

For example, for implicit linear restrictions, condition (2.5.31) yields:

$$\operatorname{rank}\begin{bmatrix} X(I_k - R^- R) \\ Q(I_k - R^- R) \end{bmatrix} = \operatorname{rank}[X(I_k - R^- R)]\,. \tag{2.5.47}$$

Similarly, for explicit linear restrictions, condition (2.5.41) gives:

$$\operatorname{rank}\begin{bmatrix} X \\ I_k - CC^- \\ Q \end{bmatrix} = \operatorname{rank}\begin{bmatrix} X \\ I_k - CC^- \end{bmatrix}. \tag{2.5.48}$$

2.5.2. Identification of a scalar linear parameter

We will now discuss the identification of a single component of β when no restriction is imposed on β. Without loss of generality, we can focus on the first component of β, *i.e.*, we consider $Q\beta = l'\beta = \beta_1$ where $Q = l' = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}' \end{bmatrix}$. Since $X\beta$ is identifiable, condition (2.5.20) of Theorem **2.5.4** takes the form

$$\operatorname{rank}\begin{bmatrix} X \\ l' \end{bmatrix} = \operatorname{rank}\big(X\big) \tag{2.5.49}$$

where x_1 is the first column of X, which is equivalent to

$$\operatorname{rank}\begin{bmatrix} x_1 & X_2 \\ 1 & \mathbf{0}' \end{bmatrix} = 1 + \operatorname{rank}\big(X_2\big) = \operatorname{rank}\begin{bmatrix} x_1 & X_2 \end{bmatrix}. \tag{2.5.50}$$

This means that the first column of X does not belong to the space spanned by the columns of X_2:

$$x_1 \notin \operatorname{Im}[X_2]\,. \tag{2.5.51}$$

This condition is sufficient for the identification of β_1. If, furthermore, separability between regression and covariance parameters holds [Assumption **2.5.2**], this condition is also necessary for identification of β_1. Note (2.5.50) is also the estimability condition given by Kounias and Chalikias (2008) in the standard linear regression model.

As a second special case, let us consider the identification of β_1 when linear restrictions of the form $R\beta = c_0$ are imposed on β. Setting $R = \begin{bmatrix} r_1 & R_2 \end{bmatrix}$ where r_1 is a scalar and

$$\bar{X} = \begin{bmatrix} X \\ R \end{bmatrix} = \begin{bmatrix} x_1 & X_2 \\ r_1 & R_2 \end{bmatrix}, \tag{2.5.52}$$

condition (2.5.31) of Theorem **2.5.5** yields the following sufficient condition for the identification of β_1 :

$$\text{rank}\begin{bmatrix} x_1 & X_2 \\ r_1 & R_2 \\ 1 & \mathbf{0}' \end{bmatrix} = 1 + \text{rank}\begin{bmatrix} X_2 \\ R_2 \end{bmatrix} = \text{rank}\begin{bmatrix} x_1 & X_2 \\ r_1 & R_2 \end{bmatrix}$$

This means that the first column of the augmented matrix $\bar{X}$ does not belong to the space spanned by the other columns of $\bar{X}$, *i.e.*

$$\begin{bmatrix} x_1 \\ r_1 \end{bmatrix} \notin \text{Im}\begin{pmatrix} X_2 \\ R_2 \end{pmatrix}.$$

If, furthermore, separability between regression and covariance parameters holds [Assumption **2.5.2**], this condition is also necessary for identification of β_1.

The identification conditions for general linear scalar parameters $Q\beta = l'\beta$ with general (explicit) restrictions of the form (2.4.23) follow on applying Theorem **2.5.6**. For completeness, we state it below. Proposition restrictions of the Let us now look at the case where $Q\beta$ is scalar ($Q = l'$). The extension of the linear independence condition (2.4.21) to situations where general linear restrictions are imposed on β leads to a more complex result. In the following proposition, we provide such an extension which is applicable to $l'\beta$ under such restrictions. Here, $C_{\cdot j}$ denotes the j-th column of C (a $k \times 1$ matrix), and $C_{(\cdot j)}$ the $k \times (m-1)$ matrix obtained by suppressing the j-th column of C. Similarly, $C_{j\cdot}$ denotes the j-th row of C (a $1 \times m$ matrix), and $C_{(j\cdot)}$ the $(k-1) \times m$ matrix obtained by suppressing the j-th row of C. Since functional dependence holds immediately when $C'l = 0$, we focus on the case where $C'l \neq 0$, so there is at least one column $C_{\cdot j}$ such that $C'_{\cdot j}l \neq 0$. We also denote $\mathbf{0}_k$ the zero $k \times k$ matrix.

Proposition 2.5.7 CONDITIONS FOR IDENTIFICATION OF A LINEAR SCALAR PARAMETER WITH EXPLICIT LINEAR RESTRICTIONS. *In the context of the model described by* (2.5.1) - (2.5.3)*, suppose* β *belongs to the affine subset* $L = \{x : x = b_0 + Cy,\ y \in \mathbb{R}^m\} \subseteq \mathbb{R}^k$. *Let* $\bar{\Theta}_{0E}(C, b_0) = \{\theta \in \Theta_0 : \beta(\theta) \in \mathscr{L}\}$, $\bar{\beta}_E(C, b_0) = \{\beta \in \mathbb{R}^k : \beta = \beta(\theta),\ \theta \in \bar{\Theta}_{0E}(C, b_0)\}$, $l = (l_1, \ldots, l_k)'$ *a* $k \times 1$ *real vector such that.*$C'l \neq 0$, $C_{\cdot j}$ *any column of* C *such that* $C'_{\cdot j}l \neq 0$, *where* $1 \leq j \leq m$, $C_{(\cdot j)}$ *the* $k \times (m-1)$ *matrix obtained by suppressing the* j*-th column of* C, $l_{(j)}$ *the* $(k-1) \times 1$ *vector obtained by dropping the* j*-th component* l_j *from* l, *and*

$$N_j = I_k - \frac{1}{C'_{\cdot j}l} C_{\cdot j} l'. \tag{2.5.53}$$

Then, each one of the following equivalent conditions entails that $l'\beta$ is $(X\beta)$-identifiable on $\bar{\Theta}_{0E}(C, b_0)$:

$$C'l = C'_{j\cdot} l_j + C'_{(j\cdot)} l_{(j)} \in \mathrm{Im}\left(C'X'\right); \tag{2.5.54}$$

$$\mathrm{rank}\,[XC] = 1 + \mathrm{rank}\left[XN_j C_{(\cdot j)}\right]; \tag{2.5.55}$$

$$XC_{\cdot j} \notin \mathrm{Im}\left[XN_j C_{(\cdot j)}\right]. \tag{2.5.56}$$

If furthermore Assumption **2.5.2** *holds and $\bar{\beta}_E(C, b_0) = \{\beta \in \mathbb{R}^k : \beta = b_0 + Ce$ for some $e \in \mathbb{R}^m\}$, each one of the conditions* (2.5.54) - (2.5.56) *is necessary for $Q\beta$ to be $(X\beta)$-identifiable on $\bar{\Theta}_{0E}(C, b_0)$.It will be interest to consider a few special cases.*

1. For $C = I_k$ (no restriction) and $l = (1, 0, \ldots, 0)'$, condition (2.5.51) follows from the above proposition: simply set $j = 1$, so

$$C_{\cdot 1} = (1, 0, \ldots, 0)',\ C_{(\cdot 1)} = \begin{bmatrix} 0' \\ I_{k-1} \end{bmatrix}, \tag{2.5.57}$$

$$C'_{\cdot 1} l = 1,\ C_{\cdot 1} l' = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} (1, 0, \ldots, 0) = \begin{bmatrix} 1 & 0' \\ 0 & \mathbf{0}_{k-1} \end{bmatrix},\ N_1 = I_k - C_{\cdot 1} l' = \begin{bmatrix} 0 & 0' \\ 0 & I_{k-1} \end{bmatrix}, \tag{2.5.58}$$

$$XC_{\cdot 1} = [x_1 \quad X_2]\, C_{\cdot 1} = x_1, \quad XN_1 C_{(\cdot 1)} = [x_1 \quad X_2] \begin{bmatrix} 0' \\ I_{k-1} \end{bmatrix} = X_2. \tag{2.5.59}$$

2. More generally, if $C = I_k$ and $l_1 \neq 0$ (without restrictions on $l_{(1)}$, the other elements of l), we have (2.5.57), $C'_{\cdot 1} l = l_1$ and

$$C_{\cdot 1} l' = \begin{bmatrix} l_1 & l'_{(1)} \\ 0 & \mathbf{0}_{k-1} \end{bmatrix},\ N_1 = I_k - \frac{1}{l_1} C_{\cdot 1} l' = \begin{bmatrix} 0 & -\frac{1}{l_1} l'_{(1)} \\ 0 & I_{k-1} \end{bmatrix}, \tag{2.5.60}$$

$$XC_{\cdot 1} = [x_1 \quad X_2]\, C_{\cdot 1} = x_1, \tag{2.5.61}$$

$$XN_1 C_{(\cdot 1)} = [x_1 \quad X_2] \begin{bmatrix} 0 & -l'_{(1)}/l_1 \\ 0 & I_{k-1} \end{bmatrix} = \begin{bmatrix} 0 & X_2 - \frac{1}{l_1} x_1 l'_{(1)} \end{bmatrix}. \tag{2.5.62}$$

Then, applying condition (2.5.56), we see that the condition

$$x_1 \notin \mathrm{Im}\left[X_2 - \frac{1}{l_1} x_1 l'_{(1)}\right] \tag{2.5.63}$$

entails that $l_1\beta_1$ is identifiable. Clearly this condition applies to β_1 on taking $l_1 = 1$. Again no assumption on the ranks of X_2 and $X_2 - \frac{1}{l_1}x_1 l'_{(1)}$ is needed.

3. When $C_{\cdot 1} = (1, 0, \ldots, 0)'$ and $C_{(\cdot 1)} = 0$, all the components of β are set by the restriction except for β_1, and condition (2.4.59) means that $x_1 \neq 0$, without further conditions on X.

Of course, Proposition **2.5.7** covers a wide array of more complex cases.

2.6. Canonical identifiable representations

Due to the close relationship between the features of the eigenvalues and the rank of a given matrix, under certain circumstances it will be more convenient to characterize some of the results in Sections 2.4 and 2.5 by using decomposition or factorization methods, such as the rank factorization theorem, the LU decomposition, QR factorization, reduced row echelon form and singular value decomposition (SVD) technique. Recall that the rank factorization theorem states that if Z is a $p \times k$ matrix of rank r then it can be factorized by

$$Z = AB,$$

where A and B are $p \times r$ and $r \times k$ matrices of rank r [see Rao (1973)]. Obviously, neither A nor B is unique.

For empirical study, it is convenient to characterize Z through a basis of $\mathrm{Im}(Z')$. Recall that a basis for a vector space $\mathscr{U}$ is a finite set of linearly independent vectors in $\mathscr{U}$ that spans $\mathscr{U}$ [see Harville (2008)].

Proposition 2.6.1 LINEAR FUNCTION EXPRESSION USING BASIS. *Let Z be any $p \times k$ matrix of rank r and C be a $k \times r$ matrix whose columns form a basis of* $\mathrm{Im}(Z')$. *Then $C'\beta$ can be expressed as a linear function of $Z\beta$.*

Proposition **2.6.1** is easily applicable since computer programs will help us find a basis of a subvector space spanned by the columns of a known matrix through different decomposition methods such as SVD, Cholesky factorization, Schur decomposition, LU matrix factorization and QR decomposition. The corresponding Matlab functions are "svd", "chol", "schur", "lu" and "qr" respectively. Note that some factorization methods are limited to specific type of matrices. For example, a matrix must be positive definite to be decomposed by Cholesky method and Schur decomposition can only factorize square matrices.

In contrast, SVD, LU decomposition and QR factorization methods can decompose an arbitrary matrix. Moreover, Matlab has a Toolbox which defines a function "colspace" that can return a basis for the column space of a given matrix. It also has a function "rref" produces the reduced row echelon form of a given matrix using Gauss-Jordan elimination with partial pivoting. Using the echelon form we can construct a basis of the column space of a matrix.

Since the rank theorem does not provide a unique factorization of a general matrix, it is interesting to check the uniqueness of other decomposition methods.

First we look at QR factorization of an arbitrary matrix. Recall that for any $p \times k$ matrix Z, we can always decompose it as

$$Z = QR,$$

where Q is a unitary matrix and R is an upper triangular matrix. The QR factorization is useful in that the linear relationships among the columns of the decomposed matrix are preserved exactly in the columns of R which is an upper triangular matrix. Since Q is unitary and can be expressed as a product of elementary matrices, if it is postmultiplied by R, the linear relationships among the columns of R will not change. Thus it is convenient to form a basis of the column space of a given matrix by checking the triangular matrix R. In Matlab, the function for QR factorization is "qr". If Z has full column rank k, then Z has the "thin QR" factorization as

$$Z = Q_1 R_1,$$

where Q_1 is $p \times k$ and composed of the first k orthonormal columns of Q and R_1 is an upper triangular matrix of $k \times k$. Furthermore, if the diagonal elements of R are positive, then both Q and R are unique [see Golub and Van Loan (2013, Theorem 5.2.2)].

Second, we can choose to decompose a $p \times k$ matrix Z through LU decomposition as follows

$$Z = LU,$$

where L is a lower triangular matrix and U is an upper triangular matrix. The Matlab function for LU decomposition is "lu". Using Matlab toolbox, we can choose L as a unit lower triangular matrix and the original decomposed matrix Z premultiplied by a permutation matrix P. Then the column relationships of upper triangular matrix U will be the same as those of the original matrix Z. Thus the LU method is similar to QR in that we can construct a basis by checking the column linear independency of U. The problem with LU decomposition is that it does not always exist. The sufficient condition for the existence of an LU decomposition of Z is that all the leading principal minors of Z are nonzero [see Meyer (2000)]. Again if Z is nonsingular and the LU decomposition exists, then the LU

decomposition is unique [see Golub and Van Loan (2013, Theorem 3.2.1)].

Third, we consider the SVD tool. For convenience, we restate the following singular value decomposition theorem. Let Z be a $p \times k$ real matrix of rank r. Also let M and N be two $p \times p$ and $k \times k$ orthogonal matrices containing the orthonormal eigenvectors of ZZ' and $Z'Z$ respectively. Let D be a $r \times r$ nonsingular diagonal matrix with its diagonal elements being the square roots of the eigenvalues of ZZ' or $Z'Z$. Suppose we can partition M and N as $M = \begin{bmatrix} M_1 & M_2 \end{bmatrix}$ and $N = \begin{bmatrix} N_1 & N_2 \end{bmatrix}$, where M_1 is $p \times r$ and N_1 is $k \times r$. Then

$$\mathrm{Im}\big(Z'\big) = \mathrm{Im}(N_1) = \ker\big(N_2{}'\big) \tag{2.6.64}$$

and

$$\ker(Z) = \ker\big(N_1{}'\big) = \mathrm{Im}(N_2). \tag{2.6.65}$$

Thus SVD provides an alternative way of thinking about the structure of the matrix Z and therefore extending the results of Proposition **2.4.2** by exploiting the SVD [Mandel (1982), Eubank and Webster (1985), Nelder (1985)]. For instance, instead of checking whether or not the kernel space of Z is a subspace of that of Q in (4.4.6), we can just examine whether the kernel space of N_1' can be spanned by the columns of Q. Additionally, if more than one of the singular values of Z is zero, Z will have a deficient rank, which will lead to the multicollinearity issue in the regression models. Frequently, another related issue may occur when some of the singular values of Z are very close to zero despite Z still has a full column rank. That fact that determinant of Z, $Z'Z$ or ZZ' is close to zero means that there exists almost a perfect linear relationship among the columns of Z, which can lead to many serious inference problems in statistical models. However, some linear combinations of parameters may still be identified in both cases and this is exactly what this paper concentrates on in the following sections. Furthermore, SVD provides a convenient way to find the basis of a linear space. Specifically, using SVD we know that a $p \times k$ matrix Z of rank r can be decomposed as

$$Z' = N_1 D M_1{}'.$$

It follows from (2.6.64) that the columns of N_1 form a basis of $\mathrm{Im}\big(Z'\big)$. Then we conclude that $N_1{}'\beta$ can be written as a linear function of $Z\beta$ from Proposition **2.6.1**. Note that although the singular values of Z are unique, the orthogonal matrices U and V are not. However, when the singular values are distinct, the columns of U and V are unique up to a sign.

Last, we can also use reduced row echelon form to find the basis of a linear space. Since conducting only elementary operations on the rows of Z' does not change the linear relationship among the columns of Z', the basis of $\mathrm{Im}\big(Z'\big)$ can be easily found. Besides, the reduced row echelon form is unique [see Meyer (2000)].

2.7. Partially linear models

The application of the identification conditions of linear parameters established in Section 2.2 to the linear regression model is just one example. In fact, these necessary and sufficient identification conditions can also be applied to a much wider class of statistical models named the partially linear model. In this section, we first analyze the identification conditions for different types of partially linear models that are commonly used in statistics. As shown hereinafter almost all statistical models such as logit and probit models, discrete choice models, quantile regression, single index models, etc., can be generalized into a framework such that either Proposition **2.3.2** or Theorem **2.3.4** is applicable. We also discuss the estimability issue the relationship with identifiability in partially linear models.

Using the same notations of Section 2.5, we hereinafter assume no additional restrictions on parameters for simplicity and all the results we derive can be easily extended to the case where linear restrictions are imposed on the linear parameters of interest. As in many classical textbooks, for instance, McCullagh and Nelder (1989), we start with a single random variable in partially linear models and denote the mean of such a univariate response y_i as μ_i, $i = 1,2,\ldots,n$, and the vector $\mathbf{x}_i'$ as the i^{th} row of X. Obviously, this is just a simplification and the rationale of the model set up of partially linear models will be the same for the multivariate case, as shown hereinafter when we demonstrate the application of the theoretical results in Section 2.2 by establishing the mappings from the fundamental parameters into the parameters of partially linear models.

2.7.1. Generalized linear models

Nelder and Wedderburn (1972) propose the generalized linear model as

$$g\big(\mathbb{E}(y_i)\big) = \mathbf{x}_i'\beta,\ i = 1,2,\ldots,n, \tag{2.7.1}$$

where the function g is differentiable and monotone in the expectation of y_i and is defined as the link function. They state that all the y_i's must belong to an exponential family

distribution in the form of

$$f(y_i;\rho_i,\phi_i)=\exp\left(\frac{y_i\rho_i-b(\rho_i)}{a(\phi_i)}\right)c(y_i,\phi_i),\text{ where }\ \phi_i\text{ is the dispersion parameter.}\quad(2.7.2)$$

The dispersion parameter ϕ_i is independent of ρ_i and thus does not depend on β. Note that ϕ_i can be either fixed or unknown. If ρ_i is the only parameter of interest, ϕ_i will represent the nuisance parameter. However, our paper distinguishes from the classical literature on the generalized linear model in two major aspects. First, we do not assume the response variable must follow a distribution that is a member of an exponential family distribution or an exponential dispersion family distribution. Rather, we only require the existence of a link function that connects the mean of y_i and the linear predictor $\mathbf{x}_i'\beta$. That is to say, only the most general functional form $g\big(\mathbb{E}(y_i)\big)=\mathbf{x}_i'\beta$ is needed in our paper. Therefore the generalized linear model can be trivially treated as a special case where such an exponential distribution assumption is indispensable. Second, the link function does not have to be either differentiable or monotone. Both differentiability and monotonicity of the link function will have no influence on the identification of linear parameters. Again the existence of the link function will suffice in this paper. Conventionally, we call ρ_i the canonical parameter which is a function of μ_i and thus the above distribution is in canonical form. If we choose $\rho_i=\mathbf{x}_i'\beta$, the link function $g(\mu_i)$ becomes the canonical link function. We can also choose non-canonical link functions but this will complicate the calculations. Obviously, if the link function has the identity form, *i.e.*, $g(\mu_i)=\mu_i$, we obtain the general linear model $\mathbb{E}(y_i)=\mathbf{x}_i'\beta$. Thus the classical linear model is just a special case of the generalized linear model. To illustrate the rationale of the generalized linear models, let's take the normal distribution as an example

$$f(y_i;\rho_i,\phi_i)=\frac{1}{\sqrt{2\pi{\sigma_i}^2}}\exp\left(-\frac{(y_i-\mu_i)^2}{2{\sigma_i}^2}\right)\quad(2.7.3)$$

$$=\exp\left(\frac{y_i\mu_i-{\mu_i}^2/2}{{\sigma_i}^2}\right)\exp\left(-\frac{y_i^2}{2{\sigma_i}^2}-\frac{\log(2\pi{\sigma_i}^2)}{2}\right).\quad(2.7.4)$$

Then the normal distribution is in the canonical form and

$$\rho_i=\mu_i,\ a(\phi_i)=\phi_i={\sigma_i}^2,\ b(\rho_i)={\mu_i}^2/2.$$

It is straightforward that the canonical link function for the normal distribution is indeed in the identity form.

Note that as in the case of the linear regression model in Section 2.5, we emphasize

herein that if the probability distribution of the observations depends on nuisance parameters and yet we can separate the parameters of interest from the nuisance parameters after the underlying probability distribution is reparameterized, the identification condition in Proposition **2.3.2** will not represent a necessary condition anymore. However, the identification condition in Theorem **2.3.4** is still both necessary and sufficient and therefore applicable in this situation.

Let's demonstrate how Proposition **2.3.2** and Theorem **2.3.4** can be applied in the framework of the generalized linear models by relating the parameters of probability distribution in the form of (2.7.2) to the parameters of the general parametric functions in Section 2.2. Let the random variables $y_1, y_2, \ldots, y_n$ be a random sample of size n. Then equation (2.7.1) can be expressed as

$$g\big(\mathbb{E}(\mathbf{y})\big) = g(\mu) = X\beta, \tag{2.7.5}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}.$$

Accordingly, the likelihood function of $\mathbf{y}$ is

$$L(\mathbf{y}; \rho, \phi) = \prod_{i=1}^{n} \exp\left(\frac{y_i\rho_i - b(\rho_i)}{a(\phi_i)}\right) c(y_i, \phi_i), \tag{2.7.6}$$

where

$$\rho = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{bmatrix}, \phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix}.$$

Given the canonical link, (2.7.6) can be rewritten as

$$L(\mathbf{y}; \theta^*) = L(\mathbf{y}; \rho, \phi) = L(\mathbf{y}; X\beta, \phi) = \prod_{i=1}^{n} \exp\left(\frac{y_i \cdot \mathbf{x}_i'\beta - b(\mathbf{x}_i'\beta)}{a(\phi_i)}\right) c(y_i, \phi_i), \tag{2.7.7}$$

where

$$\theta^* = \begin{bmatrix} \rho \\ \phi \end{bmatrix} = \begin{bmatrix} X\beta \\ \phi \end{bmatrix}.$$

Furthermore, (2.7.7) can be reparameterized as

$$L(\mathbf{y};\theta^*) = L(\mathbf{y};\rho,\phi) = L(\mathbf{y};\gamma_1(\theta),\gamma_2(\theta)) = \bar{L}(\mathbf{y};\theta) = \bar{L}(\mathbf{y};\beta,\phi), \tag{2.7.8}$$

where

$$\theta = \begin{bmatrix} \beta \\ \phi \end{bmatrix},\ \beta(\theta) = Q\beta,\ \gamma_1(\theta) = X\beta,\ \gamma_2(\theta) = \phi.$$

Note that θ, $\beta(\theta)$, $\gamma_1(\theta)$ and $\gamma_2(\theta)$ stand for the fundamental parameter, the linear combination of parameters that is of interest, the identifiable parameter and the nuisance parameter respectively in Section 2.2 whereas β and ϕ are the parameter of interest and the nuisance parameter defined in the generalized linear model. In addition, suppose the mean μ of $\mathbf{y}$ exists. Then the linear predictor $X\beta$, which is a function of μ through the link function $g(\mu)$, is identifiable. If the dispersion parameter ϕ is known, we can apply directly Proposition **2.3.2** since the distribution does not contain any nuisance parameter. On the other hand, if any element of the nuisance parameter ϕ is unknown, Proposition **2.3.2** is not applicable. Nevertheless, the canonical links for the exponential distribution family lead to a complete separation of the nuisance parameter ϕ from the parameter of interest β. That is to say, Assumption **2.3.3** is satisfied. Consequently any linear combination of parameters $Q\beta$ which is a function of $X\beta$ is also identifiable. Hence Theorem **2.3.4** applies. Let's take the normal distribution (2.7.3) as an example. The parameter of interest and the unknown nuisance parameter are

$$\rho = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix},\ \phi = \begin{bmatrix} {\sigma_1}^2 \\ {\sigma_2}^2 \\ \vdots \\ {\sigma_n}^2 \end{bmatrix},$$

which implies that the identification condition in Proposition **2.3.2** is sufficient but not necessary. However, under the identity canonical link ϕ is independent of ρ. That is to say, ϕ can be completely separated from the parameter of interest β. Therefore, we can establish the necessary and sufficient condition for identification of a linear combination of parameters $Q\beta$ by Theorem **2.3.4**.

Almost all relevant literatures which study the generalized linear model we have found so far have mainly concentrated on the estimability of the parameters and assume the design matrix has full column rank. In contrast, we impose no restrictions on the rank of the design matrix and analyze the identification condition for an arbitrary linear combination of parameters. Thus it may well be the case that we cannot identify the entire parameter vector β in the generalized linear models but only a subvector or some linear combination of

components of β. The following model types are referred to Dobson and Barnett (2008), Christensen (1997), McCullagh and Nelder (1989), Hosmer and Lemeshow (2000), Cox and Snell (1989), Davidson and MacKinnon (2004), Gouriéroux (2000), Agresti (1984), Collett (2003*a*), Collett (2003*b*) and Wood (2006). We hereby list all major types of generalized linear models and apply the theoretical identification results to each of these models.

First, we focus on log-linear model and Poisson regression. The canonical link function of the Poisson distribution is the logarithmic function and we can write the generalized linear model as

$$\mathbb{E}(y_i) = \mu_i = \mathrm{e}^{\mathbf{x}_i'\beta},\ i = 1, 2, \ldots, n. \tag{2.7.9}$$

Clearly the linear predictor $\mathbf{x}_i'\beta$ is not the mean or variance of the observation y_i. Rather it is a function of the mean or variance. Since the dispersion parameter for the Poisson distribution equals 1, it follows easily that $\mathbf{x}_i'\beta$ is identifiable by Definition **2.2.2** and Proposition **2.3.2**.

Second, we look at some binary response models, such as Probit and logit models. The canonical link function can be chosen to be the logistic function and the corresponding generalized linear model is the logit model

$$\mathbb{E}(y_i) = \mu_i = \frac{e^{\mathbf{x}_i'\beta}}{1 + e^{\mathbf{x}_i'\beta}},\ i = 1, 2, \ldots, n. \tag{2.7.10}$$

On the other hand, if we choose the inverse of standard normal probability function as the link function between the mean μ_i and the linear predictor $\mathbf{x}_i'\beta$, we have the probit model. If the binomial distribution involves nuisance parameters, the necessary and sufficient condition for the identification of $X\beta$ applies from Theorem **2.3.4** since the canonical link will satisfy Assumption **2.3.3**. Otherwise, Proposition **2.3.2** is applicable.

Third, we check nominal and ordered logit models. Although the Multinomial distribution usually does not belong to the exponential family of distributions, if it is assumed that the total number of trials follows a Poisson distribution we can rewrite the density of a Multinomial distribution as the joint density of the independent observations, each of which follows a Poisson distribution. As a consequence, the generalized linear model can be used to deal with the Multinomial data and either Proposition **2.3.2** or Theorem **2.3.4** will apply.

Fourth, we study survival time and duration models. One of the most common ways of modeling the survival data is to assume that y_i follows an exponential distribution and the canonical link function under the exponential distribution takes the reciprocal form. Since the dispersion parameter of the exponential distribution is a constant, Proposition **2.3.2** ap-

plies. An immediate example that deals with the censored or truncated data in survival analysis is the Tobit model. According to Aitkin and Clayton (1980), the censoring data can be considered as a random variable following a Poisson distribution. Using the natural log link function as the canonical link, if the first moment of the observations exists, Proposition **2.3.2** is applicable.

Since the generalized linear model method deals with most of the special distributions in statistics, such as the Poisson distribution, the binomial distribution, the negative binomial distribution, the Gaussian distribution, the inverse Gaussian distribution, the gamma distribution inclusive of the chi squared distribution and the exponential distribution as special cases, all of which can be written in the form of (2.7.2), we can obtain the necessary and sufficient condition for identification by either Proposition **2.3.2** if the distribution involves only a single parameter or Theorem **2.3.4** if the distribution contains nuisance parameters. Furthermore, this technique can be readily extended to other types of partially linear models as shown hereinafter. Hence, our identification conditions established in Section 2.5, especially Theorem **2.3.4**, are quite general and powerful.

2.7.2. Generalized linear mixed models

All the above generalized linear models assume that the response variable y_i is drawn independently. However, it is also common to measure a given subject for repeated times or sometimes we need the measurements of subjects that are correlated. Such longitudinal data will often generate dependence among the response variables. The longitudinal data analysis so far only focused on the estimation of parameters, see Diggle et al. (2002), Jiang (2007) and Dobson and Barnett (2008).In contrast, we hereby demonstrate that our identification condition can also apply to the generalized linear mixed model which is an extension to the generalized linear model. The generalized linear mixed model is constructed as

$$g\big(\mathbb{E}(y_i|\alpha)\big) = \mathbf{x}_i'\beta + \mathbf{z}_i'\alpha,\ i = 1,2,\ldots,n, \tag{2.7.11}$$

where β are the fixed effects, α are the random effects which are usually assumed to follow a distribution of an exponential family. Similar to the generalized linear model, g is the link function and our paper again does not need the restriction that the conditional distribution of y_i given α belongs to an exponential family distribution nor does it require that the distribution of the random coefficients α belong to the exponential family distribution. If we denote $\bar{g}$ as the link function and rewrite the generalized linear mixed model in the matrix form

$$\bar{g}\big(\mathbb{E}(\mathbf{y}|\alpha)\big) = X\beta + Z\alpha,$$

where Z is $n \times l$ and α is $l \times 1$, then $X\beta + Z\alpha$ is identifiable by Definition **2.2.2**. Furthermore, any arbitrary linear combination of parameter $Q\eta$, where $\eta = \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$ can be identified if it can rewritten as a function of $X\beta + Z\alpha$ and this condition is both necessary and sufficient by Proposition **2.3.2** in the absence of nuisance parameters. On the other hand, if the probability distribution contains nuisance parameters, then Proposition **2.3.2** is no loner applicable. Nonetheless, if the nuisance parameters are independent of the parameters of interest or can be separated from the latter after reparameterization of the probability distribution, the necessary and sufficient condition for identification in Theorem **2.3.4** is still valid.

Denote $T = \begin{bmatrix} X & Z \end{bmatrix}$ and suppose $T\eta = X\beta + Z\alpha$ is identifiable. According to Proposition **2.4.2** it is trivial to have the necessary and sufficient identification conditions of any arbitrary linear combination of parameters $Q\eta$.

2.7.3. Quantile regression

Quantile regression (see Koenker and Bassett (1978)) can provide useful information about the distribution of the response variables besides the classical least-squares regression, especially when the existence of significant statistical dispersion such as influential outliers makes the statistical inference less robust. Following the steps by Koenker (2005), the problem of finding the τ^{th} quantile is derived from solving the minimization problem where we intend to minimize the expected loss function $\rho_\tau(u) = u(\tau - \mathrm{I}(u < 0))$. The conditional quantile function can be specified as

$$Q_{y_i|\mathbf{x}_i}(\tau) = \mathbf{x}_i'\beta(\tau),\ i = 1, 2, \ldots, n, \text{ where } \beta(\tau) \text{ solves } \min_\beta \mathbb{E}[\rho_\tau(y_i - \mathbf{x}_i'\beta)]. \qquad (2.7.12)$$

Thus, $\mathbf{x}_i'\beta$ is the τ^{th} quantile of the distribution of the response variable y_i given $\mathbf{x}_i$. Again, Proposition **2.3.2** provides the necessary and sufficient condition for the identification of $Q\beta$ when the distribution of the observations does not depend on nuisance parameters. If the nuisance parameters are in presence and separable, we are guaranteed the necessary and sufficient condition for identification through Theorem **2.3.4**.

As for quantile regression models, the necessary and sufficient conditions for identification will be the same as those of generalized linear models because we can treat the conditional quantile function as the link function.

2.7.4. Single index models

One semiparametric model that has been widely used in econometrics is the so called single index model; see Robinson (1988), Klein and Spady (1993), Ichimura (1993) and Li and Racine (2007). The single index model resembles the generalized linear model in that it also involves the linear form $X\beta$ in the conditional expectation of the response variable. However, it differs from all the above mentioned generalized linear models because the link function is now unspecified. The single index model is given by

$$\mathbf{E}(y_i) = h\big(\mathbf{x}_i'\beta\big),\ i = 1,2,\ldots,n, \tag{2.7.13}$$

where the link function h in unknown. Li and Racine (2007) list three requirements for the identification of β and h function. First, $\mathbf{x}_i$ should not include an intercept term and must contain one continuous variable. Second, h function is differential and is not a constant function. Third, changing the values of the discrete variables in $\mathbf{x}_i$ will not lead to disjoint subsets of the support of $\mathbf{x}_i'\beta$. Suppose the link function is one-to-one. Then $Q\beta$ is identifiable if we apply Proposition **2.3.2** or Theorem **2.3.4**. Indeed, for the identification purpose, we could regard the single index model

$$\mathbf{E}(\mathbf{y}) = h(X\beta)$$

as the inverse of the link function g in generalized linear models. Thus the necessary and sufficient identification conditions can be easily established.

2.8. Examples

We look at an example that demonstrates the application of the identification conditions established in previous sections.

2.8.1. Linear regression

The following data is about monthly salaries and employment characteristics of 49 employees in a given company [see Ramanathan (2002)]:
Sample size=49; Wage=dollars per month; Gender=1 for a male and 0 for a female; Educ=years of education beyond eighth grade; Exper=number of years at the company; Age=age of employee. Researchers want to know if there exists a gender discrimination in

Wage	Gender	Educ	Exper	Age
1345	0	6	2	38
2435	1	4	18	52
1715	1	6	4	45
1461	1	6	4	58
1639	1	9	3	30
1345	0	5	8	43
1602	0	7	6	30
1144	0	4	3	33
1566	1	6	23	51
1496	1	4	15	37
1234	0	4	9	45
⋮	⋮	⋮	⋮	⋮
1839	1	4	13	32
1288	1	4	9	58
1288	0	6	4	29

terms of salaries among this group of employees. A usual specification is as follows:

$$Wage = \beta_0 + \beta_1 Gender + \beta_2 Educ + \beta_3 Exper + \beta_4 Age + u.$$

To avoid the dummy variable trap, we must satisfy the rule that the number of dummy variables is always one less than the number of categories. Otherwise, if the model includes both genders

$$Wage = \beta_0 + \beta_1 Male + \beta_2 Female + \beta_3 Educ + \beta_4 Exper + \beta_5 Age + u, \tag{2.8.14}$$

there exists exact multicollinearity among the columns of the design matrix X which is deficient in rank, where

$$X = \begin{bmatrix} 1 & 0 & 1 & 6 & 2 & 38 \\ 1 & 1 & 0 & 4 & 18 & 52 \\ 1 & 1 & 0 & 6 & 4 & 45 \\ 1 & 1 & 0 & 6 & 4 & 58 \\ 1 & 1 & 0 & 9 & 3 & 30 \\ 1 & 0 & 1 & 5 & 8 & 43 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 4 & 9 & 58 \\ 1 & 0 & 1 & 6 & 4 & 29 \\ & & & & & \end{bmatrix}.$$

Now we want to show the dummy variable trap issue is in nature an identification failure. We cannot identify all the parameters in (2.8.14) because the parameters of interest do not satisfy the necessary and sufficient conditions for identification in Theorem **2.5.4**.

Specifically, choose $Q_1 = I_6$ and denote $\beta = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \beta_5 \end{bmatrix}'$ so that the parameters of interest β in (2.8.14) can be written as $Q_1\beta$. Clearly

$$\mathbb{E}[\text{Wage}|X] = X\beta$$

which is identifiable. It is also easily seen that $\text{rank}(X) = 5$ due to the fact that Male + Female = 1. However,

$$\text{rank}\begin{bmatrix} X \\ Q_1 \end{bmatrix} = \text{rank}\begin{bmatrix} X \\ I_6 \end{bmatrix} = 6 \neq \text{rank}(X).$$

Thus identification condition (2.5.20) is not satisfied and $Q_1\beta$ which is β in the setup is not identifiable. Although we do not have full identification of β, it is possible to still identify part of β, say $\begin{bmatrix} \beta_3 & \beta_4 & \beta_5 \end{bmatrix}'$. If we set

$$Q_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ & & & & & \end{bmatrix},$$

$Q_2\beta = \begin{bmatrix} \beta_3 & \beta_4 & \beta_5 \end{bmatrix}'$. Plug both Q_2 and X in (2.5.20) and we have

$$\text{rank}\begin{bmatrix} X \\ Q_2 \end{bmatrix} = \text{rank}(X) = 5.$$

Next if we substitute "Male+Female" for the constant term in (2.8.14), we can rewrite the model as follows

$$Wage = \gamma_0 + \gamma_1 Male + \gamma_2 Female + \gamma_3 Educ + \gamma_4 Exper + \gamma_5 Age + u, \tag{2.8.15}$$

where

$$\gamma_0 = 0,\ \gamma_1 = \beta_0 + \beta_1,\ \gamma_2 = \beta_0 + \beta_2,\ \gamma_3 = \beta_3,\ \gamma_4 = \beta_4,\ \gamma_5 = \beta_5.$$

The transformation of (2.8.14) without a constant term is known to be estimable so that all

the parameters γ_i, $i = 0, 1, 2, \ldots, 5$ are identifiable. In fact we can construct a matrix Q_2

$$Q_3 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ & & & & & \end{bmatrix}$$

so that

$$Q_3\beta = \begin{bmatrix} \beta_0+\beta_1 & \beta_0+\beta_2 & \beta_3 & \beta_4 & \beta_5 \end{bmatrix}'.$$

It can be readily verified that

$$\text{rank}\begin{bmatrix} X \\ Q_3 \end{bmatrix} = \text{rank}(X) = 5.$$

Thus $Q_3\beta$ is identifiable.

Alternatively, we can substitute "1-Female" for "Male" in (2.8.14) and get

$$Wage = \lambda_0 + \lambda_1 Female + \lambda_2 Educ + \lambda_3 Exper + \lambda_4 Age + u, \tag{2.8.16}$$

where

$$\lambda_0 = \beta_0+\beta_1,\ \lambda_1 = \beta_2-\beta_1,\ \lambda_2 = \beta_3,\ \lambda_3 = \beta_4,\ \lambda_4 = \beta_5.$$

The selection matrix Q_4 is

$$Q_4 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ & & & & & \end{bmatrix}$$

and

$$Q_4\beta = \begin{bmatrix} \beta_0+\beta_1 & \beta_2-\beta_1 & \beta_3 & \beta_4 & \beta_5 \end{bmatrix}'.$$

Again

$$\text{rank}\begin{bmatrix} X \\ Q_4 \end{bmatrix} = \text{rank}(X) = 5$$

and $Q_4\beta$ is identifiable.

Similarly, "Female" in (2.8.14) can be replaced by "1-Male" and (2.8.14) is rewritten as

$$Wage = \delta_0 + \delta_1 Male + \delta_2 Educ + \delta_3 Exper + \delta_4 Age + u, \tag{2.8.17}$$

where

$$\delta_0 = \beta_0 + \beta_2,\ \delta_1 = \beta_1 - \beta_2,\ \delta_2 = \beta_3,\ \delta_3 = \beta_4,\ \delta_4 = \beta_5.$$

Then we choose

$$Q_5 = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \end{bmatrix}$$

and $Q_5\beta$ is identifiable since

$$\text{rank}\begin{bmatrix} X \\ Q_5 \end{bmatrix} = \text{rank}(X) = 5.$$

We have already seen that Theorem **2.5.4** is very useful to check identifiability of linear parameters in practice. On the other hand, we demonstrate that the basis of a vector space also provides useful information on identification given by Proposition **2.6.1**. First using function "rref" given by Matlab, we get the reduced row echelon form of X' as [3]

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & \dots & 0.3250 & 0.1500 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0.8193 & 0.4077 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0.3058 & -0.6577 \\ 0 & 0 & 0 & 0 & 1 & \dots & -0.4500 & 0.1000 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ \end{bmatrix}.$$

Since performing Gauss-Jordan elimination on rows only will not change the linear relationship among the columns of X', we can pick up the first 5 columns of X' and form a

[3] We can also use other decomposition methods such as QR, LU and SVD to find the basis of $\text{Im}(X')$.

basis of $\mathrm{Im}(X')$. Denote this submatrix of X' as B_1, *i.e.*

$$B_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 6 & 4 & 6 & 6 & 9 \\ 2 & 18 & 4 & 4 & 3 \\ 38 & 52 & 45 & 58 & 30 \end{bmatrix}.$$

Obviously, ${B_1}'\beta$ is identifiable due to Proposition **2.6.1** or condition (2.5.20). Yet, we want to make identifiable parameters have economic meanings. For instance we are interested in identification of $\beta_0+\beta_2$ which is average salary for the "Female" group without considering other explanatory variables rather than $\beta_0+\beta_2+6\beta_3+2\beta_4+38\beta_5$ which makes little sense for interpreting the parameters. Nevertheless, we can get another basis B_2 of $\mathrm{Im}(X')$ based on B_1 as follows

$$B_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Then the parameter ${B_2}'\beta$ is identifiable from Proposition **2.6.1** and

$${B_2}'\beta = \begin{bmatrix} \beta_0+\beta_2 & \beta_0+\beta_1 & \beta_3 & \beta_4 & \beta_5 \end{bmatrix}',$$

where each element has a good economic interpretation. Since the basis of any linear space is not unique, it is always possible to build an infinite number of bases of a vector space by multiply a known basis by an invertible matrix called transitional matrix. For our specific model setups (2.8.15), (2.8.16) and (2.8.17), we construct other bases of $\mathrm{Im}(X')$ by postmultiply B_2 by a transitional matrix T so that the identifiable parameters are easy to

interpret. Hence

$$B_3 \equiv B_2 T_1 = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{B_2} \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{T_1} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = Q_3{}'.$$

Similarly,

$$B_4 \equiv B_2 T_2 = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{B_2} \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{T_2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = Q_4{}'$$

and

$$B_5 \equiv B_2 T_3 = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{B_2} \underbrace{\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{T_3} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = Q_5{}'.$$

Therefore, the transposed selection matrices Q_3, Q_4 and Q_5 are actually different bases of $\mathrm{Im}\,(X')$ and $B_3'\beta$, $B_4'\beta$ and $B_5'\beta$ are all identifiable.

2.8.2. Binary choice logit model

The example here is about the determinants of women's labor force participation rates, *i.e.*, the percentage of women over age 16 who are either employed or looking for jobs [see Ramanathan (2002)]. The 1990 census data cover 50 states in terms of the following variables:

$$\begin{aligned}
\text{Wlfp} &= \text{Proportion of women over 16 who are in labor force} \\
\text{Yf} &= \text{Median earning by females (in thousands of dollars)} \\
\text{Ym} &= \text{Median earning by males (in thousands of dollars)} \\
\text{Educ} &= \text{Percentage of female high school graduates over 24} \\
\text{Unemp} &= \text{Unemployment rate} \\
\text{Marr} &= \text{Marriage rate of women at least 16} \\
\text{Divc} &= \text{Divorce rate} \\
\text{Urb} &= \text{Percent of urban population} \\
\text{Wht} &= \text{Percent of women over 16 who are white.}
\end{aligned}$$

Suppose Wlfp satisfies a logit specification as follows:

$$\log\left(\frac{Wlfp}{1-Wlfp}\right) = X\beta \tag{2.8.18}$$

$$= \beta_0 + \beta_1 Yf + \beta_2 Ym + \beta_3 Educ + \beta_4 Unemp + \beta_5 Marr + \beta_6 Divc + \beta_7 Urb + \beta_8 Wht. \tag{2.8.19}$$

Since the log function is strictly monotonic over its domain, it is easily to write (2.8.19) as

$$Wlfp = \frac{\exp(X\beta)}{1+\exp(X\beta)}.$$

If we denote the binary variable "women labor force" as "WL" then

$$\mathbb{E}(WL) = \frac{\exp(\beta_0 + \beta_1 Yf + \beta_2 Ym + \beta_3 Educ + \beta_4 Unemp + \beta_5 Marr + \beta_6 Divc + \beta_7 Urb + \beta_8 Wht)}{1+\exp(\beta_0 + \beta_1 Yf + \beta_2 Ym + \beta_3 Educ + \beta_4 Unemp + \beta_5 Marr + \beta_6 Divc + \beta_7 Urb + \beta_8 Wht)}.$$

The identification of $X\beta$ is straightforward since it is a monotone function of the mean $\mathbb{E}(WL)$. As we have demonstrated in subsection 2.8.1, if the right hand side of the model (2.8.19) is misspecified by including variables that are exact multicollinear, we will not be able to identify β in a logit setup as in (2.8.19). However, applying the identification

results we can still identify some linear combinations of the elements of β that have certain economic meanings. Furthermore, we know that such identifiable linear combinations will not be unique due to the infinite number of basis of a given vector space spanned by the columns of the design matrix. It is essential to choose a basis that can produce a linear function of β which is easy to interpret.

2.9. Conclusion

We summarize the major points and contributions of this paper as follows. First, we analyze the conditions for identification of linear parameters $Q\beta$ when the design matrix X does not have full column rank. The current literature is mainly concerned about estimability of parameters without a formal verification of parameter identification, which could lead to unreliable or even misleading statistical inference; see Dufour (2003). This paper has established the necessary and sufficient conditions of identification for an arbitrary linear combination of parameters, such as any scalar parameter, a subvector of parameters and the entire parameter vector.

Second, we demonstrate that the identification conditions are quite general. To begin with, no restrictions are imposed on the structure of the model so that our identification results can be applied to almost all types of statistical models frequently used so far, such as general linear models, generalized linear models, generalized linear mixed models, quantile regressions and single index models, etc. Therefore, our necessary and sufficient conditions provide a very powerful tool to check identification before any meaningful statistical inference should be conducted. Additionally, these conditions are still valid in spite of the presence of nuisance parameters which is a common phenomenon in statistics. When the underlying probability distribution involves nuisance parameters, the identification condition in Proposition **2.3.2** is only sufficient. However, if the parameter of interest can be completely separated from the nuisance parameter after we reparameterize the underlying distribution, as is the case for a large number of partially linear models, the identification condition in Theorem **2.3.4** is still necessary and sufficient. Furthermore, adding a group of a priori consistent linear restrictions will have no impact on the validity of the necessary and sufficient conditions for identification derived in this paper.

Third, we study the conditions for both estimability and identifiability of linear parameters. As for the linear regression models, under the restrictive classical assumptions the well-known necessary and sufficient conditions for estimability [see Reiersøl (1963) and Rao (1973)] are generally also necessary and sufficient for identification. However, considering the more general class of partially linear models which relaxes the model as-

sumptions of additivity of the error term and the complete separation of the mean from the covariance, the conditions for identification will not be applicable to estimability. In fact, the linear parameters in the partially linear models will generally no longer be estimable in the sense of Bose (1944) because the linear predictor $X\beta$ is usually not the mean any more but a nonlinear function of the mean. Nevertheless, we have shown that the classical conditions for estimability in the linear regression models are still valid for identification of the linear parameters in the partially linear models. We emphasize that even for the linear regression model, the identification of linear parameters is not restricted to the mean of the observations. It is possible to identify these parameters through other statistical properties such as covariances, quantiles or statistical estimators.

Fourth, we propose a set of equivalent conditions for identification which are intuitive and easy to apply. The standard conditions for estimability and identifiability focus on characteristics of the row spaces, which is somewhat inconvenient for empirical analysis since people conventionally investigate the relationships among the explanatory variables. Thus the introduced conditions for identification concentrating on the column spaces are straightforward. Moreover, since there are no restrictions on the selection matrix Q we generalize the identification conditions by including the statements regarding the g-inverse Q^{-}, which is quite useful in practice.

Hence, the identification conclusions in this paper are comprehensive and widely applicable. No matter what statistical properties we are considering, as long as they are identifiable any linear combinations of parameters which are functions of such properties can also be identified. The statistical inferences of such identified parameters will be the subject of future research.

Chapter 3

Necessary and sufficient conditions for identification of parametric functions in IV regressions

Xin Liang

Abstract We provide the general conditions for identification in the simultaneous equations models (SEMs). Compared to the classical SEMs setup, the design matrix of the reduced form under discussion is not assumed to have full column rank. Under such a circumstance, an arbitrary linear combination of the structural parameters may still be identifiable when the proposed identification conditions are satisfied. Moreover, we relax the classical assumption that the design matrix is exogenous or fixed and analyze the conditions for identification of structural parameters when the expectations of endogenous variables in the simultaneous equations are conditional on some chosen instruments X. By introducing the randomness into the design matrix, we extend the application of our results to a more general framework. Furthermore, we impose no exogeneity constraints on the instruments and some instruments are allowed to be endogenous. This has a quite striking influence on identification research because how to select strong instruments from an infinite set is still an open issue. As an illustration, the identification conditions can be easily applied to the classical SEMs by Richmond (1974) which are included as a special case of our general setup. In addition, the identification conditions in the SEMs of our paper are valid without any restrictions on the specification of the model itself and they are readily applicable to cases where nuisance parameters are present.

3.1. Introduction

There have been a considerable number of papers about the estimability of parameters in linear models; see Seely (1970), Milliken (1971), Eubank and Webster (1985), Baksalary and Kala (1976), Alalouf and Styan (1979*a*), Seely and Birkes (1980), Kounias and Chalikias (2008), etc. In contrast, identifiability has been relatively less studied in literature, not to mention the identification conditions of any arbitrary linear combinations of structural parameters in a system of simultaneous equations; see the review of the long history of identifiability by Dufour and Hsiao (2008). Working (1927) initiates the discussion of the nature of the identifiability issue in economics and mentions the importance of methods of analysis and the variability of demand curves. Fisher (1966) systematically studies the identification problems for a single structural equation of the simultaneous equations models (SEMs). Among the few literatures that touch on the identifiability issue, the research on the identification conditions of the parameters is somewhat limited. First of all, almost all classical analyses of identifiability assume the design matrix has a full column rank. Specifically, let's consider a simple SEMs setup:

$$y = Y\beta_1 + X_1\gamma_1 + u, \tag{3.1.1}$$

$$Y = X_1\Pi_1 + X_2\Pi_2 + v, \tag{3.1.2}$$

where y is a $T \times 1$ endogenous vector, Y is a $T \times G$ matrix of endogenous variables, $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ and X is assumed to have a full column rank with X_i of a dimension of $T \times K_i (i = 1, 2)$, β_1 and γ_1 are $G \times 1$ and $K_1 \times 1$ vectors of unknown coefficients, $\Pi = \begin{bmatrix} \Pi_1' & \Pi_2' \end{bmatrix}'$ and Π_1 and Π_2 are $K_1 \times G$ and $K_2 \times G$ matrices of unknown coefficients, u and v are error terms with zero means. The traditional way of studying the SEMs focuses on establishing the identification conditions of the structural parameters β_1 and γ_1; see Richmond (1974) and Dufour and Hsiao (2008). Plugging (3.1.2) into (3.1.1), we have

$$y = X_1\pi_1 + X_2\pi_2 + w,$$

where $\pi_1 = \Pi_1\beta_1 + \gamma_1$, $\pi_2 = \Pi_2\beta_1$ and $w = v\beta_1 + u$. The standard rank condition which is both necessary and sufficient for the identifiability of β_1 (and thus γ_1) is $\text{rank}(\Pi_2) = G$. Meanwhile, the corresponding standard order condition which is only necessary and not sufficient is that the number of excluded exogenous variables which is K_2 must be greater or equal to the number of included endogenous variables minus one which is G. It is thus well-known that the rank condition implies the order condition. Under the assumption

of full rank X, the coefficients in the reduced form of simultaneous equations are always identifiable. This requirement is, however, rather restrictive and unrealistic in applications. Quite often, the observational data will produce a design matrix that either has a deficient rank or is close to be deficient which makes the identification of the entire parameter vector of interest impossible. Second, the existing identification conditions rely heavily on pure mathematical derivations, which makes them difficult to apply; see Bowden (1973) and Rothenberg (1971). Third, almost no literature has ever studied the necessary and sufficient conditions of identifiability in the presence of nuisance parameters. As is well know, nuisance parameters are quite common in statistical models and will cause serious problems when it comes to statistical inference.

In this paper, we intend to provide the most general statements of identification by dealing with all the above mentioned limitations of current research. First, we relax the assumption that the design matrix in the reduced form of the simultaneous equations system has full column rank. Indeed, whether or not the design matrix has full rank will only have an effect on identification of the reduce form parameters and is not directly related to identification of structural parameters. As for the multivariate SEMs, even if the design matrix is deficient in rank, with the help of additional a priori information on the structural parameters, we can still achieve full identification. On the other hand, full identification of the univariate SEMs without instruments is possible if our general conditions for identification are satisfied. Furthermore, although we sometimes cannot identify the whole set of structural parameters when X has deficient rank, we may still be able to identify some linear combinations of the structural parameters, which include the subvector and the scalar parameter as special cases. As a further extension to the standard results, we do not impose the assumption that the coefficient matrix of the endogenous variables is nonsingular. Hence the rank conditions and its equivalences proposed are valid even when the reduce form equation does not have a unique expression. Second, we relax the assumption that the design matrix X is exogenous or fixed so that the expectations of the endogenous variables in the simultaneous equations are now conditional on X and the parameter identification depends on the realized value x of X. This leads to a more general setup of SEMs in that no exogeneity constraints are imposed on the instruments and some instruments are allowed to be endogenous. Moreover, we allow for collinearity among the instruments, i.e., weak instruments. Besides, our results can deal with other instrumental variable issues, such as missing instruments and many instruments. Thus our conditions for identification have an influence on the empirical research since how to select strong instruments from an infinite set of candidates is still an open issue and we are not so sure about having chosen all relevant instruments. Third, we consider the situation where nuisance parameters are involved

in the identification of parameters in simultaneous equations system. Last, we show that the general theoretical results established by our paper are easy to understand and convenient to apply to a wide range of statistical models.

The paper is organized as follows. Section 3.2 introduces a general setup of the SEMs which allows for a variety of specifications, such as a mixture of endogenous variables and exogenous variables, multicollinearity among the columns of the design matrix, introduction of instruments and nuisance parameters. Section 3.3 provides a group of equivalent conditions for identification with brief explanations on each condition. Section 3.4 discusses the applications of the established results to some well-known SEMs. It shows that the standard results can be easily generalized as special cases of our rank conditions. Section 3.5 concludes.

3.2. The general framework of simultaneous equation models

We introduce a general setup of the simultaneous equations system that imposes no constraints on the specification of the system of equations and makes no assumptions on the linear relationships among the columns of either endogenous or exogenous variables. This is followed by the discussions of identification both through the conditional expectation and the linear projector.

3.2.1. Definitions

First recall the definitions of identification of parametric functions [Dufour and Liang (2012)]. Let $(\Omega, \mathscr{A}, \mathscr{P})$ be a statistical model parameterized by θ. Denote $\psi(\theta)$ as a general parametric function.

Definition 3.2.1 IDENTIFICATION OF PARAMETRIC FUNCTIONS. *The parametric function $\psi : \Theta \to \Psi$ is identifiable if and only if*

$$\left(\psi(\theta_1) \neq \psi(\theta_2)\right) \Rightarrow \left(\mathsf{P}_{\theta_1} \neq \mathsf{P}_{\theta_2}\right), \quad \forall \theta_1, \theta_2 \in \Theta. \tag{3.2.1}$$

Definition 3.2.2 IDENTIFICATION IN TERMS OF ANOTHER PARAMETRIC FUNCTION. *Let $\beta : \Theta \to B$ and $\gamma : \Theta \to \Gamma$ be two parametric functions on the statistical model $(\Omega, \mathscr{A}, \mathscr{P})$ where $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$. If $\gamma(\theta)$ is identifiable and*

$$\left(\beta(\theta_1) \neq \beta(\theta_2)\right) \Rightarrow \left(\gamma(\theta_1) \neq \gamma(\theta_2)\right), \quad \forall \theta_1, \theta_2 \in \Theta,$$

we say that $\beta(\theta)$ is $\gamma(\theta)$-identifiable.

To facilitate the analysis of identification through instruments in the following section, we give three related definitions [see Rao and Mitra (1971)].

Definition 3.2.3 DIRECT SUM. *Let $\mathscr{A}_1$ and $\mathscr{A}_2$ be subspaces of a linear space of $\mathscr{A}$. If*

(1) *$\mathscr{A}_1 \bigcap \mathscr{A}_2$ consists of only the null vector;*

(2) *every vector $x \in \mathscr{A}$ can be expressed as $x = x_1 + x_2$, such that $x_1 \in \mathscr{A}_1$ and $x_2 \in \mathscr{A}_2$;*

then $\mathscr{A}$ is called the direct sum of $\mathscr{A}_1$ and $\mathscr{A}_2$, which is denoted as

$$\mathscr{A} = \mathscr{A}_1 \oplus \mathscr{A}_2.$$

Definition 3.2.4 PROJECTION OPERATOR. *Let any vector $x \in \mathscr{A}$ which is a direct sum of $\mathscr{A}_1$ and $\mathscr{A}_2$ such that $x = x_1 + x_2$, $x_1 \in \mathscr{A}_1$ and $x_2 \in \mathscr{A}_2$. Then the linear mapping denoted as $\mathbf{P}_{\mathscr{A}_1 \| \mathscr{A}_2} : x \to x_1$ is called the projection operator of x on $\mathscr{A}_1$ along $\mathscr{A}_2$.*

Definition 3.2.5 ORTHOGONAL PROJECTION OPERATOR. *In the special case where $\mathscr{A}_2$ is the orthogonal complement of $\mathscr{A}_1$ in $\mathscr{A}$, the projection operator is called the orthogonal projection operator denoted as $\mathbf{P}_L$.*

We will simply call the "orthogonal projection operator" the "projection operator" hereinafter if such a simplification does not cause any confusion.

Now let's consider the following simultaneous equation model setup

$$Z = \overline{Y}A + W, \tag{3.2.2}$$

where Z is $T \times H$, $\overline{Y}$ is $T \times G$, the unknown coefficients matrix A is $G \times H$ and the error term W is $T \times H$. We assume $\mathbb{E}(W) = 0$ and call Z the endogenous variables. Taking expectations on both sides of equation (3.2.2), we obtain

$$\mathbb{E}(Z) = \mathbb{E}(\overline{Y})A. \tag{3.2.3}$$

(3.2.3) is called a linear system and the linearity is in terms of the unknown coefficient A. We replicate some of the related well-known statements for convenience. First, any value of the parameter A that satisfies the linear system (3.2.3) is called a solution to the linear system. Second, the linear system (3.2.3) is consistent if and only if it has at least one solution; see Harville (2008).

Remark 3.2.6 The setup of $\overline{Y}$ is inclusive since it can either contain only exogenous variables or consists of both endogenous and exogenous variables. Besides, it can be either a random variable or a fixed variable as in the standard linear regression model. Furthermore, we impose no assumptions on the rank of $\mathbb{E}(\overline{Y})$ so that if $\mathbb{E}(\overline{Y})$ is deficient in rank, it is impossible to identify the entire matrix A without further information. However, we can still achieve a unique solution of some linear combination of A provided that certain conditions are satisfied, which becomes a more interesting issue for empirical studies and sometimes the best possible choice we can make. In addition, the general framework (3.2.2) takes into account the situation when the underlying probability distribution family depends on both the parameters of interest and the nuisance parameters.

3.2.2. Identification of a linear system

We turn to the relationship between identification of coefficient matrix A of (3.2.3) and the unique solution to such a linear system. Let θ denote all deep parameters in the linear SEMs (3.2.2). Also let $\gamma_1 : \Theta \to \Gamma_1$, $\gamma_2 : \Theta \to \Gamma_2$ and $\beta : \Theta \to B$ be three parametric functions, which stand for the identifiable parameter, the nuisance parameter and the parameter of interest respectively. First of all, we claim that the existence of a unique solution to (3.2.3) is only sufficient and not necessary for identification on A. We then provide both necessary and sufficient condition for identification under some stronger assumption.

Theorem 3.2.7 SUFFICIENT CONDITION OF IDENTIFICATION OF PARAMETER MATRIX. *If there exists a unique solution to (*3.2.3*), the parameter A in the SEMs (*3.2.2*) is identifiable.*

Since the general setup (3.2.2) allows for the situation where the underlying distribution of the error term W depends on nuisance parameters, the existence of a unique solution to (3.2.3) is not necessary for identification of A. Recall from the conclusion of necessary and sufficient condition for identification with nuisance parameters [Dufour and Liang (2012, Theorem 2.3.4)], if the underlying distribution of W involves nuisance parameters, identification of A can be achieved through other statistical properties rather than the mean of Z. In other words, we can identify A using the useful information provided by the nuisance parameters. To make the result in Theorem **3.2.7** both necessary and sufficient, we need to impose an additional assumption that the nuisance parameter can be completely separated from the parameter of interest; see Dufour and Liang (2012, Assumption 2.3.3).

Assumption 3.2.8 NUISANCE PARAMETER SEPARABILITY. *For any $\theta \in \Theta$ and $\bar{\gamma}_2 \in$*

$\gamma_2(\Theta)$, *we can find* $\theta^* \in \Theta$ *such that*

$$\beta(\theta^*) = \beta(\theta), \quad \gamma_1(\theta^*) = \gamma_1(\theta) \text{ and } \gamma_2(\theta^*) = \bar{\gamma}_2. \tag{3.2.4}$$

This assumption means that the value of the nuisance parameter $\gamma_2(\theta)$ is not affected by the values of the identifiable parameter $\gamma_1(\theta)$ and the parameter of interest $\beta(\theta)$. In other words, the parametric function $\gamma_2(\theta)$ is "free" in terms of $\gamma_1(\theta)$ and $\beta(\theta)$. In the case of model (3.2.2), it follows that we can set

$$\gamma_1(\theta) = \mathbb{E}(Z),\ \beta(\theta) = A.$$

Note that the nuisance parameter $\gamma_2(\theta)$ can be part of the characterizations of the distribution of W. However, as long as the value of the nuisance parameter is not affected by the values of $\mathbb{E}(Z)$ and A, the unique solution to (3.2.3) is both necessary and sufficient for identification of A whether or not the nuisance parameter is identifiable. This can lead to rigorous statistical inferences.

Theorem 3.2.9 NECESSARY AND SUFFICIENT CONDITION FOR IDENTIFICATION OF PARAMETER MATRIX. *Suppose Assumption* **3.2.8** *is satisfied. The parameter A in the SEMs (*3.2.2*) is identifiable if and only if there exist a unique solution to (*3.2.3*).*

Therefore, despite the presence of nuisance parameters, it is still possible to use the techniques discussed by Dufour and Liang (2012, Theorem 2.3.4) so that the probability distributions can be reparameterized through a complete separation of nuisance parameters and parameters of interest as in the case of the generalized linear models and other partially linear models in Dufour and Liang (2012). Hence we can achieve identification of parameters of interest without concerning whether or not the nuisance parameters are identifible, which eases the burden of dealing with nuisance parameters and leads to the valid statistical inferences.

3.2.3. Identification through instruments

Often we can modify the assumption of zero unconditional mean of W by introducing some random matrix X with dimension $T \times K$ whose K variables are called instrumental variables or instruments. If we assume $\mathbb{E}(W|X) = 0$ then equation (3.2.3) becomes

$$\mathbb{E}(Z|X) = \mathbb{E}(\overline{Y}|X)A. \tag{3.2.5}$$

Since the conditional expectation $\mathbb{E}(\overline{Y}|X)$ can be interpreted as the orthogonal projection of $\bar{Y}$ on any square integrable function of X [see Gouriéroux and Monfort (1995)], both $\mathbb{E}(Z|X)$ and $\mathbb{E}(\overline{Y}|X)$ can be nonlinear functions of X. On the other hand, we can express equation (3.2.5) in terms of the projection operator $\mathbf{P}_L$ as

$$\mathbf{P}_L(Z|X) = \mathbf{P}_L(\overline{Y}|X)A. \tag{3.2.6}$$

Likewise, nuisance parameters are allowed in the conditional distribution of W and we can set $\gamma_1(\theta)$, $\gamma_2(\theta)$ and $\beta(\theta)$ equal to $\mathbb{E}(Z|X)$ (or $\mathbf{P}_L(Z|X)$), the nuisance parameter and A in the setups (3.2.5) and (3.2.6). Similar to the analysis in Subsection 3.2.2, we give the following results for identification of A both through the conditional expectation and the linear projector.

Proposition 3.2.10 IDENTIFICATION THROUGH THE CONDITIONAL EXPECTATION. *If there exists a unique solution to (*3.2.5*), the parameter A in the SEMs (*3.2.2*) is identifiable. Furthermore, suppose Assumption* **3.2.8** *is satisfied, the existence of such a unique solution is both necessary and sufficient for identification of A in (*3.2.2*).*

Proposition 3.2.11 IDENTIFICATION THROUGH THE LINEAR PROJECTOR. *If there exist a unique solution to (*3.2.6*), the parameter A in the SEMs (*3.2.2*) is identifiable. Furthermore, suppose Assumption* **3.2.8** *is satisfied, the existence of such a unique solution is both necessary and sufficient for identification of A in (*3.2.2*).*

Since the linear projection operator $\mathbf{P}_L(\overline{Y}|X)$ is derived by minimizing the mean squared prediction error in a linear regression of $\overline{Y}$ on X, it is a special case of $\mathbb{E}(\overline{Y}|X)$. Under the assumption of invertibility of the covariance matrix of each column of (X), we have

$$\mathbf{P}_L(\overline{Y}_j|X_i) = \mathbb{E}(\overline{Y}_j) + \mathbb{C}(X_i, \overline{Y}_j)\mathbb{V}(X_i)^{-1}\big(X_i - \mathbb{E}(X_i)\big),\ i = 1,2,\ldots,K,\ j = 1,2,\ldots,G, \tag{3.2.7}$$

where $\mathbb{C}(X_i, \overline{Y}_j)$ is the covariance of X_i and $\overline{Y}_j$. The immediate issue regarding identifiability of A in (3.2.5) and (3.2.6) is whether the identification conditions in both situations are equivalent. That is to say, can the identifiability of A in (3.2.5) lead to the identifiability of A in (3.2.6) and vice versa?

For illustration purpose, we consider only the univariate regressor x_i and scalar parameter case where the regressor x_i and the error term u_i are uncorrelated and the conditional mean of the error term is zero. First, let's assume a linear regression model as follows:

$$y_i = \alpha(x_i^2 - 1) + u_i,\ i = 1,2,\ldots,n,$$

where α is the scalar parameter and $x_i \sim \mathrm{N}[0,1]$. Then

$$\mathbb{C}(x_i, y_i) = \mathbb{E}(u_i x_i) + \mathbb{E}\big(\alpha x_i (x_i^2 - 1)\big) = \alpha \mathbb{E}(x_i^3) - \alpha \mathbb{E}(x_i) = 0$$

and thus

$$\mathbf{P}_L(y_i|x_i) = \mathbb{E}(y_i) = 0.$$

However, $\mathbb{E}(y_i|x_i) = \alpha(x_i^2 - 1)$ which is a random variable is not equal to zero unless $x_i = \pm 1$. Clearly, the scalar parameter α can be identified through the conditional expectation $\mathbb{E}(y_i|x_i)$ but not the linear projector $\mathbf{P}_L(y_i|x_i)$ when $x_i \neq \pm 1$. Since $\mathbb{E}(y_i|x_i)$ is a function of the random variable x_i and is itself random, it is not equal to a zero constant unless x_i takes on some specific realized values. In contrast, the linear projector $\mathbf{P}_L(y_i|x_i) = \mathbb{E}(y_i) + \mathbb{C}(x_i, y_i)\mathbb{V}(x_i)^{-1}(x_i - \mathbb{E}\big(x_i\big))$, can be a constant or a function of x_i depending on the moments of x_i and y_i. Thus identification of A through the conditional expectation does not imply that through the linear projector. On the other hand, we want to check if the reverse is true. We first claim that the following statement holds.

Lemma 3.2.12 FAILURE OF IDENTIFICATION THROUGH THE CONDITIONAL EXPECTATION IMPLIES FAILURE OF IDENTIFICATION THROUGH THE LINEAR PROJECTOR.

$$\big(\mathbb{E}(\overline{Y}|X) = 0\big) \Rightarrow \big(\mathbf{P}_L(\overline{Y}|X) = 0\big), \tag{3.2.8}$$

or equivalently

$$\big(\mathbf{P}_L(\overline{Y}|X) \neq 0\big) \Rightarrow \big(\mathbb{E}(\overline{Y}|X) \neq 0\big). \tag{3.2.9}$$

Although (3.2.8) means that if A cannot be uniquely determined through the conditional expectation it cannot be uniquely determined through the linear projector either, we cannot conclude from (3.2.9) that identification of A through $\mathbf{P}_L(\overline{Y}|X)$ implies that through $\mathbb{E}(\overline{Y}|X)$ because $\mathbb{E}(\overline{Y}|X) \neq 0$ or $\mathbf{P}_L(\overline{Y}|X) \neq 0$ does not provide any information on the rank of $\mathbb{E}(\overline{Y}|X)$ or $\mathbf{P}_L(\overline{Y}|X)$ and thus does not necessarily lead to a unique solution of A in (3.2.6) or (3.2.5). Specifically, we need to impose restrictions on the relationships among the columns of both $\mathbb{E}(\overline{Y}|X)$ and $\mathbf{P}_L(\overline{Y}|X)$ so that they have full column rank. Otherwise, it can be the case that both of $\mathbb{E}(\overline{Y}|X)$ and $\mathbf{P}_L(\overline{Y}|X)$ are nonzero but with different ranks. It then follows that the elements of A that can be identified through the conditional expectation may be different from those identified through the linear projector. Hence, the identifiability conditions for A in (3.2.5) and (3.2.6) are not equivalent despite the fact that the linear

projection operator is a special case of the conditional expectation. The critical point is that $\mathbb{E}(\overline{Y}|X)$ is a function of X and its value depends on the realized values of X. In contrast, the linear projector $\mathbf{P}_L(\overline{Y}|X)$ involves both the realized values of X and the unconditional second order moments of X and $\overline{Y}$ which may contain some useful information that will help identify the parameters of interest but is not included in the conditional expectation. Consequently, the failure of identification of A through the conditional expectation $\mathbb{E}(\overline{Y}|X)$ implies the failure of identification of A through the linear projector $\mathbf{P}_L(\overline{Y}|X)$ but the reverse is not true.

We now take a brief look at the identification conditions of linear parameters in the general setup (3.2.2). Without loss of generality, let's assume the random variable $\overline{Y}$ contains some of the columns of the exogenous matrix X. Under such a circumstance, we can partition $\overline{Y}$ as

$$\overline{Y} = \begin{bmatrix} Y_1 & X_1 \end{bmatrix},$$

where Y_1 is endogenous with a dimension of $T \times G_1$, X_1 is exogenous with a dimension of $T \times K_1$ and $X_1 \subset X$. Note that $X_1 = X$ is just a trivial case in this setup.

Also A can be partitioned conformably as

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix},$$

where A_1 is $G_1 \times H$ and A_2 is $K_1 \times H$. Thus equation (3.2.5) can be rewritten as

$$\begin{aligned} \mathbb{E}(Z|X) &= \mathbb{E}(\overline{Y}|X)A \\ &= \begin{bmatrix} \mathbb{E}(Y_1|X) & X_1 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \\ &= \mathbb{E}(Y_1|X)A_1 + X_1A_2, \end{aligned} \tag{3.2.10}$$

where

$$\mathbb{E}(\overline{Y}|X) = \begin{bmatrix} \mathbb{E}(Y_1|X) & X_1 \end{bmatrix}.$$

It is clear that since both $\mathbb{E}(Z|X)$ and $\mathbb{E}(Y_1|X)$ are the first moments conditional on X, they are identifiable for any realized value x of X.

Remark 3.2.13 The exogenous matrix X_1 in the structural form (3.2.10) is assumed to have

full column rank in the classical literature. Meanwhile, if we are further given a system of reduced equations with extra exogenous matrix X_2 being excluded from (3.2.10), the augmented matrix $[\ X_1 \quad X_2\ ]$ is assumed to be full in rank; see Fisher (1966), Rothenberg (1971), Richmond (1974) and Bekker and Wansbeek (2001). However, this is not the case in our general setup as we only require the existence of the conditional moment $\mathbb{E}(\overline{Y}|X)$. Our interest hereby is to establish the necessary and sufficient conditions of identifiability of arbitrary linear combinations of parameters $Q(X)A = \Theta$, where $Q(X)$ is a function of X with a dimension of $q \times G$, provided that $\mathbb{E}(\overline{Y}|X)A$ is identifiable, which is indeed the case.

Clearly, the propositions regarding necessary and sufficient conditions for identification that we have already established in Dufour and Liang (2012) can be readily applied in the SEMs under discussion. In particular, we can treat both $Q(X)A$ and $\mathbb{E}(\overline{Y}|X)A$ as parametric functions of A. Then the necessary and sufficient condition for the identification of $Q(X)A$ is that $Q(X)A$ can be expressed as a function of $\mathbb{E}(\overline{Y}|X)A$.

3.3. Necessary and sufficient conditions for identification

We can simplify the process of identifying any arbitrary linear combination of parameter A in Section 3.2 and establish the necessary and sufficient conditions for identification which are more intuitive and easier to check in practice. We first propose a necessary and sufficient condition for identification of an arbitrary linear combinations of coefficient matrix A followed by a group of equivalent statements based on rank conditions and characteristics of vector subspaces. Let $\mathbb{E}(\overline{Y}|X)$ and $Q(X)$ be $T \times G$ and $q \times G$ matrices of real numbers. Suppose Assumption **3.2.8** is satisfied and we can set

$$\gamma_1(\theta) = \mathbb{E}(\overline{Y}|X)A,\ \beta(\theta) = Q(X)A.$$

Meanwhile assume $\mathbb{E}(\overline{Y}|X)A$ is identifiable and (3.2.5) is consistent.

Definition 3.3.1 COLUMN SPACE. *Let A be a $m \times n$ matrix. The column space of A denoted as* $\mathrm{Im}(A)$ *is the set*

$$\mathrm{Im}(A) \equiv \{y \in \mathbb{R}^m : Ax = y\}.$$

Proposition 3.3.2 NECESSARY AND SUFFICIENT CONDITION FOR A UNIQUE EXPRESSION OF LINEAR COMBINATIONS OF PARAMETER MATRIX. *$Q(X)A$ has a unique expression if and only if*

$$\mathrm{Im}\Big(\big(Q(X)\big)'\Big) \subseteq \mathrm{Im}\Big(\big(\mathbb{E}(\overline{Y}|X)\big)'\Big).$$

Proposition 3.3.3 EQUIVALENT CONDITIONS FOR IDENTIFIABILITY. *Suppose Assumption* **3.2.8** *holds. Then* $Q(X)A$ *is* $\big(\mathbb{E}(\overline{Y}|X)A\big)$*-identifiable if and only if one of the following statements holds:*

$$\begin{gathered}\textit{There is a function } g : \mathrm{Im}\big[\mathbb{E}(\overline{Y}|X)\big] \mapsto \mathrm{Im}\big(Q(X)\big) \textit{ such that} \\ Q(X)A = g\big(\mathbb{E}(\overline{Y}|X)A\big),\ \forall A \in \mathbb{R}^{G\times H};\end{gathered} \tag{3.3.1}$$

$$\big(\forall A_1,\, A_2 \in \mathbb{R}^{G\times H}\big)\Big[\big(\mathbb{E}(\overline{Y}|X)A_1 = \mathbb{E}(\overline{Y}|X)A_2\big) \Rightarrow \big(Q(X)A_1 = Q(X)A_2\big)\Big]; \tag{3.3.2}$$

$$\ker\big(\mathbb{E}(\overline{Y}|X)\big) \subseteq \ker\big(Q(X)\big); \tag{3.3.3}$$

$$\mathrm{Im}\Big(\big(Q(X)\big)'\Big) \subseteq \mathrm{Im}\Big(\big(\mathbb{E}(\overline{Y}|X)\big)'\Big); \tag{3.3.4}$$

$$Q(X) = D\mathbb{E}(\overline{Y}|X), \textit{ for some matrix } D; \tag{3.3.5}$$

$$\mathrm{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ Q(X) \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)\big); \tag{3.3.6}$$

$$\mathrm{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) + V_1 Q(X) \\ Q(X) \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)\big), \textit{ for some } T \times q \textit{ matrix } V_1; \tag{3.3.7}$$

$$\mathrm{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ Q(X) + V_2\mathbb{E}(\overline{Y}|X) \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)\big), \textit{ for some } q \times T \textit{ matrix } V_2; \tag{3.3.8}$$

$$\mathrm{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ SQ(X) \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)\big), \textit{ where } S \textit{ is } s \times q \textit{ and } s = \mathrm{rank}\big(SQ(X)\big) = \mathrm{rank}\big(Q(X)\big). \tag{3.3.9}$$

$$\begin{gathered}\mathrm{rank}\Big\{\mathbb{E}(\overline{Y}|X)\Big(I - \big(Q(X)\big)^{-}Q(X)\Big)\Big\} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)\big) - \mathrm{rank}\big(Q(X)\big), \\ \textit{for some g-inverse } \big(Q(X)\big)^{-};\end{gathered} \tag{3.3.10}$$

$$Q(X) = Q(X)\big(\mathbb{E}(\overline{Y}|X)\big)^{-}\mathbb{E}(\overline{Y}|X), \textit{ for some g-inverse } \big(\mathbb{E}(\overline{Y}|X)\big)^{-}. \tag{3.3.11}$$

First of all, Proposition **3.3.2** is an extension to the estimability result by Seely (1977) who discusses the equivalent conditions for estimability of the linear combination of a parameter vector.

Furthermore, as discussed in Section 3.2, identification through the linear projector is different from identification through the conditional expectation. However, since the linear projector $\mathbf{P}_L(\overline{Y}|X)$ is the optimal prediction achieved by projecting $\overline{Y}$ on the subspace spanned by any linear function of X compared to $\mathbb{E}(\overline{Y}|X)$ which is the optimal prediction of $\overline{Y}$ on the subspace spanned by any function of X, both Proposition **3.3.2** and Proposi-

tion **3.3.3** are still valid if we substitute the linear projector $\mathbf{P}_L(\overline{Y}|X)$ for the conditional expectation $\mathbb{E}(\overline{Y}|X)$. Additionally, if Assumption **3.2.8** is not satisfied, i.e, the underlying distribution depends on nuisance parameter which is not separable from the identifiable parameters $\mathbb{E}(\overline{Y}|X)A$, the equivalent conditions in Proposition **3.3.3** are only sufficient for identification of $Q(X)A$.

Of the above identification conditions, (3.3.1) and (3.3.2) are the direct paraphrases of the definition of identification. The identification conditions based on the characteristics of linear subspaces in (3.3.2), (3.3.3) and (3.3.4) are quite straightforward because we just need to check whether any row of $Q(X)$ belongs to the subspace spanned by the rows of matrix $\mathbb{E}(\overline{Y}|X)$. If this the case, then $Q(X)A$ is identifiable. Otherwise, it cannot be identified. Additionally, the rank conditions in (3.3.6), (3.3.7), (3.3.8) and (3.3.9) provide alternative ways of identifying $Q(X)A$, which are convenient to use and can simplify the calculations in certain statistical models. Conditions (3.3.10) and (3.3.11) are the more general statements for the identification of $Q(X)A$ when either $\mathbb{E}(\overline{Y}|X)$ or $Q(X)$ has a deficient rank, which further demonstrates that our paper does not impose any restrictions on the rank of X and on that of its functions. All the equivalent statements are necessary and sufficient conditions for the identification of $Q(X)A$. In particular, if the column space of $\big(Q(X)\big)'$ is contained in that of $\big(\mathbb{E}(\overline{Y}|X)\big)'$, then $Q(X)A$ can be identified. In other words, each of the columns of $\big(Q(X)\big)'$ can be expressed as a linear combination of the columns of $\big(\mathbb{E}(\overline{Y}|X)\big)'$ if and only if $Q(X)A$ is identifiable.

As discussed in the first chapter, we can establish identification conditions with linear restrictions on parameter matrix A. Consider the restriction in the implicit form:

$$R(X)A = C_0 \tag{3.3.12}$$

where R is an $m \times G$ matrix (with $m \geq 1$) and $\mathrm{Im}(C_0) \subseteq \mathrm{Im}\big(R(X)\big)$. We make no rank assumption on $R(X)$ and the following theorem on the identification of $Q(X)A$ is straightforward.

Theorem 3.3.4 CONDITIONS FOR IDENTIFICATION OF LINEAR PARAMETERS IN IV REGRESSION WITH IMPLICIT LINEAR RESTRICTIONS. *Let $Q(X)$ and $R(X)$ be $q \times G$ and $m \times G$ real matrices, respectively,* $\mathrm{Im}(C_0) \subseteq \mathrm{Im}\big(R(X)\big)$, *and*

$$\widetilde{X} = \begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ R(X) \end{bmatrix}. \tag{3.3.13}$$

Suppose the implicit linear restriction (3.3.12) *holds, and let* $\bar{\Theta}_{0I}\big(R(X), C_0\big) = \{\theta \in$

$\Theta_0 : R(X)A(\theta) = C_0\}$, $\bar{A}_I\big(R(X), C_0\big) = \{Z \in \mathbb{R}^{G\times H} : Z = A(\theta),\ \theta \in \bar{\Theta}_{0I}\big(R(X), C_0\big)\}$. *Then, each one of the following equivalent conditions entails that* $Q(X)A$ *is* $\big(\mathbb{E}(\overline{Y}|X)A\big)$*-identifiable on* $\bar{\Theta}_{0I}\big(R(X), C_0\big)$*, for any* $\mathrm{Im}(C_0) \subseteq \mathrm{Im}\big(R(X)\big)$*:*

$$\ker(\widetilde{X}) \subseteq \ker\big(Q(X)\big)\,; \tag{3.3.14}$$

$$\mathrm{Im}\big(Q(X)'\big) \subseteq \mathrm{Im}(\widetilde{X}')\,; \tag{3.3.15}$$

$$Q(X) = B\widetilde{X}, \textit{ for some matrix } B\,; \tag{3.3.16}$$

$$\mathrm{rank}\begin{bmatrix} \widetilde{X} \\ Q(X) \end{bmatrix} = \mathrm{rank}(\widetilde{X})\,; \tag{3.3.17}$$

$$\mathrm{rank}\begin{bmatrix} \widetilde{X} \\ Q(X) + V_1\widetilde{X} \end{bmatrix} = \mathrm{rank}(\widetilde{X}), \textit{ for any } q \times (T+m) \textit{ matrix } V_1\,; \tag{3.3.18}$$

$$\mathrm{rank}\begin{bmatrix} \widetilde{X} + V_2Q(X) \\ Q(X) \end{bmatrix} = \mathrm{rank}(\widetilde{X}), \textit{ for any } (T+m) \times q \textit{ matrix } V_2\,; \tag{3.3.19}$$

$$\mathrm{rank}\begin{bmatrix} \widetilde{X} \\ SQ(X) \end{bmatrix} = \mathrm{rank}(\widetilde{X}), \textit{ for any matrix } S \textit{ such that } \mathrm{rank}\big(SQ(X)\big) = \mathrm{rank}\big(Q(X)\big)\,; \tag{3.3.20}$$

$$Q(X) = Q(X)\widetilde{X}^{-}\widetilde{X}, \textit{ for some g-inverse } \widetilde{X}^{-}\,; \tag{3.3.21}$$

$$\mathrm{rank}[\,(I_G - Q(X)^{-}Q(X))\,] = \mathrm{rank}(\widetilde{X}) - \mathrm{rank}\big(Q(X)\big), \textit{ for some g-inverse } Q(X)^{-}\,. \tag{3.3.22}$$

If furthermore Assumption **3.2.8** *holds and* $\bar{A}_I\big(R(X), C_0\big) = \{A \in \mathbb{R}^{G\times H} : R(X)A = C_0\}$*, each one of the conditions* (3.3.14) - (3.3.22) *is necessary for* $Q(X)A$ *to be* $\big(\mathbb{E}(\overline{Y}|X)A\big)$*-identifiable on* $\bar{\Theta}_{0I}\big(R(X), C_0\big)$.

Now consider linear restrictions in explicit form:

$$A = B_0 + CL \text{ for some } L \in \mathbb{R}^{l\times H} \tag{3.3.23}$$

where B_0 is some $G \times H$ vector and C is some $G \times l$ real matrix. This yields the following characterizations of identifiability for $Q(X)A$.

Theorem 3.3.5 CONDITIONS FOR IDENTIFICATION OF LINEAR PARAMETERS IN IV REGRESSION WITH EXPLICIT LINEAR RESTRICTIONS. *Let* Q *be a* $q \times G$ *real matrices. Suppose the explicit linear restriction* (3.3.23) *holds, and let* $\bar{\Theta}_{0E}(C, B_0) = \{\theta \in \Theta_0 : A(\theta) = B_0 + CL \textit{ for some } L \in \mathbb{R}^{l\times H}\}$, $\bar{A}_E(C, B_0) = \{A \in \mathbb{R}^{G\times H} : A = A(\theta),\ \theta \in \bar{\Theta}_{0E}(C, B_0)\}$. *Then, each one of the following equivalent conditions entails that* $Q(X)A$

is $\big(\mathbb{E}(\overline{Y}|X)A\big)$*-identifiable on* $\bar{\Theta}_{0E}(C,B_0)$*, for any* $B_0 \in \mathbb{R}^{G\times H}$*:*

$$\ker\big(\mathbb{E}(\overline{Y}|X)C\big) \subseteq \ker\big(Q(X)C\big); \tag{3.3.24}$$

$$\mathrm{Im}(C'Q(X)') \subseteq \mathrm{Im}\big(C'(\mathbb{E}(\overline{Y}|X))'\big); \tag{3.3.25}$$

$$Q(X)C = D\big(\mathbb{E}(\overline{Y}|X)\big)C, \textit{ for some matrix } D; \tag{3.3.26}$$

$$\mathrm{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X)C \\ Q(X)C \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)C\big); \tag{3.3.27}$$

$$\mathrm{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X)C \\ Q(X)C + V_1\big(\mathbb{E}(\overline{Y}|X)\big)C \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)C\big), \textit{ for any } q\times T \textit{ matrix } V_1; \tag{3.3.28}$$

$$\mathrm{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X)C + V_2Q(X)C \\ Q(X)C \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)C\big), \textit{ for any } T\times q \textit{ matrix } V_2; \tag{3.3.29}$$

$$\mathrm{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X)C \\ SQ(X)C \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)C\big), \textit{ for any matrix } S \textit{ such that } \mathrm{rank}\big(SQ(X)C\big) = \mathrm{rank}\big(Q(X)C\big); \tag{3.3.30}$$

$$Q(X)C = Q(X)C\big(\mathbb{E}(\overline{Y}|X)C\big)^{-}\mathbb{E}(\overline{Y}|X)C, \textit{ for some g-inverse } \big(\mathbb{E}(\overline{Y}|X)C\big)^{-}; \tag{3.3.31}$$

$$\mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)C[I_l - (Q(X)C)^{-}Q(X)C]\big) = \mathrm{rank}\big(\mathbb{E}(\overline{Y}|X)C\big) - \mathrm{rank}(Q(X)C), \textit{ for some g-inverse } (Q(X)C)^{-}. \tag{3.3.32}$$

If furthermore Assumption **3.2.8** *holds and* $\bar{A}_E(C,B_0) = \{A \in \mathbb{R}^{G\times H} : A = B_0 + CL \textit{ for some } L \in \mathbb{R}^{l\times H}\}$*, each one of the conditions* (3.3.24) - (3.3.32) *is necessary for* $Q(X)A$ *to be* $\big(\mathbb{E}(\overline{Y}|X)A\big)$*-identifiable on* $\bar{\Theta}_{0E}(C,B_0)$*, for any* $B_0 \in \mathbb{R}^{G\times H}$*.*

A simple but useful application is the case when we are interested in identifying any arbitrary scalar parameter of A. Let's denote the i-th column of A as

$$\alpha_i = \begin{bmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{iG} \end{bmatrix}, \; i = 1,2,\ldots,H.$$

Without loss of generality, we choose to identify the first element α_{11}. Suppose we can partition Y as

$$Y = \begin{bmatrix} y & Y_2 \end{bmatrix},$$

where y is the first column of Y. Without loss of generality, choose

$$Q' = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Then the necessary and sufficient conditions for identification of $Q(X)\alpha_1 = \alpha_{11}$ by Proposition **3.3.3** is

$$\begin{aligned} \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} &= \operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y_2|X) \\ 1 & O \end{bmatrix} \\ &= 1 + \operatorname{rank}\Big(\mathbb{E}(Y_2|X)\Big) \\ &= \operatorname{rank}\Big(\mathbb{E}(Y|X)\Big). \end{aligned}$$

Equivalently, the necessary and sufficient condition for the identification of α_{11} is that conditional on X the expectation of the first column y does not belong to the subspace spanned by the remaining $G-1$ components of Y.

3.4. Applications

Due to the generality of (3.2.2), we can discuss identification of the structural parameters when the SEMs take different specific forms. In particular, an interesting empirical situation arises when the design matrix X of the reduced form equations does not have full column rank. Furthermore, it will be helpful if we can establish identification conditions for structural parameters even though the reduced form equation does not have a unique expression since the coefficient matrix of the endogenous variables is allowed to be singular. We will study identification issue for two scenarios, namely univariate structural equation and multivariate structural equations.

3.4.1. Univariate structural equation

We start with identification of the simple SEMs setup (3.1.1) and (3.1.2) where the structural form consists of a single equation. We relax the classical assumption that X has full column rank. Furthermore, the linearity assumption (3.1.2) can be relaxed and either linear or nonlinear specifications are allowed. Hence, any particular form of partially linear models can be part of the simultaneous equations system. However, we require (3.1.1) to be in

the linear form in order to employ the linearity property of conditional expectations. Let's denote

$$\overline{Y} = \begin{bmatrix} Y & X_1 \end{bmatrix}$$

and

$$\beta = \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}.$$

To simplify analysis, we do not impose a prior restrictions on the structural parameters. Suppose both the conditional means of u and v on X exist. Also assume Y is the endogenous variable and $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ consists of all exogenous instrumental variables. Then the SEMs can be constructed as a special case of (3.2.2). We have discussed briefly the standard rank and order conditions for identification of this setup when X has full column rank in the introduction. However, our intention herein is to demonstrate how the necessary and sufficient identification conditions can be applied to such a conventional SEMs setup. Taking expectations on both sides of (3.1.1) conditional on X we have

$$\mathbb{E}(y|X) - \mathbb{E}(u|X) = \mathbb{E}(\overline{Y}|X)\beta \equiv \mathbb{E}(Y|X)\beta_1 + X_1\gamma_1, \tag{3.4.1}$$

and

$$\begin{aligned} \mathbb{E}(Y|X) - \mathbb{E}(v|X) &= X\Pi \\ &= X_1\Pi_1 + X_2\Pi_2, \end{aligned} \tag{3.4.2}$$

which are identifiable for any realized value of X. Since we are not concerned about the identification of Π in (3.4.2) whether or not X has a full column rank has no impact on the analysis. In fact, only the existence of the conditional expectations of y and u will justify the application of Proposition **3.3.3**.

We are interested in identifying arbitrary linear combinations of β denoted as $Q_1(X)\beta_1 + Q_2(X)\gamma_1$. According to the general rank condition (3.3.6), the necessary and sufficient condition for identification of $Q_1(X)\beta_1 + Q_2(X)\gamma_1$ can be expressed as

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X_1 \\ Q_1(X) & Q_2(X) \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}. \tag{3.4.3}$$

Let's take a close look at the identification issue in this univariate structural equation setup for different situations. First, we want to identify the entire structural parameter in (3.1.1). It follows from (3.4.3) that the condition for identification of β is

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \\ I_G & O \\ O & I_{K_1} \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}, \tag{3.4.4}$$

which means that

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix} = G + K_1. \tag{3.4.5}$$

Hence, all the columns of $\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}$ are linearly independent. It follows from (3.4.2) where we assume $\mathbb{E}(v|X) = 0$ for simplicity that

$$\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}\begin{bmatrix} I_G \\ -\Pi_1 \end{bmatrix} = X_2\Pi_2.$$

In combination with (3.4.5), we have

$$\operatorname{rank}\begin{bmatrix} I_G \\ -\Pi_1 \end{bmatrix} = G = \operatorname{rank}(X_2\Pi_2).$$

Given the assumption that X has full column rank, it follows that

$$\operatorname{rank}(\Pi_2) = G \tag{3.4.6}$$

and it is the classical rank condition for identification. On the other hand, provided that X has full rank, it is trivial to show (3.4.6) implies (3.4.5). Since

$$\begin{aligned}(\operatorname{rank}(\Pi_2) = G) &\Rightarrow (\operatorname{rank}(X_2\Pi_2) = G) \\ &\Rightarrow \operatorname{rank}\left(\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}\begin{bmatrix} I_G \\ -\Pi_1 \end{bmatrix}\right) = G\end{aligned}$$

and

$$\operatorname{rank}\left(\mathbb{E}(Y|X)\right) = G,$$

the column subspaces of $\mathbb{E}(Y|X)$ and X_1 are essentially disjoint. Otherwise,

$$\operatorname{rank}\left(\mathbb{E}(Y|X)\right) < G.$$

Hence (3.4.5) holds.

Second, it is easy to examine if either β_1 or γ_1 is identifiable by checking the following rank conditions:

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \\ I_G & O \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}$$

and

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \\ O & I_{K_1} \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix},$$

which is equivalent to check whether

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix} = G + \operatorname{rank}(X_1) \tag{3.4.7}$$

and

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix} = K_1 + \operatorname{rank}(\mathbb{E}(Y|X)). \tag{3.4.8}$$

It then follows that (3.4.7) and (3.4.8) imply respectively

$$\operatorname{rank}(\mathbb{E}(Y|X)) = G \tag{3.4.9}$$

and

$$\operatorname{rank}(X_1) = K_1. \tag{3.4.10}$$

Note that both (3.4.9) and (3.4.10) are only necessary conditions. They become both necessary and sufficient conditions for identification when we further assume that the column subspaces of $\mathbb{E}(Y|X)$ and X_1 are essentially disjoint.

Third, we can check whether a subvector $\beta_{1,p}$ of β_1 is identifiable, where $\beta_{1,p}$ stands for the first p elements of β_1. It is then straightforward to choose the selection matrix

$$Q_{\beta,p} = \begin{bmatrix} I_p & O \end{bmatrix}$$

and verify whether the following condition holds

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ Q_{\beta,p} \end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(\overline{Y}|X)\big).$$

Similarly, the condition for identification of a subvector $\gamma_{1,q}$ of γ_1 can be given as

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ Q_{\beta,q} \end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(\overline{Y}|X)\big),$$

where

$$Q_{\beta,q} = \begin{bmatrix} O & I_q \end{bmatrix}.$$

Fourth, the identification of any linear combination of β denoted as $l'\beta$ can be checked by the following condition

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ l_1' \end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(\overline{Y}|X)\big).$$

In particular, it is handy to verify whether a scalar parameter of β can be identified. Without loss of generality, let's give the rank condition for the identification of the first element β_{11} of β_1 as

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ e_1' \end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(\overline{Y}|X)\big). \tag{3.4.11}$$

Clearly, (3.4.11) can be rewritten as

$$\operatorname{rank}\big(\mathbb{E}(\overline{Y}|X)\big) = \operatorname{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}_1|X) & \mathbb{E}(\overline{Y}_2|X) \\ 1 & O \end{bmatrix} = 1 + \operatorname{rank}\big(\mathbb{E}(\overline{Y}_2|X)\big),$$

where $\overline{Y}_1$ is the first column of $\overline{Y}$. Hence, the necessary and sufficient condition for identification of the first element of β_1 is that the the first column of $\mathbb{E}(\overline{Y}|X)$ does not belong to the subspace spanned by the columns of $\mathbb{E}(\overline{Y}_2|X)$; see Dufour and Liang (2012).

Remark 3.4.1 We can discuss the conditions for identification without the instruments X_2 and the results will be similar. Let's consider only (3.1.1) and denote

$$\overline{Y} \equiv \begin{bmatrix} Y & X_1 \end{bmatrix}.$$

Following the same steps, the conditional expectation of (3.1.1) on X_1 instead of X leads to

$$\mathbb{E}(y|X_1) - \mathbb{E}(u|X) = \mathbb{E}(\overline{Y}|X_1) \equiv \mathbb{E}(Y|X_1)\beta_1 + X_1\gamma_1.$$

Hence, the conditions for identification of β, any subvector and an arbitrary linear combination of β will be the same as above except that we condition all the random variables on X_1.

We have shown the generality of (3.4.3) by discussing identification of the whole structural parameter β, any subvector, any linear combination and any scalar parameter. Meanwhile, the classical results about identification can be readily generalized as special cases of Proposition **3.3.3**, which will be discussed hereinafter.

On the other hand, we can check identification of the structural parameters in a slightly different way by including the instrument X_2. Let's express (3.1.1) and (3.1.2) in the following compact matrix form:

$$\begin{bmatrix} y & Y \end{bmatrix} \begin{bmatrix} 1 & O \\ -\beta_1 & I_G \end{bmatrix} = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} \gamma_1 & \Pi_1 \\ O & \Pi_2 \end{bmatrix} + \begin{bmatrix} u & v \end{bmatrix}, \tag{3.4.12}$$

or

$$\widetilde{Y}\widetilde{B} = X\widetilde{\Gamma} + \widetilde{U}, \tag{3.4.13}$$

where

$$\widetilde{Y} = \begin{bmatrix} y & Y \end{bmatrix}, \widetilde{U} = \begin{bmatrix} u & v \end{bmatrix},$$
$$\widetilde{B} = \begin{bmatrix} 1 & O \\ -\beta_1 & I_G \end{bmatrix}, \widetilde{\Gamma} = \begin{bmatrix} \gamma_1 & \Pi_1 \\ O & \Pi_2 \end{bmatrix}.$$

Obviously, $\widetilde{B}$ is nonsingular. Rewrite (3.4.12) and (3.4.13) in the structural forms:

$$\begin{bmatrix} y & Y & X_1 & X_2 \end{bmatrix} \begin{bmatrix} 1 & O \\ -\beta_1 & I_G \\ -\gamma_1 & -\Pi_1 \\ O & -\Pi_2 \end{bmatrix} = \begin{bmatrix} u & v \end{bmatrix}$$

and

$$\begin{bmatrix} \widetilde{Y} & X \end{bmatrix} \begin{bmatrix} \widetilde{B} \\ -\widetilde{\Gamma} \end{bmatrix} = \widetilde{U}. \tag{3.4.14}$$

Again taking conditional expectation on both sides of (3.4.14), we have

$$\begin{bmatrix} \mathbb{E}(\widetilde{Y}|X) & X \end{bmatrix} \begin{bmatrix} \widetilde{B} \\ -\widetilde{\Gamma} \end{bmatrix} = \mathbb{E}(\widetilde{U}|X).$$

It then follows that the condition for identification of $Q_1(X)\beta_1 + Q_2(X)\gamma_1$ is

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \\ O & -Q_1(X) & -Q_2(X) & O \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix}. \tag{3.4.15}$$

We emphasize that (3.4.3) can generalize (3.4.15) since they are only equivalent if we impose further assumption on the structure of (3.4.14). In contrast, the rank condition (3.4.3) holds despite the relationships among the columns of $\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix}$. That is to say, without such additional assumptions, (3.4.15) may not be applicable but (3.4.3) still holds. This is due to the fact that the setup of (3.4.14) implicitly assumes that the reduced form is unique, i.e., $\widetilde{B}$ is invertible while the validity of (3.4.3) does not rely on such an assumption. If the column subspaces of $\begin{bmatrix} \mathbb{E}(y|X) & X_2 \end{bmatrix}$ and $\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}$ are essentially disjoint, it follows that

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \\ O & -Q_1(X) & -Q_2(X) & O \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & X_2 \\ O & O \end{bmatrix} + \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \\ -Q_1(X) & -Q_2(X) \end{bmatrix}$$

and

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & X_2 \end{bmatrix} + \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}.$$

Hence (3.4.15) and (3.4.3) indeed equivalent. Without such an assumption, (3.4.15) is a special case of (3.4.3) as demonstrated by the following example.

Let's consider identification conditions in the following situations. First, the condition for identification of β can be written as

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \\ O & I_G & O & O \\ O & O & I_{K_1} & O \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix}. \quad (3.4.16)$$

That is to say,

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix} = G + K_1 + \operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & X_2 \end{bmatrix}.$$

Remark 3.4.2 If $\mathbb{E}(y|X)$ belongs to the column space of $\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix}$ but cannot be spanned by the column space of X_2, we cannot identify β using (3.4.16) since

$$RHS = \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix} \leq \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix} + \operatorname{rank}(X_2) \leq G + K_1 + \operatorname{rank}(X_2)$$

and

$$LHS = G + K_1 + \operatorname{rank}\begin{bmatrix} \mathbb{E}(y|X) & X_2 \end{bmatrix} = G + K_1 + 1 + \operatorname{rank}(X_2) > G + K_1 + \operatorname{rank}(X_2).$$

Thus we have contradiction and (3.4.16) will not apply. However, (3.4.4) is still valid and so is (3.4.3).

Second, we provide the condition for identification of β_1 as

$$\text{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \\ O & I_G & O & O \end{bmatrix} = \text{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix},$$

which means

$$\text{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix} = G + \text{rank}\begin{bmatrix} \mathbb{E}(y|X) & X_1 & X_2 \end{bmatrix}. \tag{3.4.17}$$

Likewise, the condition for identification of γ_1 is

$$\text{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_1 & X_2 \end{bmatrix} = K_1 + \text{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(Y|X) & X_2 \end{bmatrix}. \tag{3.4.18}$$

Third, we check identification of any scalar parameter by verifying the following rank condition

$$\text{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(\overline{Y}|X) & X_2 \\ O & e_i' & O \end{bmatrix} = \text{rank}\begin{bmatrix} \mathbb{E}(y|X) & \mathbb{E}(\overline{Y}|X) & X_2 \end{bmatrix}. \tag{3.4.19}$$

3.4.2. Multivariate structural equations

Consider a system of G simultaneous equations in the standard structural form:

$$YB + X\Gamma = U, \tag{3.4.20}$$

where Y is a $T \times G$ matrix of endogenous variables, X is a $T \times K$ random matrix of exogenous variables, B and Γ are $G \times G$ and $K \times G$ matrices of unknown coefficients. Assume (3.4.20) satisfies the following regularity conditions:

$$\mathbb{E}(U|X) \text{ exists;} \tag{3.4.21}$$

$$\text{Var}(U_t|X) = \Sigma \text{ exists and is nonsingular, where } U' = \begin{bmatrix} U_1 & U_2 & \dots & U_T \end{bmatrix}; \tag{3.4.22}$$

$$\mathbb{E}(X'U) = 0; \tag{3.4.23}$$

the distribution of U can be fully characterized through its first two moments. (3.4.24)

We take a closer look at the above regularity assumptions. First, there is no restriction on the rank of X. It can be deficient in rank compared to the standard setup of the SEMs where X is assumed to have full column rank. Second, we impose no assumptions on the coefficient matrix B. It can be singular or a non-square matrix with or without full rank. Thus (3.4.20) allows for the situation when we do not have a unique expression for the reduce form. Third, we allow for the case where the distribution of the structural shocks can have a mean-variance structure that is not separable, i.e., the variance can be functionally dependent on the mean. This extends the analysis of identification in the classical literature where the distribution is either normal or has a separate mean-variance structure; see Rothenberg (1971) and Richmond (1974).

Meanwhile, we add a system of linear restrictions about the structural parameters

$$C_1 B + C_2 \Gamma = D, \tag{3.4.25}$$

where C_1, C_2 and D are $c \times G$, $c \times K$ and $c \times G$ know matrices. Taking conditional expectation on both sides of (3.4.20) gives

$$\mathbb{E}(Y|X)B + X\Gamma = \mathbb{E}(U|X). \tag{3.4.26}$$

We rewrite (3.4.25) and (3.4.26) in the matrix form

$$\begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix} \begin{bmatrix} B \\ \Gamma \end{bmatrix} = \begin{bmatrix} \mathbb{E}(U|X) \\ D \end{bmatrix}. \tag{3.4.27}$$

It follows that the LHS of (3.4.27) is identifiable due to construction. Therefore we can apply (3.3.6) in Proposition **3.3.3** and state that the necessary and sufficient condition for identification of arbitrary linear combinations $Q_1(X)B + Q_2(X)\Gamma$ is

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \\ Q_1(X) & Q_2(X) \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix} \tag{3.4.28}$$

Remark 3.4.3 Since we do not impose restrictions on the rank of either X or B, (3.4.28) is valid despite the linear relationships among the columns of X and B.

To demonstrate the generality of (3.4.28), let's consider the following cases. First, we

can check the identification of $\begin{bmatrix} B \\ \Gamma \end{bmatrix}$ by choosing the selection matrix I_{G+K}. The rank condition for identification of $\begin{bmatrix} B \\ \Gamma \end{bmatrix}$ is

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \\ I_G & O \\ O & I_K \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix}, \tag{3.4.29}$$

which entails that

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix} = G+K.$$

This coincides with the requirement for a unique solution $\begin{bmatrix} B \\ \Gamma \end{bmatrix}$ to (3.4.27), i.e., $\begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix}$ has full column rank.

Second, it is straightforward to check the identification of either B or Γ by verifying their respective rank conditions as follows:

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \\ I_G & O \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix} \tag{3.4.30}$$

and

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \\ O & I_K \end{bmatrix} = \operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix}. \tag{3.4.31}$$

Both (3.4.30) and (3.4.31) can be further simplified by

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix} = G + \operatorname{rank} \begin{bmatrix} X \\ C_2 \end{bmatrix} \tag{3.4.32}$$

$$\operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix} = K + \operatorname{rank} \begin{bmatrix} \mathbb{E}(Y|X) \\ C_1 \end{bmatrix}. \tag{3.4.33}$$

Since

$$\text{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix} \leq \text{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ C_1 \end{bmatrix} + \text{rank}\begin{bmatrix} X \\ C_2 \end{bmatrix},$$

it follows from (3.4.32) that

$$\text{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ C_1 \end{bmatrix} = G. \tag{3.4.34}$$

Likewise, (3.4.33) entails

$$\text{rank}\begin{bmatrix} X \\ C_2 \end{bmatrix} = K. \tag{3.4.35}$$

We herein point out that both (3.4.34) and (3.4.35) become only necessary for identification of B and Γ respectively. They are both necessary and sufficient unless the column subspaces of $\begin{bmatrix} \mathbb{E}(Y|X) \\ C_1 \end{bmatrix}$ and $\begin{bmatrix} X \\ C_2 \end{bmatrix}$ are essentially disjoint; see Harville (2008, Theorem 17.2.4).

Third, it is also convenient to check identification of any row of $\begin{bmatrix} B \\ \Gamma \end{bmatrix}$. Without loss of generality, let's consider the first row. Then the necessary and sufficient condition for identification becomes

$$\text{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \\ e_1' & O \end{bmatrix} = \text{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X \\ C_1 & C_2 \end{bmatrix}. \tag{3.4.36}$$

As mentioned earlier, the standard results on identification of the SEMs in the literature are only restricted to the setup where the exogenous matrix X has full column rank so that there is a unique solution to the reduced form equation. In contrast, we have demonstrated that the general rank condition (3.3.6) or (3.4.28) in case of the multivariate structural equation is valid without imposing assumption on the rank of X. Furthermore, whether or not we are able to have a unique reduced form will have no influence on the validity of our rank condition and its equivalence in Proposition **3.3.3**. Therefore, it is quite easy to check identification of any possible combinations of the structural parameters.

3.4.3. Generalization of standard results

Identification in the SEMs has been studied extensively in the classical literature; Fisher (1966), Rothenberg (1971) and Richmond (1974). We herein only focus on the conditions for identification in the linear framework of SEMs and leave the discussions of identification for nonlinear SEMs in Dufour and Liang (2013*b*).

Let's first generalize the standard conditions for identification in the univariate structural equation (3.1.1) and (3.1.2). Assume X is exogenous and has full column rank. Suppose we want to check identification of β_1. Let's construct a selection matrix Q with dimension $G \times (G + K_1)$ and partition it as follows:

$$Q = \begin{bmatrix} I_G & O \end{bmatrix}.$$

Then

$$Q\beta = \begin{bmatrix} I_G & O \end{bmatrix} \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix} = \beta_1.$$

From (3.3.6), it follows that the necessary and sufficient condition for identification of β_1 is

$$\text{rank} \begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ Q \end{bmatrix} = \text{rank}\big(\mathbb{E}(\overline{Y}|X)\big). \tag{3.4.37}$$

Since

$$\begin{aligned} \text{rank} \begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ Q \end{bmatrix} &\equiv \text{rank} \begin{bmatrix} \mathbb{E}(Y|X) & X_1 \\ I_G & O \end{bmatrix} \\ &= \text{rank} \begin{bmatrix} X\Pi & X_1 \\ I_G & O \end{bmatrix} \\ &= G + \text{rank}(X_1) \\ &= G + K_1, \end{aligned}$$

(3.4.37) means that

$$\text{rank}\big(\mathbb{E}(\overline{Y}|X)\big) = G + K_1. \tag{3.4.38}$$

Meanwhile, plugging (3.4.2) into (3.4.1) gives

$$\begin{aligned} \mathbb{E}(y|X) &= X\Pi\beta_1 + X_1\gamma_1 \\ &= \begin{bmatrix} X\Pi & X_1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix} \\ &= \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} \Pi_1 & I_{K_1} \\ \Pi_2 & O \end{bmatrix} \begin{bmatrix} \beta_1 \\ \gamma_1 \end{bmatrix}. \end{aligned} \tag{3.4.39}$$

Provided that $\text{rank}(\Pi_2) = G$, it follows that

$$\begin{aligned}
\text{rank}\big(\mathbb{E}(\overline{Y}|X)\big) &\equiv \text{rank}\begin{bmatrix} \mathbb{E}(Y|X) & X_1 \end{bmatrix} \\
&= \text{rank}\begin{bmatrix} X\Pi & X_1 \end{bmatrix} \\
&= \text{rank}\left(\begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} \Pi_1 & I_{K_1} \\ \Pi_2 & O \end{bmatrix} \right) \\
&= \text{rank}\begin{bmatrix} \Pi_1 & I_{K_1} \\ \Pi_2 & O \end{bmatrix} \\
&= G + K_1. \qquad (3.4.40)
\end{aligned}$$

Since the matrix $\begin{bmatrix} X\Pi & X_1 \end{bmatrix}$ has full column rank $G+K_1$, the structural parameters β_1 (and thus γ_1) are identifiable. Under the assumption of X having full column rank, it is clear that

$$\Big(\text{rank}\begin{bmatrix} X\Pi & X_1 \end{bmatrix} = G + K_1\Big) \iff \Big(\text{rank}(\Pi_2) = G\Big). \qquad (3.4.41)$$

Note that the second part of (3.4.41) is the standard rank condition which is a special case of (3.4.37) when X is assumed to have full column rank. On top of that, due to the structure of (3.1.1) and (3.1.2), γ_1 is identifiable only after β_1 is identified. However, we can identify both β_1 and γ_1 simultaneously rather than sequentially by applying (3.3.6). Specifically, let's choose

$$Q = I_{G+K_1}.$$

Since $\mathbb{E}(\overline{Y}|X)\beta$ is identifiable, we can check whether the following rank condition holds

$$\text{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ I_{G+K_1} \end{bmatrix} = \text{rank}\big(\mathbb{E}(\overline{Y}|X)\big). \qquad (3.4.42)$$

Clearly

$$\text{rank}\begin{bmatrix} \mathbb{E}(\overline{Y}|X) \\ I_{G+K_1} \end{bmatrix} = G + K_1.$$

Now the question reduces to checking whether

$$\text{rank}\big(\mathbb{E}(\overline{Y}|X)\big) = G + K_1.$$

This coincides with the necessary and sufficient identification condition discussed in

(3.4.41).

At first glance, our general rank condition (3.4.42) seems to be the same as the standard one. Nonetheless, there is a crucial difference between them. The validation of (3.4.42) does not depend on the assumption that the reduced form design matrix X has full column rank. Without such an assumption, the standard rank condition for identification of β_1 (and thus γ_1) fails but (3.4.42) still holds. On the other hand, let's suppose either of the rank conditions in (3.4.41) does not hold. Then we cannot identify β_1 and neither can we identify γ_1 but it is straightforward to check identification of an arbitrary linear combination of β as shown in subsection 3.4.1.

Next, we look at the standard setup (3.4.20). In comparison to the standard identification results, we need to make additional assumptions besides the ones in subsection 3.4.2. Thus we assume the following regularity conditions:

$$\mathbb{E}(U) = 0; \tag{3.4.43}$$

$$\text{Var}(U_t) = \Sigma \text{ exists and is nonsingular, where } U' = \begin{bmatrix} U_1 & U_2 & \dots & U_T \end{bmatrix}; \tag{3.4.44}$$

$$X \text{ is exogenous and has full column rank;} \tag{3.4.45}$$

$$\mathbb{E}(X'U) = 0; \tag{3.4.46}$$

$$B \text{ is nonsingular;} \tag{3.4.47}$$

$$\text{the distribution of the error terms can be fully characterized through its first two moments.} \tag{3.4.48}$$

We emphasize herein that X is assumed to be exogenous is for simplicity. Otherwise, we can condition on all related moments on X. Besides, we simplify the notation by assuming (3.4.43). In fact, the existence of $\mathbb{E}(U)$ suffices. Moreover, the normality assumption about the distribution of the error term is not required in our paper compared to Rothenberg (1971) who imposes normal distribution with a separate mean-variance structure. However, we do require that B and Γ are independent of Σ so that if assumption (3.4.48) is satisfied the identification of B and Γ implies that of Σ but the reverse is not true; see Bekker and Wansbeek (2001). As for the univariate random variable with normal distribution, the identification of one moment entails that of another and vice versa; see Dufour and Liang (2012). In addition, without assumption (3.4.48), the general rank condition (3.3.6) and its equivalences are only sufficient and not necessary since the structural parameters may be identified through higher order moments or other statistical properties rather than the first two moments.

Since B is nonsingular, we obtain the reduced form:

$$Y = X\Pi + V, \tag{3.4.49}$$

where

$$\Pi = -\Gamma B^{-1} \tag{3.4.50}$$

and $V = UB^{-1}$. For illustration purpose, we rewrite the above structural form and reduced form of the SEMs by observing that:

$$B'Y_t + \Gamma' X_t = U_t, \quad t = 1, 2, \ldots, T \tag{3.4.51}$$

and

$$Y_t = \Pi' X_t + V_t, \quad t = 1, 2, \ldots, T, \tag{3.4.52}$$

where $V_t = \left(B'\right)^{-1} U_t$ and Y_t, X_t and V_t are respectively the t-th row of Y, X and V. Since the random vector U_t is assumed to have a zero mean and a positive definite covariance matrix Σ, we have

$$\mathbb{E}(V_t) = 0$$

and

$$\mathrm{Var}(V_t) \equiv \Omega = \left(B^{-1}\right)' \Sigma \left(B^{-1}\right). \tag{3.4.53}$$

Denote the $G^2 + GK$ structural parameters as

$$\eta = \begin{pmatrix} \mathrm{vec}(B) \\ \mathrm{vec}(\Gamma) \end{pmatrix}$$

and assume the M linear restrictions take the form

$$\Phi \eta = d \tag{3.4.54}$$

which does not depend on Σ. Let's rewrite (3.4.50) as

$$\psi(\eta) \equiv \begin{bmatrix} \Pi & I_K \end{bmatrix} \begin{bmatrix} B \\ \Gamma \end{bmatrix} = 0. \tag{3.4.55}$$

and denote η_0 such that $\psi(\eta_0)$ and $\Omega(\eta_0)$ satisfy (3.4.55) and (3.4.53).

Remark 3.4.4 (3.4.54) and (3.4.55) are both identifiable and consist of a linear system of equations in η. This becomes clear if we take partial derivatives of $\psi(\eta)$ with respect to η

and the resulting linear operator can be denoted as

$$\Psi \equiv \left[\begin{array}{cc} I_G \otimes \Pi & I_{GK} \end{array} \right]$$

Therefore

$$\psi(\eta) = \Psi\eta$$

and the SEMs are a linear system of equations

$$\left[\begin{array}{c} \Psi \\ \Phi \end{array} \right] \eta = \left[\begin{array}{c} O \\ d \end{array} \right].$$

To identify the $G^2 + GK$ vector η, we choose the selection matrix $Q(X)$ in (3.3.6) to be the identity matrix I_{G^2+GK}. It follows that the necessary and sufficient condition for identification of η becomes

$$\text{rank} \left[\begin{array}{c} \Psi \\ \Phi \\ I_{G^2+GK} \end{array} \right] = \text{rank} \left[\begin{array}{c} \Psi \\ \Phi \end{array} \right], \tag{3.4.56}$$

which is equivalent to

$$\text{rank} \left[\begin{array}{c} \Psi \\ \Phi \end{array} \right] = G^2 + GK. \tag{3.4.57}$$

Note that Richmond (1974, Theorem 5) reaches the same conclusion as (3.4.57) which is just a special case of (3.3.6).

Let's denote the following $(GK + M) \times (G^2 + GK)$ matrix as

$$J \equiv \left[\begin{array}{c} \Psi \\ \Phi \end{array} \right]$$

and $J_{(1)}$ as the submatrix of J by dropping off its first column. Then it is easy to check identification of the i-th scalar parameter η_i by verifying the following condition

$$\text{rank} \left[\begin{array}{c} J \\ e_i' \end{array} \right] = \text{rank}(J).$$

Without loss of generality, let's consider identification of the first element η_1 by checking

$$\text{rank} \left[\begin{array}{c} J \\ e_1' \end{array} \right] = \text{rank}(J). \tag{3.4.58}$$

Since

$$\text{rank}\begin{bmatrix} J \\ e_1' \end{bmatrix} = \text{rank}\begin{bmatrix} O & J_{(1)} \\ 1 & O \end{bmatrix} = 1 + \text{rank}(J_{(1)}),$$

(3.4.58) leads to

$$\text{rank}(J) = 1 + \text{rank}(J_{(1)}), \tag{3.4.59}$$

which is the same conclusion by Richmond (1974, Theorem 3). It is noticeable that the standard results such as (3.4.57) and (3.4.59) are based on the linear SEMs setup. In fact, the linear form of both (3.4.54) and (3.4.55) can be generalized to nonlinear specifications. If we relax (3.4.48) or the assumption that the mean-variance structure is separate or the linear restrictions (3.4.54) does not depend on the covariance matrix Σ, identification of the structural parameters will also rely on the second moment of the reduced shocks which becomes a nonlinear function of η. In the presence of nonlinearity the standard rank conditions will fail. However, (3.3.6) and other statements in Proposition **3.3.3** are still valid for checking identification of nonlinear SEMs; see Dufour and Liang (2013*b*).

Applying (3.3.6), we can also derive identification condition for an arbitrary structural equation of (3.4.20). Let $A = \begin{bmatrix} B' & \Gamma' \end{bmatrix}$ and denote $A_i{}'$ the i-th row of A, $i = 1, 2, \ldots, G$. Suppose the linear restrictions are

$$\phi_i A_i = 0, \ , i = 1, 2, \ldots, G,$$

where ϕ_i is a $M_i \times (G+K)$ constant matrix.

Clearly if empirical studies are interested in identification of some linear combination $l'\eta$, it is straightforward to write the rank condition as

$$\text{rank}\begin{bmatrix} J \\ l' \end{bmatrix} = \text{rank}(J). \tag{3.4.60}$$

3.5. Conclusion

We study the necessary and sufficient conditions for the SEMs with particular attention paid to the linear system, where both the structural equations and the prior restrictions are linear in the structural parameters. We introduce a general framework for the system of simultaneous equations which provides the flexibility of including a variety of the SEMs setups commonly encountered in the literature, such as Fisher (1966), Rothenberg (1971), Richmond (1974) and Bekker and Wansbeek (2001), etc. In contrast to the classical SEMs,

we relax a series of assumptions on the usual model specifications so that the standard identification conditions can be easily generalized as special cases of our rank condition and its equivalences. First of all, the design matrix in the reduced form may not have full column rank. Since we are concerned about the identification of structural parameters but the full rank constraint is only conductive to identifying the reduced form parameters and is not related to the identification of structural parameters, relaxation of this assumption does not cause direct effect on identification of deep parameters. With the help of additional prior information, it is possible to achieve a complete identification of the whole structural model. Otherwise, we are still able to identify some linear combinations of parameters provided that the structure of the SEMs under consideration satisfies certain identification conditions we have proposed, i.e., parametric function identification is feasible when the design matrix is deficient in rank. Furthermore, the coefficient matrix of the endogenous variables can be singular. Therefore, our general rank condition is applicable even if there does not exist a unique reduced form of the SEMs. Indeed, since we focus directly on the structural equations, the validity of the general identification conditions holds despite the existence of a unique reduced form. Besides, we relax the assumption that the design matrix is either exogenous or fixed. Thus all endogenous variables are now conditioned on the instruments and identification conditions now rely on the realized values of these instrumental variables. Note that we allow for some of the instruments to be endogenous and this has an influential impact on the study of instrumental variables since it is always a challenging issue to deal with missing instruments and weak instruments and how to select the strong instruments from a large set of candidates is still open. Moreover, our SEMs setup takes into account the presences of nuisance parameters and their role in identification. As argued by Dufour and Liang (2012), the necessary and sufficient conditions are valid as long as the nuisance parameters can be completely separated from the parameters of interest (i.e., structural parameters or deep parameters). Otherwise, the identification conditions established therein are only sufficient and not necessary. What needs to be mentioned is that the standard normality assumption on the distribution of the structural shocks implies a separate mean-variance structure. Thus the classical identification conditions are special cases of our results. Additionally, we demonstrate the generality of proposed identification conditions by considering two types of SEMs setups: univariate structural equation and multivariate structural equations, and provide conditions for identification of the entire structural parameter vector, a subvector or any arbitrary scalar element of the structural parameter vector and any linear combination of its components.

The identification conditions established are global in the sense that they can check identification of the structural parameters over the entire parameter space or constrained

space under some prior information. If the SEMs are nonlinear in the structural parameters, we generally can only obtain local identification within some neighborhoods of given parameter values. However, it is still possible to achieve global identification for nonlinear models inclusive of nonlinear SEMs using other techniques based on global univalence theorems. These interesting issues will be discussed as separate topics in our following papers.

Chapter 4

Necessary and sufficient conditions for nonlinear parametric function identification

Jean-Marie Dufour and Xin Liang

Abstract We study identification conditions of a general nonlinear function $\beta(\theta)$ in terms of another nonlinear parametric function $\gamma(\theta)$ which is identifiable in nonlinear models and call it "parametric function identification". The setup is fundamentally semiparametric, and the basic assumption is that structural parameters of interest determine a number of identifiable parameters through a nonlinear equation (such as a conditional or unconditional moment equation). Such models are quite common in econometrics, and include for example nonlinear models typically estimated by GMM, and dynamic stochastic general equilibrium (DSGE) models (used in macroeconomics and finance). This paper considers the general case where not all model parameters are identifiable, with the purpose of characterizing nonlinear parameter transformations which are identifiable. The literature on this problem is very thin, and only deals with the identification of the full parameter vector in the equation of interest. The contributions of this paper lies in the following aspects. First, we propose both necessary and sufficient conditions for local parametric function identifications when the classical assumption of full rank Jacobian matrix fails. It is noted that such necessary and sufficient conditions for parametric function identification are valid under rather weak assumptions and they hold with or without linear and nonlinear restrictions. The framework under investigation is quite general since $\gamma(\theta)$ is not limited to either unconditional moment equations or conditional moment equations, as studied in the

literature. Besides, we allow for parametric functions to have infinite dimensions although we focus on identification for parametric models. Second, we introduce to the literature the concept of local identification around a parameter value and distinguish it from the classical definition of local identification at a parameter value which is restrictive for statistical inference. This new concept will help achieve local parametric function identification in the restricted parameter space, which helps justify the existence of a consistent estimator. Third, we provide a group of equivalent conditions for parametric function identification that are intuitive and easy to check in practice. The useful rank condition for parametric function identification is just one of these equivalences. Fourth, we verify that the established identification conditions can be readily applied to the likelihood model setups through Kullback-Leibler divergence and local parametric function identification can be achieved by higher order moments. Fifth, this paper can help empirical research for identification since the general parametric function identification conditions proposed are easily applicable to the simultaneous models and the DSGE models. It shows that most of the classical rank identification conditions in the literature can be generalized as special cases.

4.1. Introduction

The earliest known study in English on identification problem is attributed to Wright (1915); see Stock and Trebbi (2003) who mention the fact that Lenoir (1913) appears to be the first to notice identification issue in French, which is noted by both Fox (1968) and Christ (1985). Although the important results of Wright (1915) have been discussed by Working (1927), identification has not become a heated topic in econometrics until it is rekindled by Koopmans and Reiersøl (1950) who formally study the identification issue through reformulating the specification problem. Some later influential work on identification with particular statistical model framework includes Fisher (1966) and Rothenberg (1971). Of course mathematicians and statisticians have a long history of interest in the unique solution to a given system; see Birkhoff (1934), Whyburn (1942) and Arens (1946) who analyze the properties of homeomorphism group in the topological space. If such a uniqueness does exist it is named (global) univalence. Parthasarathy (1983) gives a detailed background on univalence and its related theorems. Since homeomorphism is sufficient for injectivity and hence univalence, a large amount of fruitful work on global homeomorphism and local homeomorphism and the relationship between them is closely related and includes Church (1962), Church (1963), Gale and Nikaidô (1965), Plastock (1974), Kojima and Saigal (1979), Garcia and Zangwill (1979), Pourciau (1982). In essence, identification can be treated as a problem of univalence over the parameter space because if there is a

one-to-one function between the parameter θ of interest and the probability distribution P_θ, then θ is (globally) identifiable.

Generally speaking, identifiability of parameters should be treated independently of and prior to statistical inference issues as advocated by Manski (1995). Meaningful consistent estimators of linear parameters in statistical models entail identifiability of such parameters. Unless identification holds, any statistical inference derived and interpretations induced thereon could become unreliable or even misleading. Although theoretical verification of identifiability should always be conducted before we perform statistical inference, it is often difficult to do so, especially for nonlinear regression models such as simultaneous equations models (SEMs). Therefore, the literature has to assume identifiability in most cases. Another reason of performing estimation first may be that sometimes the interval estimators can provide useful information about the identifiability of certain parameters. That is to say, if the confidence intervals are unbounded or cover the entire parameter space, then we can conclude that these parameters are not identifiable. Global identification of parameters could be achieved in linear regression models and usually we can obtain only local identification for nonlinear regression models; see the discussions in Rothenberg (1971). The most recent paper that discusses global identification in nonlinear models is by Komunjer (2012) who relaxes the classical Gale-Nikaidô-Fisher-Rothenberg conditions by assuming a nonnegative determinant of the Jacobian matrix. As a trade-off, other restrictions, such as properness and countably many solutions to a mapping at its critical values, have to be imposed to obtain sufficient conditions for global identification. On the other hand, an increasing number of literature focus on global or local identification in nonparametric and semiparametric models based on conditional moment restrictions, i.e., conditional on the instrumental variables; see Florens (2003), Newey and Powell (2003), Florens, Johanes and Van Bellegem (2012), Chernozhukov, Imbens and Newey (2007), Chen et al. (2011). Chen et al. (2011) state a sufficient condition for local identification which is close to one of the identification conditions we propose in the paper but they only consider sufficient condition for local identification at θ_0 through the rank of the Jacobian matrix of the moment conditions. Meanwhile, we recognize that global identification in the context of nonlinear models is quite challenging. Rothenberg (1971) gives a sufficient condition for global identification under strong assumptions such as multivariate normality, known covariance matrix and full rank information matrix. Our paper focuses on conditions for both global identification and local identification.

Another important issue studied in the paper is what we call "parametric function identification" or "function identification" for simplicity when it causes no ambiguity, which means that the parameter vector θ may not be identifiable but a function of θ can still be

identified. Surprisingly, the literature on this topic is quite scarce. For instance, Fisher (1966) considers identification of a single equation within the SEMs setups while Richmond (1974) concentrates on global identification of either the entire parameter vector θ or an element of it. But neither paper discusses identification of a general nonlinear function of θ. Meanwhile, Rothenberg (1971) studies identification of a scalar of θ but his results are limited to likelihood models [see also Bowden (1973)] and do not cover identification of nonlinear functions. To the best of our knowledge, we are the first to look at identification conditions of a general nonlinear function $\beta(\theta)$ in terms of another nonlinear function $\gamma(\theta)$ which is identifiable in nonlinear models. On one hand, the concept of "parametric function identification" extends that of partial identification in the sense of Phillips (1989) and Bekker and Wansbeek (2001) who are interested in identification of either a scalar element or a subvector of θ without mentioning identification of a function of θ. In contrast, we provide both necessary and sufficient conditions for identification of a general function of θ which can be linear or nonlinear in θ. This makes our identification results applicable to some well-known statistical and econometric models, such as the likelihood models, the simultaneous equations models, the DSGE models, etc. On the other hand, we use the term "parametric function identification" to distinguish it from the concept of set-valued identification region, which is also named partial identification by Manski (1995), Manski (2003) and Tamer (2010). Nevertheless, Manski (1995) employs partial identification to compare with point identification and does not mean partially identified models as Phillips (1989) and Bekker and Wansbeek (2001) do. In essence, we concern ourselves with point identification of a parametric function of θ rather than putting identification bounds on the parameters. Hence "parametric function identification" will help clarify the confusion caused by partial identification in the literature.

Empirically, identification issue has gained popularity in the literature especially when researchers try to derive reliable statistical inferences in the presence of weak instruments and weak identification. For instance, the New Keynesian Phillips curves (NKPC) have been studied extensively and particular attentions concentrate on the hybrid NKPC by Galí and Gertler (1999) and its variations; see Mavroeidis (2005), Rudd and Whelan (2006), Zhang, Osborn and Kim (2007), Tsoukis, Kapetanios and Pearlman (2011) and Nymoen, Swensen and Tveter (2012). Moreover, recent literature has also made progress in using identification robust methods to deal with weak identification problem in the NKPC; see Dufour, Khalaf and Kichian (2006), Nason and Smith (2008), Martins and Gabriel (2009), Dees, Pesaran, Smith and Smith (2009), Kleibergen and Mavroeidis (2009), etc. Nevertheless, there does not exist a simple and unified approach that can check identification of a function of deep parameters. This paper makes a breakthrough in examining differ-

ent types of identification issues under a general framework, i.e., we provide identification conditions to analyze both identification failure and weak identification with the flexibility to study any functional form (linear and nonlinear) of structural parameters.

Our paper makes contributions to the analysis of identification conditions in the following aspects. First, we establish both necessary and sufficient conditions for parametric function identification with and without linear or nonlinear restrictions. It will be interesting to achieve function identification rather than identification of the whole parameter vector because often identifying a subvector or a linear combination of parameter vector is more feasible and efficient when identification of each element of the parameter vector becomes impossible without further information about the underlying distributions. Note that our conditions for identification of functions of parameters include the local identification of the entire parameters vector as a special case. Second, we introduce to the literature a new concept of local identification around θ_0 and distinguish it from the classical definition of local identification at θ_0. Some extensions of the standard identification conditions have been made. Third, a group of intuitive equivalent necessary and sufficient conditions for parametric function identification are introduced using vector space properties since empirical researchers usually examine a statistical model through its explanatory variables. Therefore, our results are easier to verify in practice.

This paper is organized as follows. We give the classical definitions of global and local identifications for parameters of interest and related standard results with slight extensions in Section 4.2. Section 4.3 introduces the concept of local identification around a point and both necessary and sufficient conditions for local identification based on the modified inverse function theorem. Section 4.4 provides the local identification conditions for general parametric functions and alternative conclusions on local identification based on linear parameters. Section 4.5 investigates identification conditions in likelihood models. Section 4.6 applies the identification results in the context of the SEMs and the DSGE models. Section 4.7 demonstrates how to easily check identification failure and weak identification issues using our proposed general rank condition (and its equivalences) through real macroeconomic models. Section 4.8 concludes.

4.2. Framework

We start with the well-known definitions of observational equivalence, global identification and local identification at θ_0 as in the classical literature. Let $(\Omega, \mathscr{A}, \mathscr{P})$ be a statistical model, where Ω is a sample space, $\mathscr{A}$ is a σ-algebra of subsets of Ω, $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ is a family of probability measures on $(\Omega, \mathscr{A})$, the open connected set $\Theta \in \mathbb{R}^k$ is a topological

parameter space. Also let $\gamma : \Theta \to \Gamma$ be some parametric function of θ. For instance, we can take $\gamma(\theta)$ to be the moment equation as follows:

$$\gamma(\theta) = \mathbb{E}[g(Y_t;\theta)]$$

or

$$\gamma(\theta) = \mathbb{E}[g(Y_t;\theta)|Z_t].$$

Note that $\gamma(\theta)$ does not have to take the form of a moment condition as the literature usually assumes. In fact, it can be any nonlinear transformation of moments. Furthermore, the probability distribution of $g(Y_t;\theta)$ is allowed to depend on both θ and the nuisance parameter $\nu \in \Xi$ which can have an infinite dimension. In other words, ν can be any assumption regarding the distribution of observed variables – such as $g(Y_t;\theta)$ in the case of a moment equation– which is not completely identified through θ. Therefore, the distribution of observed variables is not necessarily specified up to a finite number of parameters and our framework covers semiparametric setups.

4.2.1. Definitions

We now define different types of identifications as follows. Suppose the nuisance parameter ν can be separated from the parameter of interest θ and denote the probability distribution as $\mathsf{P}_{\bar{\theta}} \equiv \mathsf{P}_{(\theta,\nu)}$, $\forall \nu \in \Xi$.

Definition 4.2.1 OBSERVATIONAL EQUIVALENCE. *If $\bar{\theta}_1 = (\theta_1, \nu_1)$ and $\bar{\theta}_2 = (\theta_2, \nu_2)$ satisfy the condition $\mathsf{P}_{\bar{\theta}_1} = \mathsf{P}_{\bar{\theta}_2}$,we say that $\bar{\theta}_1$ and $\bar{\theta}_2$ are* observationally equivalent.

Although observational equivalence has been defined in classical papers [see Koopmans and Reiersøl (1950) and Rothenberg (1971)], Definition **4.2.1** is the first in the literature to explicitly take into account nuisance parameters and therefore contains semiparametric frameworks.

Definition 4.2.2 GLOBAL IDENTIFICATION AT θ_0 AND OVER Θ. *The parameter θ is globally identifiable at $\theta_0 \in \Theta$ if and only if*

$$\left(\theta \neq \theta_0\right) \Rightarrow \left(\mathsf{P}_{(\theta,\nu_1)} \neq \mathsf{P}_{(\theta_0,\nu_2)}\right), \quad \forall \theta \in \Theta,\, \forall \nu_1,\, \nu_2 \in \Xi\,. \tag{4.2.1}$$

On the other hand, the parameter θ is said to be globally identifiable over Θ if and only if

$$\left(\theta_1 \neq \theta_2\right) \Rightarrow \left(\mathsf{P}_{(\theta_1,\nu_1)} \neq \mathsf{P}_{(\theta_2,\nu_2)}\right), \quad \forall \theta_1,\, \theta_2 \in \Theta,\, \forall \nu_1,\, \nu_2 \in \Xi\,. \tag{4.2.2}$$

Empirically, it is feasible to identify a parametric function of θ even if θ itself is not identifiable and we denote such a transformation $\gamma(\theta)$.

Definition 4.2.3 GLOBAL IDENTIFICATION OF PARAMETRIC FUNCTIONS AT θ_0 AND OVER Θ . *The parametric function $\gamma : \Theta \to \Gamma$ is globally identifiable at $\theta_0 \in \Theta$ if and only if*

$$\big(\gamma(\theta) \neq \gamma(\theta_0)\big) \Rightarrow \big(\mathsf{P}_{(\theta,\nu_1)} \neq \mathsf{P}_{(\theta_0,\nu_2)}\big), \quad \forall \theta \in \Theta, \forall \nu_1, \nu_2 \in \Xi . \tag{4.2.3}$$

On the other hand, the parametric function $\gamma : \Theta \to \Gamma$ is globally identifiable over Θ if and only if

$$\big(\gamma(\theta_1) \neq \gamma(\theta_2)\big) \Rightarrow \big(\mathsf{P}_{(\theta_1,\nu_1)} \neq \mathsf{P}_{(\theta_2,\nu_2)}\big), \quad \forall \theta_1, \theta_2 \in \Theta, \forall \nu_1, \nu_2 \in \Xi . \tag{4.2.4}$$

Definition **4.2.3** states that $\gamma(\theta)$ is globally identifiable if and only if we can empirically distinguish different values of $\gamma(\theta)$ from different probability distributions of the observed data. It separates itself from the standard global identification at θ_0 in the literature [see Gouriéroux and Monfort (1995)] by considering nuisance parameters in the framework. Moreover, it is convenient to generalize Definition **4.2.3** by studying identification of a parametric function of θ in terms of another identifiable parametric function, which we call parametric function identification.

Definition 4.2.4 GLOBAL IDENTIFICATION IN TERMS OF ANOTHER PARAMETRIC FUNCTION AT θ_0 AND OVER Θ. *Let $\beta : \Theta \to B$ and $\gamma : \Theta \to \Gamma$ be two parametric functions. $\beta(\theta)$ is globally identifiable in terms of $\gamma(\theta)$ at $\theta_0 \in \Theta$ if and only if*

$$\big(\beta(\theta) \neq \beta(\theta_0)\big) \Rightarrow \big(\gamma(\theta) \neq \gamma(\theta_0)\big), \quad \forall \theta \in \Theta. \tag{4.2.5}$$

On the other hand, $\beta(\theta)$ is globally identifiable in terms of $\gamma(\theta)$ over Θ if and only if

$$\big(\beta(\theta_1) \neq \beta(\theta_2)\big) \Rightarrow \big(\gamma(\theta_1) \neq \gamma(\theta_2)\big), \quad \forall \theta_1, \theta_2 \in \Theta. \tag{4.2.6}$$

Note that the statement in Definition **4.2.4** is equivalent to saying that $\beta(\theta)$ can be expressed as a transformation of $\gamma(\theta)$ in Θ, which entails that $\beta(\theta)$ can be globally identified as long as $\gamma(\theta)$ is globally identifiable. Similarly, we can define local identifications at θ_0 as follows.

Definition 4.2.5 LOCAL IDENTIFICATION OF PARAMETERS AT θ_0. *The parameter θ is locally identifiable at $\theta_0 \in \Theta$ if and only if there exists an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0*

such that

$$\left(\theta \neq \theta_0\right) \Rightarrow \left(\mathsf{P}_{(\theta,\nu_1)} \neq \mathsf{P}_{(\theta_0,\nu_2)}\right), \quad \forall \theta \in \mathscr{V}(\theta_0),\, \forall \nu_1,\, \nu_2 \in \Xi\,. \tag{4.2.7}$$

Definition 4.2.6 LOCAL IDENTIFICATION OF PARAMETRIC FUNCTIONS AT θ_0. *The parametric function $\gamma : \Theta \to \Gamma$ is locally identifiable at $\theta_0 \in \Theta$ if and only if there exists an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 such that*

$$\left(\gamma(\theta) \neq \gamma(\theta_0)\right) \Rightarrow \left(\mathsf{P}_{(\theta,\nu_1)} \neq \mathsf{P}_{(\theta_0,\nu_2)}\right), \quad \forall \theta \in \mathscr{V}(\theta_0),\, \forall \nu_1,\, \nu_2 \in \Xi\,. \tag{4.2.8}$$

Definition 4.2.7 LOCAL IDENTIFICATION IN TERMS OF ANOTHER PARAMETRIC FUNCTION AT θ_0. *Let $\beta : \Theta \to B$ and $\gamma : \Theta \to \Gamma$ be two parametric functions. $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ at $\theta_0 \in \Theta$ if and only if there exists an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 such that*

$$\left(\beta(\theta) \neq \beta(\theta_0)\right) \Rightarrow \left(\gamma(\theta) \neq \gamma(\theta_0)\right), \quad \forall \theta \in \mathscr{V}(\theta_0). \tag{4.2.9}$$

For instance, we can take $\gamma(\theta)$ and $\beta(\theta)$ to be the reduced form parameters and the structural parameters in the dynamic stochastic general equilibrium (DSGE) models. We want to retrieve useful information about $\beta(\theta)$ from the identifiable structural parameter $\gamma(\theta)$. That is to say, we need to check whether $\beta(\theta)$ depends on θ only through $\gamma(\theta)$.

We need the following definition to achieve both necessary and sufficient conditions for local identification; see Fisher (1966), Rothenberg (1971) and Shapiro (1986).

Definition 4.2.8 REGULAR POINT. *Let θ be a parameter vector in the parameter space $\Theta \subset \mathbb{R}^k$ and $M(\theta)$ a matrix whose elements are continuous functions of θ. Then θ_0 is a regular point of the matrix $M(\theta)$ if there is an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 such that $M(\theta)$ has the same rank for all $\theta \in \mathscr{V}(\theta_0)$.*

In fact, the above definitions are valid without specifying the dimension of the parameter space Θ. Nevertheless, we focus on the establishment of identification conditions through the rank of Jacobian matrices which requires that Θ has a finite dimension.

4.2.2. Standard results

We now provide the classical condition for global and local identifications at θ_0 [see Fisher (1966), Rothenberg (1971), Bowden (1973) and Bekker et al. (1994)] that can help establish more general identification conditions hereinafter, which will in turn lay a solid foundation

for our derivation of the necessary and sufficient local identification conditions for nonlinear models, such as the DSGE models. For reference purpose, we replicate some of them below. Rothenberg (1971) has established the well-known global and local identification conditions based on the likelihood function and the rank of the Fisher information matrix under a series of assumptions; Rothenberg (1971, Theorem 3).

Theorem 4.2.9 NECESSARY AND SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION. *Let θ_0 be a regular point of the Fisher information matrix. Then θ is locally identifiable at θ_0 if and only if the Fisher information matrix evaluated at θ_0 is nonsingular.*

Moreover, Rothenberg (1971) proposes conditions for identification of a scalar parameter of the simultaneous equations models (SEMs); see Rothenberg (1971, Theorem 8, Corollary 3). Suppose there exist M continuously differentiable restrictions on the structural parameters

$$\phi(\theta) = \mathbf{0} \tag{4.2.10}$$

and we denote the Jacobian matrix of ϕ as $J_\phi(\theta)$.

Theorem 4.2.10 NECESSARY AND SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION OF A SCALAR PARAMETER. *Suppose*

(1) *Θ is an open set in $\mathbb{R}^k$;*

(2) *the parameter space Θ is restricted to (*4.2.10*) and denoted by $\bar{\Theta}$;*

(3) *$f(y;\theta)$ depends on θ only through the g reduced form parameters $\eta = \rho(\theta)$ such that $f(y;\theta) = f^*(y;\eta)$ and ρ is continuously differentiable in Θ with Jacobian matrix $J_\rho(\theta)$;*

(4) *$\Xi \subseteq \mathbb{R}^g$ is the image of $\bar{\Theta}$ under ρ;*

(5) *every $\eta \in \Xi$ is globally identifiable.*

Let $Q(\theta) \equiv \begin{bmatrix} J_\rho(\theta) \\ J_\phi(\theta) \end{bmatrix}$ and $Q_1(\theta) \equiv \begin{bmatrix} Q(\theta) \\ e_1{}' \end{bmatrix}$, where e_1 is the unit vector with one as its first element and zero elsewhere. Let θ_0 is a regular point of of both $Q(\theta)$ and $Q_1(\theta)$. Also denote θ_1 and $\theta_{0,1}$ as the first element of θ and θ_0. Then the first element θ_1 is locally identifiable at $\theta_{0,1}$ if and only if $Q(\theta)$ and $Q_1(\theta)$ have the same rank.

Now consider a system of G structural equations:

$$YB + X\Gamma = U, \tag{4.2.11}$$

where Y is a $T \times G$ matrix of endogenous variables, X is a $T \times K$ matrix of exogenous variables, B and Γ are $G \times G$ and $K \times G$ matrices of unknown coefficients. Let U_t, $t = 1,2,\ldots,T$, be the element of U and assume it is normally distributed with zero mean and covariance matrix Ω.

Theorem 4.2.11 NECESSARY AND SUFFICIENT CONDITION FOR GLOBAL IDENTIFICATION OF A SINGLE EQUATION. *Let* $A \equiv \begin{bmatrix} B' & \Gamma' \end{bmatrix}$ *and* A_i *be the i-th column of* A', $i = 1,2,\ldots,G$. *Suppose there exist linear restrictions* M_i *on* A_i *such that* $\phi_i A_i = \mathbf{0}$, $i = 1,2,\ldots,G$, *where* ϕ_i *is a constant matrix. Then* A_i *is globally identifiable at the true value* $A_{0,i}$ *if and only if* $\operatorname{rank}(\phi_i A') = G$.

Bekker and Wansbeek (2001) restate the main conclusions of Rothenberg (1971) and generalize the identified parametric function to include higher order moments for nonnormal distributions. They also provide a different condition for local identification of a scalar parameter under similar assumptions of Theorem **4.2.10**; see Bekker and Wansbeek (2001, Theorem 9).

Theorem 4.2.12 ALTERNATIVE SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION OF A SCALAR PARAMETER. *Suppose the assumptions of Theorem* **4.2.10** *are satisfied. Let* $Q_{(i)}(\theta)$ *be the Jacobian matrix* $Q(\theta)$ *without the i-th column and* θ_0 *be a regular point of of both* $Q_{(i)}(\theta)$ *and* $Q(\theta)$. *If* $\operatorname{rank}\left(Q_{(i)}(\theta)\right) < \operatorname{rank}(Q(\theta))$, *then* θ_i *is locally identifiable at* $\theta_{0,i}$.

Recently, Chen et al. (2011) propose a sufficient condition for local identification at θ_0 using the rank of the Jacobian matrix of the moment conditions. For convenience, we state it as a corollary as follows.

Corollary 4.2.13 SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION WITH MOMENT CONDITIONS. *Let* $g(\theta) : \Theta \subseteq \mathbb{R}^k \mapsto \mathbb{R}^p$ *be moment functions. If* $g(\theta)$ *is differentiable at* θ_0 *and* $\operatorname{rank}[J_g(\theta_0)] = p$, *where* $J_g(\theta)$ *is the Jacobian matrix of* $g(\theta)$, *then* θ *is locally identifiable at* θ_0.

In comparison to the above mentioned classical identification conclusions, we herein propose somewhat different conditions for local identification at θ_0. Specifically, we do not limit a priori the locally identifiable function as the likelihood functions of the observations as in Rothenberg (1971) or the moment conditions as in Chen et al. (2011). Rather it can take the form of any locally identifiable functions such as the Kullback-Leibler divergence, the first order partial derivative of the Kullback-Leibler divergence, the auto-covariance function as in Komunjer and Ng (2011), the quantile function of the probability function

or any moment of the data which depends on the deep parameter θ. Thus the following conclusions can be treated as a complement to the classical literature on local identification at θ_0.

Theorem 4.2.14 SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION AT θ_0. *Let $\theta_0 \in \Theta$ and $\gamma : \Theta \subseteq \mathbb{R}^k \mapsto \mathbb{R}^G$ be a parametric function which is locally identifiable at θ_0. Suppose $\gamma(\theta)$ is differentiable at θ_0 with Jacobian matrix $J_\gamma(\theta_0)$. Then the condition*

$$\text{rank}[J_\gamma(\theta_0)] = k \tag{4.2.12}$$

implies that θ is locally identifiable at θ_0.

It is worth noting that the parametric function $\gamma(\theta)$ in Theorem **4.2.14** is in a general form without any specification restrictions. Hence, Theorem **4.2.14** includes the statement by Chen et al. (2011) as a special case if $\gamma(\theta)$ takes the form of the moment function.

4.2.3. Necessary and sufficient condition for local identification at a point

As a further extension to the classical condition for identification at θ_0, we state a condition for local identification at θ_0 which is both necessary and sufficient with rigorous proof.

Theorem 4.2.15 NECESSARY AND SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION AT θ_0. *Let $\theta_0 \in \Theta$ and $\gamma : \Theta \subseteq \mathbb{R}^k \mapsto \mathbb{R}^G$ be a parametric function which is locally identifiable at θ_0. Suppose $\gamma(\theta)$ is continuously differentiable in some open neighborhood of θ_0 with Jacobian matrix $J_\gamma(\theta)$ and θ_0 is a regular point of $J_\gamma(\theta)$. Then θ is locally identifiable at θ_0 if and only if*

$$\text{rank}[J_\gamma(\theta_0)] = k. \tag{4.2.13}$$

Chen et al. (2011) propose a sufficient condition for identification through nonlinear moment equations and they focus on identification of the linear parameter θ, rather than an arbitrary nonlinear function of θ which we call parametric function identification and will be studied in detail in Section 4.4. In comparison, Theorem **4.2.15** establishes both both necessary and sufficient condition for local identification. Hence, we need to impose stronger assumptions in Theorem **4.2.15** than those in both Corollary **4.2.13** and Theorem **4.2.14**. First of all, the parametric function $\gamma(\theta)$ has to be continuously differentiable in some open neighborhood of θ_0. In constrast, the sufficient condition in Theorem **4.2.14** only requires that $\gamma(\theta)$ is differentiable at θ_0. Second, the assumption that θ_0 is a regular point of $J_\gamma(\theta)$ is crucial for the establishment of necessity; see Rothenberg (1971)

and Shapiro (1986). On the other hand, the validity of sufficient condition alone in Theorem **4.2.14** does not need such a regularity assumption. Nevertheless, Theorem **4.2.15** is still classical because it relies on full rank Jacobian matrix to obtain local identification at θ_0. In the following sections, we will derive more powerful identification conclusions based on Theorem **4.2.15** that substitute much weaker conditions for the full rank or non-singular Jacobian matrix assumption.

4.3. Local identification around a point

In this section, we introduce the definition of local identification around θ_0 which implies the usual meaning of local identification at θ_0 and establish both necessary and sufficient condition for local identification in this stronger concept. As discussed, more often we have to settle for local identification due to the difficulty of globally identifying the parameter of interest. Nevertheless, if parameters can be identified within a neighborhood of some other given parameter values rather than over the entire parameter space, we can still achieve *"global"* identification in the sense that any two distinct parameter points will entail two different parameterized probability distributions within the whole neighborhood which can be treated as a restricted parameter space.

As discussed in Section 4.2, the following definitions also take into account the semi-parametric setup by allowing the nuisance parameter space Ξ to be infinitely dimensional based on the assumption that the parameter of interest θ can be separated form the nuisance parameter ν.

Definition 4.3.1 LOCAL IDENTIFICATION OF PARAMETERS AROUND θ_0. *The parameter θ is locally identifiable around $\theta_0 \in \Theta$ if and only if there exists an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 such that*

$$\left(\theta_1 \neq \theta_2\right) \Rightarrow \left(\mathsf{P}_{(\theta_1,\nu_1)} \neq \mathsf{P}_{(\theta_2,\nu_2)}\right), \quad \forall \theta_1,\ \theta_2 \in \mathscr{V}(\theta_0),\ \forall \nu_1,\ \nu_2 \in \Xi\,. \tag{4.3.1}$$

Compared to Definition **4.2.5** which states that θ is locally identifiable at θ_0 if and only if there does not exist any other parameter in such an open neighborhood of θ_0 which is observationally equivalent to θ_0, Definition **4.3.1** emphasizes that θ is locally identifiable around θ_0 if and only if no point in the neighborhood of θ_0 is observationally equivalent to any other distinct point in the same neighborhood. As a consequence, local identification around θ_0 is stronger than local identification at θ_0 and thus implies the latter. Equivalently

we can rewrite (4.3.1) as

$$\left(\mathsf{P}_{(\theta_1,\nu_1)} = \mathsf{P}_{(\theta_2,\nu_2)}\right) \Rightarrow \left(\theta_1 = \theta_2\right), \quad \forall \theta_1,\, \theta_2 \in \mathscr{V}(\theta_0),\, \forall \nu_1,\, \nu_2 \in \Xi\,. \tag{4.3.2}$$

Similarly, we define local identification of parametric function around θ_0 and local identification through another identifiable parametric function.

Definition 4.3.2 LOCAL IDENTIFICATION OF PARAMETRIC FUNCTIONS AROUND θ_0. *The parametric function $\gamma : \Theta \to \Gamma$ is locally identifiable around $\theta_0 \in \Theta$ if and only if there exists an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 such that*

$$\left(\gamma(\theta_1) \neq \gamma(\theta_2)\right) \Rightarrow \left(\mathsf{P}_{(\theta_1,\nu_1)} \neq \mathsf{P}_{(\theta_2,\nu_2)}\right), \quad \forall \theta_1,\, \theta_2 \in \mathscr{V}(\theta_0),\, \forall \nu_1,\, \nu_2 \in \Xi\,. \tag{4.3.3}$$

Definition 4.3.3 LOCAL IDENTIFICATION IN TERMS OF ANOTHER PARAMETRIC FUNCTION AROUND θ_0. *Let $\beta : \Theta \to B$ and $\gamma : \Theta \to \Gamma$ be two parametric functions. $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ around $\theta_0 \in \Theta$ if and only if there exists an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 such that*

$$\left(\beta(\theta_1) \neq \beta(\theta_2)\right) \Rightarrow \left(\gamma(\theta_1) \neq \gamma(\theta_2)\right), \quad \forall \theta_1,\, \theta_2 \in \mathscr{V}(\theta_0). \tag{4.3.4}$$

Assuming $\gamma(\theta)$ is locally identifiable around θ_0 and continuously differentiable in Theorem **4.2.15**, we can derive the the conditions for local identification of θ around θ_0.

We next state a modified inverse function theorem based on which the conditions for local identification around a point θ_0 can be established. Note that the continuity of the Jacobian matrix at θ_0 plays a crucial role in the conditions for the local identification around θ_0 as demonstrated in the proof. In fact, the inverse function theorem by Rudin (1976) should be modified slightly by more accurate assumptions on the continuity and differentiability of the parametric functions. First, recall the definitions of continuous differentiability of a function at a point and on an open set in Rudin (1976, Definition 9.20). Let $\mathscr{E}$ be an open set of $\mathbb{R}^n$ and assume $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ be a differentiable mapping on $\mathscr{E}$. Then f is continuously differentiable at $a \in \mathscr{E}$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\|J_f(x) - J_f(a)\| < \varepsilon$$

for all points $x \in \mathscr{E}$ and $\|x - a\| < \delta$. If f is continuously differentiable at every point of $\mathscr{E}$, then f is said to be continuously differentiable on $\mathscr{E}$. Next, we summarize such a variation of the inversion function theorem as follows.

Theorem 4.3.4 POINTWISE INVERSE FUNCTION THEOREM. *Let $f : \mathbb{R}^m \mapsto \mathbb{R}^m$ be a differentiable mapping in some open set $\mathscr{E} \subseteq \mathbb{R}^m$ with continuous differentiability at some $a \in \mathscr{E}$. Suppose $J_f(a)$ is invertible and $f(a) = b$. Then there exist open sets $\mathscr{A}$ and $\mathscr{B}$ in $\mathbb{R}^m$ such that $a \in \mathscr{A}$, and $b \in \mathscr{B}$, f is one-to-one on $\mathscr{A}$ and $f(\mathscr{A}) = \mathscr{B}$. If g is the inverse of f and defined in $\mathscr{B}$ by*

$$g(f(x)) = x, \ (x \in \mathscr{A}), \tag{4.3.5}$$

then g is differentiable in $\mathscr{B}$. Furthermore, if assume $f : \mathbb{R}^m \mapsto \mathbb{R}^m$ is continuously differentiable in some open set $\mathscr{E} \subseteq \mathbb{R}^m$ and keep other assumptions unchanged, then g is continuously differentiable in $\mathscr{B}$.

As pointed out by Rudin (1976) and Nijenhuis (1974), the differentiability of f in $\mathscr{E}$, the invertibility of $J_f(a)$ and the continuity of J_f at a are sufficient to guarantee f being one-to-one on $\mathscr{A}$ and the existence of the inverse function g. The assumption of the continuous differentiability of f on $\mathscr{E}$ is only required to show that the inverse function g is continuously differentiable in $\mathscr{B}$, which has no impact on the local identification of θ in this paper. Hence the assumptions that are required to establish the local identification around θ_0 are weaker than those in the inverse function theorem by Rudin (1976).

Theorem 4.3.5 CONDITIONS FOR LOCAL IDENTIFICATION AROUND θ_0. *Let $\theta_0 \in \Theta$ and $\gamma : \Theta \mapsto \mathbb{R}^G$ be a parametric function which is locally identifiable around θ_0. Suppose $\gamma(\theta)$ is differentiable in some open neighborhood of θ_0 with continuous differentiability at θ_0. Then the condition*

$$\text{rank}[J_\gamma(\theta_0)] = k \tag{4.3.6}$$

implies that θ is locally identifiable around θ_0. Furthermore, suppose $\gamma(\theta)$ is continuously differentiable in some open neighborhood of θ_0 and θ_0 is a regular point of $J_\gamma(\theta)$. Then θ is locally identifiable around θ_0 if and only if $\text{rank}[J_\gamma(\theta_0)] = k$.

The establishment of Theorem **4.3.5** depends on the inverse function theorem [see Rudin (1976)]. However, as pointed out by Krantz and Parks (2002, Theorem 3.3.2) and Clarke (1976), we can have a stronger version of the inverse function theorem where $\gamma(\theta)$ is p-th order continuously differentiable. For convenience, let's recall the definition of the p-th order continuous differentiability from Lieb and Loss (2000).

Definition 4.3.6 p-TH ORDER CONTINUOUS DIFFERENTIABILITY. *Let $f : \Psi \subseteq \mathbb{R}^n \mapsto \mathbb{R}^m$ and $\frac{\partial^p f_j}{\partial x_{i_1} \ldots \partial x_{i_p}}$ be the p-th partial derivative of f_j with respect to $x_{i_1}, \ldots, x_{i_p}$, where $j = 1, 2, \ldots, m$. If $\frac{\partial^p f_j}{\partial x_{i_1} \ldots \partial x_{i_p}}$ exist at all $x \in \Psi$ and are continuous functions on Ψ, then f is said to be p-th order continuously differentiable in Ψ and denoted as $C^p(\Psi)$.*

Hence we give the following proposition based on a stronger version of the inverse function theorem.

Theorem 4.3.7 NECESSARY AND SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION AROUND θ_0 WITH THE p-TH ORDER CONTINUOUSLY DIFFERENTIABLE FUNCTION. *Let $\theta_0 \in \Theta$ and $\gamma : \Theta \mapsto \mathbb{R}^G$ be a parametric function which is locally identifiable around θ_0. Suppose $\gamma(\theta)$ is p-th order continuously differentiable in some open neighborhood of θ_0 and θ_0 is a regular point of $J_\gamma(\theta)$. Then θ is locally identifiable around θ_0 if and only if*

$$\text{rank}[J_\gamma(\theta_0)] = k. \tag{4.3.7}$$

In connection to the proof of necessity of Theorem **4.2.15** and Theorem **4.3.5**, we establish herein a statement that justifies the reverse argument of the inverse function theorem .

Proposition 4.3.8 INVERTIBILITY OF THE JACOBIAN MATRIX. *Let $\gamma : \Theta \mapsto \mathbb{R}^G$ be a parametric function which is locally identifiable around θ_0 and continuously differentiable on $\mathscr{V}(\theta_0)$. Suppose the inverse function of $\gamma(\theta)$ exists and is continuously differentiable. If θ is locally identifiable around θ_0, then* $\text{rank}[J_\gamma(\theta)] = k$, $\forall \theta \in \mathscr{V}(\theta_0)$.

4.4. Parametric function identification

Rothenberg (1971) proposes the the classical identification conditions for the parameter vector of interest or a scalar parameter through the rank of the Jacobian matrix of a given locally identifiable function. Bowden (1973) generalizes some of the conclusions of Rothenberg (1971) based on Kullback's information criterion without imposing a series of regularity assumptions on the density function. Bekker and Wansbeek (2001) attack on parametric function identification but only focus on the scalar parameter. Recently, Dasgupta et al. (2007) manage to extend the local identification conditions by Rothenberg (1971) to a function of parameters using the Fisher information matrix. More often, people are interested in the identification of a linear combination of parameters either because it is infeasible to identify the entire parameter vector or because the information induced from the identification of a function of the parameters will suffice to conduct statistical inference. Therefore we establish herein a theorem that generalizes all the above classical identification results through the local identification condition of a parametric function in terms of another locally identifiable parametric function appealing to the linear subspace properties of their kernels. Such a parametric function of interest can be the entire parameter vector

θ, any scalar component of θ, an arbitrary linear combination of θ or any mapping of θ that can be expressed as a transformation of an identifiable parametric function. Moreover, it can take the form of the score function, the Kullback-Leibler divergence and the first order partial derivative of the Kullback-Leibler divergence, which includes the parametric function identification conditions in Bowden (1973) and Dasgupta et al. (2007) as special cases. Furthermore, it is also worth emphasizing that the general condition is just one of a series of powerful equivalent statements in Dufour and Liang (2012) which are readily modified for local identification.

4.4.1. Identification of general parametric functions

First recall the definitions of the image space and the kernel space; see Abadir and Magnus (2005).

Definition 4.4.1 IMAGE SPACE. *Let A be a $m \times n$ matrix. The image space of A denoted as* $\mathrm{Im}(A)$ *is the set*

$$\mathrm{Im}(A) \equiv \{y \in \mathbb{R}^m : Ax = y \text{ for some } x \in \mathbb{R}^n\}.$$

Definition 4.4.2 KERNEL SPACE. *Let A be a $m \times n$ matrix. The kernel space of A denoted as* $\ker(A)$ *is the set*

$$\ker(A) \equiv \{x \in \mathbb{R}^n : Ax = 0\}.$$

Theorem 4.4.3 CONDITIONS FOR LOCAL IDENTIFICATION OF PARAMETRIC FUNCTIONS AT θ_0. *Suppose $\gamma : \mathbb{R}^k \mapsto \mathbb{R}^G$ is differentiable at θ_0 and locally identifiable at θ_0. Also assume $\beta : \mathbb{R}^k \mapsto \mathbb{R}^H$ is continuously differentiable in an open neighborhood of θ_0 and θ_0 is a regular point of $J_\beta(\theta)$. Then the condition*

$$\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big) \tag{4.4.1}$$

implies that $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ at θ_0. Furthermore, suppose $\gamma(\theta)$ and $\beta(\theta)$ are continuously differentiable in an open neighborhood of θ_0 and θ_0 is a regular point of both $J_\gamma(\theta)$ and $J_\beta(\theta)$. Then $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ at θ_0 if and only if $\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big)$.

More often it will be interesting to obtain identification condition of linear parameters $Q\theta$ which is a special case of the general setup $\beta(\theta)$.

Corollary 4.4.4 CONDITIONS FOR LOCAL IDENTIFICATION OF LINEAR PARAMETRIC FUNCTIONS AT θ_0. *Suppose $\gamma : \mathbb{R}^k \mapsto \mathbb{R}^G$ is differentiable at θ_0 and locally identifiable at θ_0. Let Q be a $H \times k$ matrix. Then the condition*

$$\mathrm{Im}\big(Q'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big) \tag{4.4.2}$$

implies that $Q\theta$ is locally identifiable in terms of $\gamma(\theta)$ at θ_0. Furthermore, suppose $\gamma(\theta)$ is continuously differentiable in an open neighborhood of θ_0 and θ_0 is a regular point of $J_\gamma(\theta)$. Then $Q\theta$ is locally identifiable in terms of $\gamma(\theta)$ at θ_0 if and only if $\mathrm{Im}\big(Q'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big)$.

Corollary **4.4.4** is a generalization of Theorem **4.2.15**, which becomes clear if we think of the parametric function $Q\theta$ as θ. It is trivial that θ is continuously differentiable in Θ and its Jacobian matrix has a constant rank k, i.e.,

$$Q = I_k, \quad \forall \theta \in \Theta,$$

which means that

$$\mathrm{Im}\big(Q'\big) = \mathbb{R}^k, \quad \forall \theta \in \Theta.$$

If

$$\mathrm{Im}\big(Q'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big),$$

it follows that $\mathrm{Im}\big(J_\gamma(\theta_0)'\big) = \mathbb{R}^k$. Hence

$$\mathrm{rank}\left[J_\gamma(\theta_0)\right] = \mathrm{rank}\left[J_\gamma(\theta_0)'\right] = \dim\left(J_\gamma(\theta_0)'\right) = k,$$

which is the result of Theorem **4.2.15**. On the other hand, if we assume

$$\mathrm{rank}\left[J_\gamma(\theta_0)\right] = k,$$

the image space of $J_\gamma(\theta_0)'$ equals $\mathbb{R}^k$ which contains any other image space whose element is of k dimension. Meanwhile, $\mathrm{Im}\big(Q'\big) = \mathbb{R}^k$ and we have

$$\left[\mathrm{Im}\big(Q'\big) = \mathrm{Im}\big(J_\gamma(\theta_0)'\big)\right] \Rightarrow \left[\mathrm{Im}\big(Q'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big)\right].$$

So (4.2.12) and (4.4.2) are equivalent if $Q\theta$ is taken to be θ.

It is natural to extend the above conditions for local identification at θ_0 to local identification around θ_0. As it turns out, such a generalization will require stronger assumptions.

Particularly, the establishment of sufficiency needs θ_0 to be a regular point of both $J_\gamma(\theta)$ and $J_\beta(\theta)$ compared to Theorem **4.4.3** where we achieve sufficiency assuming that θ_0 is a regular point of only $J_\beta(\theta)$.

Theorem 4.4.5 NECESSARY AND SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION OF PARAMETRIC FUNCTIONS AROUND θ_0. *Let $\gamma : \mathbb{R}^k \mapsto \mathbb{R}^G$ and $\beta : \mathbb{R}^k \mapsto \mathbb{R}^H$ be two parametric functions. Assume $\gamma(\theta)$ is locally identifiable around θ_0. Suppose $\gamma(\theta)$ and $\beta(\theta)$ are continuously differentiable in an open neighborhood of θ_0 and θ_0 is a regular point of both $J_\gamma(\theta)$ and $J_\beta(\theta)$. Then $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ around θ_0 if and only if*

$$\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big). \tag{4.4.3}$$

Remark 4.4.6 Different from the case of the local identification of the parameter θ, we need the assumption that θ_0 is a regular point of the Jacobian matrices to establish both the necessary and sufficient for the local identification condition of $\beta(\theta)$ around θ_0.

As a direct consequence of Theorem **4.4.5**, we provide two useful results on the differentiability given $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ around θ_0.

Proposition 4.4.7 DIFFERENTIABILITY OF THE REPARAMETERIZED FUNCTION. *Let $\gamma : \mathbb{R}^k \mapsto \mathbb{R}^G$ and $\beta : \mathbb{R}^k \mapsto \mathbb{R}^H$ be two parametric functions. Assume $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ around θ_0. Suppose $\gamma(\theta)$ and $\beta(\theta)$ are continuously differentiable in an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 and θ_0 is a regular point of both $J_\gamma(\theta)$ and $J_\beta(\theta)$. Then there exists a mapping $\bar{\beta} : \mathscr{U} \mapsto \mathscr{V}$, where $\mathscr{U} \subseteq \{\gamma(\theta) : \theta \in \mathscr{V}(\theta_0)\}$ and $\mathscr{V} \subseteq \{\beta(\theta) : \theta \in \mathscr{V}(\theta_0)\}$ and $\bar{\beta}$ is differentiable in γ.*

Proposition 4.4.8 DIFFERENTIABILITY OF THE REPARAMETERIZED FUNCTION WITH p-TH ORDER CONTINUOUS DIFFERENTIABILITY. *Let $\gamma : \mathbb{R}^k \mapsto \mathbb{R}^G$ and $\beta : \mathbb{R}^k \mapsto \mathbb{R}^H$ be two parametric functions. Assume $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ around θ_0. Suppose $\gamma(\theta)$ and $\beta(\theta)$ are p-th order continuously differentiable in an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 and θ_0 is a regular point of both $J_\gamma(\theta)$ and $J_\beta(\theta)$. Then there exists a mapping $\bar{\beta} : \mathscr{U} \mapsto \mathscr{V}$, where $\mathscr{U} \subseteq \{\gamma(\theta) : \theta \in \mathscr{V}(\theta_0)\}$ and $\mathscr{V} \subseteq \{\beta(\theta) : \theta \in \mathscr{V}(\theta_0)\}$ and $\bar{\beta}$ is p-th order continuously differentiable in γ.*

In many statistical cases, the information from the observations alone is inadequate to identify the parameters of interest and some prior information has to be imposed. Let's denote such a prior information regarding the parameters as

$$\xi(\theta_0) = c, \tag{4.4.4}$$

where $\xi : \mathbb{R}^k \mapsto \mathbb{R}^R$ and c is a known constant. Let $J_\xi(\theta)$ be the Jacobian matrix of $\xi(\theta)$. Along the lines to Theorem **4.4.3** and Theorem **4.4.5**, we state the general local identification condition with restrictions as follows.

Corollary 4.4.9 LOCAL IDENTIFICATION OF PARAMETRIC FUNCTIONS AT θ_0 WITH RESTRICTIONS. *Suppose $\gamma(\theta)$, $\beta(\theta)$ and $\xi(\theta)$ are continuously differentiable in an open neighborhood of θ_0 and θ_0 is a regular point of $J_\gamma(\theta)$, $J_\beta(\theta)$ and $J_\xi(\theta)$. Suppose $\gamma(\theta)$ is locally identifiable at θ_0. Then*

$$\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\begin{bmatrix} J_\gamma(\theta_0)' & J_\xi(\theta_0)' \end{bmatrix}$$

is both necessary and sufficient for $\beta(\theta)$ to be locally identifiable at θ_0.

Corollary 4.4.10 LOCAL IDENTIFICATION OF PARAMETRIC FUNCTIONS AROUND θ_0 WITH RESTRICTIONS. *Suppose $\gamma(\theta)$, $\beta(\theta)$ and $\xi(\theta)$ are continuously differentiable in an open neighborhood of θ_0 and θ_0 is a regular point of $J_\gamma(\theta)$, $J_\beta(\theta)$ and $J_\xi(\theta)$. Suppose $\gamma(\theta)$ is locally identifiable around θ_0. Then*

$$\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\begin{bmatrix} J_\gamma(\theta_0)' & J_\xi(\theta_0)' \end{bmatrix}$$

is both necessary and sufficient for $\beta(\theta)$ to be locally identifiable around θ_0.

Since linear or loglinear approximation has been a popular method to deal with the solution to the DSGE models, we provide an interesting necessary and sufficient condition of local identification for the product of the deep parameter and its Jacobian matrix evaluated at some point in a subset of the parameter space.

Proposition 4.4.11 EQUIVALENT LOCAL IDENTIFICATION OF A PARAMETRIC FUNCTION AND THE PRODUCT OF ITS JACOBIAN AND THE PARAMETER. *Let $\gamma : \mathbb{R}^k \mapsto \mathbb{R}^G$ be a differentiable parametric function. Suppose θ_0 is a regular point of $J_\gamma(\theta)$. Then the local identification of $\gamma(\theta)$ around (or at) θ_0 entails the local identification of $J_\gamma(\theta_0)\theta$ around (or at) θ_0 and vice versa.*

Proposition **4.4.11** does not impose any restrictions on the rank of the Jacobian matrix $J_\gamma(\theta_0)$ in contrast to Theorem **4.2.15**. As a consequence the Jacobian matrix $J_\gamma(\theta)$ evaluated at θ_0 can be deficient in rank. More importantly, the identification of the product of the parameters and the partial derivatives of a nonlinear function with respect to these parameters can be extended to higher order derivative case.

4.4.2. Alternative formulations of local identification conditions based on linearization

In order to show the generality of our proposed equivalent statements regarding the necessary and sufficient identification conditions for an arbitrary linear combination of parameters, it is straightforward to slightly modify our global identification results and concentrate on the identification of parameters of nonlinear models within the given neighborhoods of specified parameter values; see Dufour and Liang (2012). Specifically, if $\gamma(\theta)$ is locally identifiable around θ_0 and the subspace condition (4.4.3) holds, $\beta(\theta)$ is locally identifiable around θ_0 by Theorem **4.4.5**. Nevertheless, such a kernel property is not straightforward and may be hard to check in practice since statisticians usually look at a model through explanatory variables. Thus we propose a group of equivalent identification conditions based on the characteristics of subspaces.

Proposition 4.4.12 EQUIVALENT NECESSARY AND SUFFICIENT CONDITIONS FOR LOCAL IDENTIFICATION AROUND θ_0. *Let $\gamma : \Theta \mapsto \mathbb{R}^G$ and $\beta : \Theta \mapsto \mathbb{R}^H$ be at differentiable at θ_0 with derivatives $J_\gamma(\theta_0)$ and $J_\beta(\theta_0)$ repectively. Then the following properties are equivalent:*

$$\mathrm{Im}[J_\beta(\theta_0)'] \subseteq \mathrm{Im}[J_\gamma(\theta_0)']; \tag{4.4.5}$$

$$\ker[J_\gamma(\theta_0)] \subseteq \ker[J_\beta(\theta_0)]; \tag{4.4.6}$$

$$J_\beta(\theta_0) = F(\theta_0)J_\gamma(\theta_0), \textit{ for some matrix } F(\theta_0); \tag{4.4.7}$$

$$\mathrm{rank}\begin{bmatrix} J_\gamma(\theta_0) \\ J_\beta(\theta_0) \end{bmatrix} = \mathrm{rank}(J_\gamma(\theta_0)); \tag{4.4.8}$$

$$\mathrm{rank}\begin{bmatrix} J_\gamma(\theta_0) \\ J_\beta(\theta_0) + V_1(\theta_0)J_\gamma(\theta_0) \end{bmatrix} = \mathrm{rank}(J_\gamma(\theta_0)), \textit{ for any } q \times p \textit{ matrix } V_1(\theta_0); \tag{4.4.9}$$

$$\mathrm{rank}\begin{bmatrix} J_\gamma(\theta_0) + V_2(\theta_0)J_\beta(\theta_0) \\ J_\beta(\theta_0) \end{bmatrix} = \mathrm{rank}(J_\gamma(\theta_0)), \textit{ for any } p \times q \textit{ matrix } V_2(\theta_0); \tag{4.4.10}$$

$$\mathrm{rank}\begin{bmatrix} J_\gamma(\theta_0) \\ S(\theta_0)J_\beta(\theta_0) \end{bmatrix} = \mathrm{rank}(J_\gamma(\theta_0)), \tag{4.4.11}$$

where $S(\theta_0)$ is a $q_1 \times q$ matrix such that $q_1 = \mathrm{rank}[S(\theta_0)J_\beta(\theta_0)] = \mathrm{rank}[J_\beta(\theta_0)]$;

$$\mathrm{rank}\Big\{J_\gamma(\theta_0)(I - J_\beta(\theta_0)^- J_\beta(\theta_0))\Big\} = \mathrm{rank}(J_\gamma(\theta_0)) - \mathrm{rank}(J_\beta(\theta_0)), \textit{ for some g-inverse } J_\beta(\theta_0)^-; \tag{4.4.12}$$

$$J_\beta(\theta_0) = J_\beta(\theta_0)J_\gamma(\theta_0)^- J_\gamma(\theta_0), \textit{ for some g-inverse } J_\gamma(\theta_0)^-. \tag{4.4.13}$$

We herein emphasize that (4.4.5) is just one of a group of equivalences that can be used as necessary and sufficient conditions for local identification under different assumptions. For instance, given $\gamma(\theta)$ is locally identifiable at θ_0, if we further assume continuous differentiability of $\gamma(\theta)$ and $\beta(\theta)$ and regularity of θ_0, all the equivalent conditions in Proposition **4.4.12** are both necessary and sufficient for local identification of $\beta(\theta)$ in terms of $\gamma(\theta)$ at θ_0 according to Theorem **4.4.3**. If we substitute the assumption of local identification of $\gamma(\theta_0)$ around θ_0 with that at θ_0 while keeping unchanged all other restrictions, Theorem **4.4.5** shows that these equivalent conditions are both necessary and sufficient for local identification of $\beta(\theta)$ in terms of $\gamma(\theta)$ around θ_0. Furthermore, the elements of $J_\gamma(\theta)$ and $J_\beta(\theta)$ are generally nonlinear transformations of θ. However, they can also be either independent of θ or linear functions of θ. If this is the case, we are back to the classical linear regression models where the necessary and sufficient identification conditions of an arbitrary linear combination of parameters are global. Thus the above modified local identification conditions are indeed an extension to our global identificaion results. As can be seen later, such an extension is crucial for the analysis of identifiability of parameters in the DSGE models.

We next explain in detail the meaning of the above equivalent conditions and demonstrate that they are intuitive and easy to apply. (4.4.6) and (4.4.5) are equivalent due to the duality between the kernel space and the image space, see Gouriéroux and Monfort (1995). (4.4.6) plays an important role in the proof the fundamental results, such as Theorem **4.4.3** and Theorem **4.4.5** whereas (4.4.6) is more intuitive since empirical research using regression analysis usually looks at the design matrix in columns. As for the linear regression models, Rao and Mitra (1971) and among others propose estimability conditions that are similar to (4.4.5) and (4.4.7). In fact, Reiersøl (1963) and Seely (1977) show that these conditions are necessary and sufficient for both estimability and identifiability of the whole parameter vector. In contrast, we extend the classical results on identification from linear parameters to nonlinear parameters. Specifically, (4.4.5) and (4.4.7) are concerned about identification of parameter in a more general form. Both $\gamma(\theta)$ and $\beta(\theta)$ can be either linear or nonlinear functions of θ. If they are both linear in θ, then the Jacobian matrices $J_\gamma(\theta)$ and $J_\beta(\theta)$ become constant and we are back to the identification for linear parameters. Besides, $\beta(\theta)$ can be the whole parameter θ, any element of θ or an arbitrary linear combination of θ, see Dufour and Liang (2012). On the other hand, if either $\gamma(\theta)$ or $\beta(\theta)$ is nonlinear or both of them are nonlinear, then we can check the relationship between the subvector spaces of their Jacobian matrices. That is to say, if $J_\beta(\theta_0)'$ can be spanned by the columns of $J_\gamma(\theta_0)'$, $\beta(\theta)$ is identifiable in terms of $\gamma(\theta)$. Otherwise, we cannot locally identify $\beta(\theta)$ around (or at) θ_0. Moreover the rank condition (4.4.8) and

its variations (4.4.9), (4.4.10) and (4.4.11) provide a straightforward of checking the identification of $\beta(\theta)$ given that $\gamma(\theta)$ is locally identifiable. As it turns out that (4.4.8) can be easily applied to some important statistical and economic models, such as the simultaneous equation model and the DSGE models. What is worth mentioning is that these rank conditions are no longer in the classical sense since they are both necessary and sufficient for parametric function identification of $\beta(\theta)$ rather than θ and we impose no restrictions on the rank of the Jacobian matrices of $\gamma(\theta)$ and $\beta(\theta)$ evaluated at θ_0 which can be deficient in our setup. Finally, the generalized inverse conditions (4.4.12) and (4.4.12) are natural extensions to the estimability results by Alalouf and Styan (1979*a*) in a nonlinear framework for parametric function identification. In the following discussions we will first focus particularly the rank condition (4.4.8). Then we look through other equivalent local identification conditions.

4.5. Identification in likelihood models

We now concentrate on parametric models and assume θ contains all model parameters so that no nuisance parameter is involved in the parametric model setups. To demonstrate the applications of the results hereinbefore, such as Theorem **4.2.15**, we can take the locally identifiable parametric function $\gamma(\theta)$ as the Kullback-Leibler distance and its first order derivative. We show they are both locally identifiable and the subsequent consequences coincide with the conclusion by Rothenberg (1971).

Recall first the definition of the Kullback-Leibler divergence [see also Gouriéroux and Monfort (1995)]. Let $(\Omega, \mathscr{A}, \mathscr{P})$ be a statistical model on a set of observations $\mathscr{Y}$, where Ω is a sample space, $\mathscr{A}$ is a σ-algebra of subsets of Ω, $\mathscr{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ is a family of probability measures on $(\Omega, \mathscr{A})$, Θ is an open subset of $\mathbb{R}^k$. Assume the probability distributions have densities with respect to the same dominating measure μ. Let $f(y;\theta)$ and $f(y;\theta_0)$ be two densities related to P_θ and P_{θ_0} respectively. Then the measure of the distance between P_θ and P_{θ_0}

$$D_{KL}(\theta|\theta_0) \equiv \mathbb{E}_{\theta_0}\log\frac{f(y;\theta_0)}{f(y;\theta)} = \int_{\mathscr{Y}} \log\frac{f(y;\theta_0)}{f(y;\theta)} f(y;\theta_0)\mu(\mathrm{d}y)$$

is called the Kullback-Leibler divergence.

Proposition 4.5.1 GLOBAL IDENTIFICATION OF THE KULLBACK-LEIBLER DIVERGENCE OVER Θ. *The Kullback-Leibler divergence $D_{KL} : \Theta \mapsto \mathbb{R}$ is globally identifiable over Θ.*

It follows immediately that $D_{KL}: \Theta \mapsto \mathbb{R}$ is globally identifiable at θ_0 and is locally identifiable around θ_0. Furthermore, we can demonstrate that the first order partial derivative of the Kullback-Leibler divergence is locally identifiable at θ_0 based on Lebesgue's dominated convergence theorem [see Rudin (1976) and Casella and Berger (2002)], which will lead to the well-known classical rank condition for identification.

Proposition 4.5.2 GLOBAL IDENTIFICATION OF THE FIRST ORDER DERIVATIVE OF THE KULLBACK-LEIBLER DIVERGENCE AT θ_0. *Suppose the following assumptions are satisfied:*

(1) $f(y;\theta)$ *is strictly positive and differentiable on* Θ $\forall y \in \mathscr{Y}$;

(2) $D_{KL}(\theta|\theta_0)$ *is differentiable on* Θ;

(3) *There exists a Lebesgue integrable function* $g_1(y;\theta)$ *on* $\mathscr{Y}$ *and a constant* c_1 *such that*
$$\left|\left(\frac{\partial \log f(y;\theta)}{\partial \theta}\right)_{\theta=\theta'}\right| \leq g_1(y;\theta) \qquad \forall \theta' \text{ such that } |\theta'-\theta| \leq c_1 \text{ and } \forall y \in \mathscr{Y};$$

(4) *There exists a Lebesgue integrable function* $g_2(y;\theta)$ *on* $\mathscr{Y}$ *and a constant* c_2 *such that*
$$\left|\left(\frac{\partial f(y;\theta)}{\partial \theta}\right)_{\theta=\theta^*}\right| \leq g_2(y;\theta) \qquad \forall \theta^* \text{ such that } |\theta^*-\theta| \leq c_2 \text{ and } \forall y \in \mathscr{Y}.$$

Then the first order derivative of the Kullback-Leibler divergence $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ *is globally identifiable at* θ_0.

If replacing Θ with $\mathscr{V}(\theta_0)$ in conditions (1) and (2) in Proposition **4.5.2**, we conclude that the first order derivative of the Kullback-Leibler divergence $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ is locally identifiable at θ_0. Moreover, the first order derivative of the Kullback-Leibler divergence, however, is usually not locally identifiable around θ_0 (and thus not globally identifiable over Θ) unless the probability distribution parameterized at any $\theta \in \mathscr{V}(\theta_0)$ is equal to the true P_{θ_0}. Additionally, the second order derivative of $D_{KL}(\theta|\theta_0)$ evaluated at θ_0, which is the Fisher information matrix $\mathscr{I}(\theta_0)$, is not locally identifiable at θ_0. Suppose

$$\mathsf{P}_\theta = \mathsf{P}_{\theta_0}.$$

Then

$$\begin{aligned}\frac{\partial^2 D_{KL}(\theta|\theta_0)}{\partial \theta \partial \theta'} &= -\int_{\mathscr{Y}} \frac{\partial^2 \log f(y;\theta)}{\partial \theta \partial \theta'} f(y;\theta_0)\mu(\mathrm{d}y) \\ &= -\mathbb{E}_\theta \frac{\partial^2 \log f(y;\theta)}{\partial \theta \partial \theta'}\end{aligned}$$

$$= \mathscr{I}(\theta),$$

which is generally not equal to $\mathscr{I}(\theta_0)$. Interestingly, using the definition of the Kullback-Leibler divergence, we can provide an alternative way of showing the identity property of the Fisher information matrix [1].

Since the first order derivative of the Kullback-Leibler divergence is locally identifiable at θ_0 from Proposition **4.5.2**, if we take it as $\gamma(\theta)$ in Theorem **4.2.15** we can establish the condition for the local identification of θ through the rank of the Hessian matrix of the Kullback-Leibler divergence by applying Theorem **4.2.15**.

Theorem 4.5.3 CONDITIONS FOR LOCAL IDENTIFICATION THROUGH THE KULLBACK-LEIBLER DIVERGENCE. *Suppose $D_{KL}(\theta|\theta_0)$ is twice differentiable at θ_0 with Hessian matrix $H_{KL}(\theta_0|\theta_0)$. Then the condition*

$$\text{rank}[H_{KL}(\theta_0|\theta_0)] = k \tag{4.5.1}$$

implies that θ is locally identifiable at θ_0. Furthermore, suppose $D_{KL}(\theta|\theta_0)$ is twice continuously differentiable in $\mathscr{V}(\theta_0)$ with Hessian matrix $H_{KL}(\theta|\theta_0)$, and θ_0 is a regular point of $H_{KL}(\theta|\theta_0)$. Then θ is locally identifiable at θ_0 if and only if $\text{rank}[H_{KL}(\theta_0|\theta_0)] = k$.

As shown in the appendix, the first order partial derivative of $D_{KL}(\theta|\theta_0)$ with respect to θ evaluated at θ_0 equals zero and thus locally identifiable at θ_0. Therefore the Hessian matrix of $D_{KL}(\theta|\theta_0)$ has full rank at θ_0 if and only if θ is locally identifiable at θ_0 from Theorem **4.2.15**. The definition of the Kullback-Leibler divergence leads to

$$\frac{\partial^2 D_{KL}(\theta|\theta_0)}{\partial\theta\partial\theta'} = -\mathbb{E}_{\theta_0}\frac{\partial^2 \log f(y;\theta)}{\partial\theta\partial\theta'}, \ \forall\theta \in \Theta.$$

Thus the Hessian matrix $H_{KL}(\theta|\theta_0)$ evaluated at θ_0 is equal to the Fisher information matrix at θ_0. This is the exact necessary and sufficient local identification condition for θ at θ_0 by Rothenberg (1971). Therefore we just establish the classical condition using the Kullback-Leibler divergence and the properties of the score vector. Note that the classical local identification condition by Rothenberg (1971) can be treated as a special case of Theorem **4.2.15** as shown above. However, there is a crucial distinction of our results from those by Rothenberg (1971). To achieve the sufficient condition for local identification of θ at θ_0 Rothenberg (1971) assumes continuous differentiability of the density function in the entire parameter space. In contrast, we only require that the Kullback-Leibler divergence is

[1] Please refer to Corollary A.1 in Appendix.

differentiable at θ_0 in Theorem **4.5.3**. Thus our sufficient condition for local identification at θ_0 is much less restrictive than that of Rothenberg (1971) and consequently implies the latter. Besides, regularity assumption plays no role in the establishment of sufficiency but is essential for necessity as shown in the appendix. This issue has been noticed by Bowden (1973).

Remark 4.5.4 Bowden (1973) proposes a similar statement of identifiability based directly on the rank of the Fisher information matrix. Nonetheless, it is worth pointing out that our result differs in the following aspects. First, we only assume twice differentiability of $D_{KL}(\theta|\theta_0)$ at θ_0 to establish sufficiency compared to Bowden (1973) who requires that the density function is twice continuously differentiable in a neighborhood of θ_0. As demonstrated in the proof of Theorem **4.5.3** , the assumption of continuous differentiability is only required to show necessity. Thus we somehow generalize the sufficient result of Bowden (1973). Second, our rank condition is derived from the fact that the first order derivative $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ is locally identifiable at θ_0 from Theorem **4.2.15** whereas the conclusion in Bowden (1973) still depends on the full rank of the Fisher information matrix as Rothenberg (1971) does. This verifies our argument that the parametric function $\gamma(\theta)$ can be quite inclusive and when it takes the form of the first order derivative of the Kullback-Leibler divergence, Theorem **4.2.15** includes Theorem **4.5.3** and thus the argument by Bowden (1973) as special cases. Third, we make a distinction between local identification around θ_0 and at θ_0 and point out that the first order derivative of the Kullback-Leibler divergence is only locally identifiable at θ_0 but not around θ_0. Consequently, we excludes the possibility of establishing local identification of θ around θ_0 and thus global identification over Θ through the characteristics of the Kullback-Leibler divergence.

Since $D_{KL}(\theta|\theta_0)$ is a function of θ and $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ is globally identifiable at θ_0, we can apply the conclusions of parametric function identification in Section 4.4 to identify a parametric function $\beta(\theta)$ at θ_0.

Corollary 4.5.5 CONDITIONS FOR LOCAL PARAMETRIC FUNCTION IDENTIFICATION THROUGH THE KULLBACK-LEIBLER DIVERGENCE. *Suppose $D_{KL}(\theta|\theta_0)$ is twice differentiable at θ_0 with Hessian matrix $H_{KL}(\theta_0|\theta_0)$ and $\beta(\theta)$ is differentiable at θ_0. Then the condition*

$$\mathrm{Im}\left(J_\beta(\theta_0)'\right) \subseteq \mathrm{Im}\left(H_{KL}(\theta_0|\theta_0)'\right)$$

implies that $\beta(\theta)$ is locally identifiable in terms of $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ at θ_0. Furthermore, suppose $D_{KL}(\theta|\theta_0)$ is twice continuously differentiable and $\beta(\theta)$ is continuously differentiable in an open neighborhood of θ_0. Also assume θ_0 is a regular point of both $H_{KL}(\theta|\theta_0)$ and

$J_\beta(\theta)$. *Then* $\beta(\theta)$ *is locally identifiable in terms of* $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ *if and only if* $\text{Im}\left(J_\beta(\theta_0)'\right) \subseteq \text{Im}\left(H_{KL}(\theta_0|\theta_0)'\right)$.

4.6. Applications

In this section, we apply our results of (global) local identification and function identification to some econometric and macroeconomic models, including the simultaneous equations models (SEMs) and the DSGE models. We demonstrate that the standard conditions for identification so far in the literature can be easily generalized as special cases of identification conditions, in particular the results in Subsection 4.4.1 and their equivalent statements in Subsection 4.4.2.

4.6.1. Simultaneous equations models

Consider a system of G simultaneous equations (4.2.11) in the structural form that satisfies the following conditions:

$$\mathbb{E}(U_t) = \mu \text{ exists, where } t = 1, 2, \ldots, T \text{ and } U' = \begin{bmatrix} U_1 & U_2 & \ldots & U_T \end{bmatrix}; \tag{4.6.1}$$

$$\text{Var}(U_t) = \Sigma \text{ exists and is nonsingular;} \tag{4.6.2}$$

$$\mathbb{E}(X'U) = 0; \tag{4.6.3}$$

$$\text{the first and second moments of all elements of } Y \text{ and } X \text{ exist;} \tag{4.6.4}$$

$$\text{the distribution of the error terms can be fully characterized through its first two moments.} \tag{4.6.5}$$

Note that the necessary and sufficient conditions for identification in a general setup of the SEMs without any constraints on the ranks of X and B have been discussed by Dufour and Liang (2013*a*). Let's denote the $(2G^2 + GK + G)$ structural parameters as

$$\theta \equiv \begin{pmatrix} \text{vec}(B) \\ \text{vec}(\Gamma) \\ \mu \\ \text{vec}(\Sigma) \end{pmatrix}. \tag{4.6.6}$$

Meanwhile, suppose we are given M continuously differentiable constraints based on a

prior information on structural parameters

$$\phi(\theta) = \begin{bmatrix} \phi_1(\theta) \\ \phi_2(\theta) \\ \vdots \\ \phi_M(\theta) \end{bmatrix} = 0 \tag{4.6.7}$$

whose Jacobian matrix is denoted as $J_\phi(\theta)$. Also let's denote the partial derivatives of $\phi(\theta)$ with respect to vec(B), vec(Γ), μ and vec(Σ) as $J_{\phi,B}(\theta)$, $J_{\phi,\Gamma}(\theta)$, $J_{\phi,\mu}(\theta)$ and $J_{\phi,\Sigma}(\theta)$.

The SEMs framework (4.2.11) under assumptions (4.6.1) through (4.6.5) differs from the classical setup in the following aspects. First, the normality assumption (4.6.5) about the distribution of the error term is not required herein compared to Rothenberg (1971) who imposes normal distribution with a separate mean-variance structure. Second, we impose no constraints on the functional forms of a prior restrictions in comparison to Richmond (1974) who assumes that the additional restrictions on the structural parameters are all in linear form. Third, the a prior restrictions can depend on the covariance matrix or the second order moments of the structural shocks compared to Bekker and Wansbeek (2001) who assume that the a prior restrictions are independent of the covariance matrix and functions of only B and Γ. Hence our framework allows for nonlinearity in both the simultaneous structural equations and the a prior restrictions.

For illustration purpose, we rewrite the above structural form and reduced form of the SEMs observation by observation

$$B'Y_t + \Gamma' X_t = U_t, \quad t = 1, 2, \ldots, T \tag{4.6.8}$$

which can be expressed in the matrix form

$$\begin{bmatrix} B' & \Gamma' \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} = U_t, \tag{4.6.9}$$

where Y_t, X_t and U_t are respectively the t-th row of Y, X and U. Due to the assumptions (4.6.1), (4.6.2) and (4.6.4), we write the moment functions

$$m_1(\theta) \equiv \begin{bmatrix} B' & \Gamma' \end{bmatrix} \begin{bmatrix} \mu_{Y_t} \\ \mu_{X_t} \end{bmatrix} - \mu = 0 \tag{4.6.10}$$

and

$$m_2(\theta) \equiv \begin{bmatrix} B' & \Gamma' \end{bmatrix} \begin{bmatrix} \Sigma_{Y_t} & \Sigma_{Y_t X_t} \\ \Sigma_{X_t Y_t} & \Sigma_{X_t} \end{bmatrix} \begin{bmatrix} B \\ \Gamma \end{bmatrix} - \Sigma = 0. \tag{4.6.11}$$

Let's denote θ_0 such that (4.6.7), (4.6.10) and (4.6.11) are satisfied. Clearly, both $m_1(\theta)$ and $m_2(\theta)$ and identifiable. However, the first moment equation m_1 is linear in θ but the second moment equation m_2 is a nonlinear function of θ. Take partial derivatives of $m_1(\theta)$ and $m_2(\theta)$ and denote the $(G+G^2)\times(2G^2+GK+G)$ Jacobian matrix as

$$J_m(\theta) \equiv \begin{bmatrix} J_{m_1}(\theta) \\ J_{m_2}(\theta) \end{bmatrix} = \begin{bmatrix} I_G \bigotimes {\mu_{Y_t}}' & I_G \bigotimes {\mu_{X_t}}' & -I_G & O \\ C_1(\theta) & C_2(\theta) & O & -I_{G^2} \end{bmatrix},$$

where

$$\begin{aligned} C_1(\theta) &= (I_{G^2} + M_{G^2})(I_G \otimes B'\Sigma_{Y_t}) + I_G \otimes \Gamma'\Sigma_{X_t Y_t} + (\Gamma'\Sigma_{X_t Y_t} \otimes I_G) M_{G^2}, \\ C_2(\theta) &= (I_{G^2} + M_{G^2})(I_G \otimes \Gamma'\Sigma_{X_t}) + I_G \otimes B'\Sigma_{Y_t X_t} + (B'\Sigma_{Y_t X_t} \otimes I_G) M_{GK}. \end{aligned}$$

M stands for the commutation matrix [2]; see Abadir and Magnus (2005). Hence, we can readily give the necessary and sufficient condition for identification of an arbitrary parametric function $\beta(\theta)$ according to the general rank condition (4.4.8) as follows:

$$\text{rank} \begin{bmatrix} J_m(\theta_0) \\ J_\phi(\theta_0) \\ J_\beta(\theta_0) \end{bmatrix} = \text{rank} \begin{bmatrix} J_m(\theta_0) \\ J_\phi(\theta_0) \end{bmatrix}. \tag{4.6.12}$$

It is noticeable that without assumption (4.6.5), (4.6.12) and its equivalences are only sufficient and not necessary since the structural parameters may be identified through higher order moments or other statistical properties rather than the first two moments. Furthermore, the validity of (4.6.12) and thus (4.4.8) and its equivalent statements do not depend on the assumptions of full rank X and nonsingular B compared to the standard rank conditions in literature which will fail when either X is deficient in rank or B is singular.

Remark 4.6.1 We do not impose the assumption that μ is completely separable from Σ compared to the zero mean normality assumption in the classical literature. Therefore, the covariance matrix can be a differentiable function in the mean, which is denoted as $\Sigma(\mu)$.

[2] The commutation matrix K_{mn} of a $m \times n$ matrix A is a permutation matrix such that

$$K_{mn}\text{vec}A = \text{vec}A'.$$

If this the is case, the structural parameters that need to be identified can be reduced to

$$\theta \equiv \begin{pmatrix} \text{vec}(B) \\ \text{vec}(\Gamma) \\ \mu \end{pmatrix}$$

and the Jacobian matrix of the moment equations becomes

$$J_m(\theta) = \begin{bmatrix} I_G \bigotimes {\mu_{Y_t}}' & I_G \bigotimes {\mu_{X_t}}' & -I_G \\ C_1(\theta) & C_2(\theta) & C_3(\theta) \end{bmatrix},$$

where $C_3(\theta)$ is the $G^2 \times G$ matrix $-J_\Sigma(\mu)$. It follows that the rank condition will be the same as (4.6.12).

Specifically, let's consider identification issue for the following cases. First, we check identification of θ. Set

$$\beta(\theta) = \theta.$$

From (4.6.12), it follows that the necessary and sufficient condition for identification of θ is

$$\text{rank} \begin{bmatrix} J_m(\theta_0) \\ J_\phi(\theta_0) \\ I_{2G^2+GK+G} \end{bmatrix} = \text{rank} \begin{bmatrix} J_m(\theta_0) \\ J_\phi(\theta_0) \end{bmatrix}$$

which is equivalent to

$$\text{rank} \begin{bmatrix} J_m(\theta_0) \\ J_\phi(\theta_0) \end{bmatrix} = 2G^2 + GK + G. \tag{4.6.13}$$

We emphasize that (4.6.13) is a generalization of Bekker and Wansbeek (2001, Theorem 10). First of all, we do not put any constraint on the rank of B. Therefore, (4.6.13) is valid despite the nonexistence of a unique reduced form equation. Besides, the a prior restriction $\phi(\theta)$ also depends on the covariance matrix Σ. On top of that, the distribution of the structural shocks does not have to be normal. In fact, upon normality assumption and independence of $\phi(\theta)$ from Σ, we only need to focus on the identification of B and Γ without concerning about the identification of the covariance matrix Σ since the normal distribution is fully characterized by its first two moments and the identification of B and Γ implies that of Σ. As for the scalar case under normality assumption, the identification

of one moment entails that of another moment and vice versa; see the appendix of Dufour and Liang (2012).

Second, it is easy to obtain the condition for identification of a subvector of θ, such as $\text{vec}(B)$ by setting

$$\beta(\theta) = \text{vec}(B)$$

and the condition becomes

$$\begin{aligned}
&\text{rank}\begin{bmatrix} I_G \bigotimes \mu_{Y_t}' & I_G \bigotimes \mu_{X_t}' & -I_G & O \\ C_1(\theta_0) & C_2(\theta_0) & O & -I_{G^2} \\ J_{\phi,B}(\theta_0) & J_{\phi,\Gamma}(\theta_0) & J_{\phi,\mu}(\theta_0) & J_{\phi,\Sigma}(\theta_0) \\ I_{G^2} & O & O & O \end{bmatrix} \\
&= \text{rank}\begin{bmatrix} I_G \bigotimes \mu_{Y_t}' & I_G \bigotimes \mu_{X_t}' & -I_G & O \\ C_1(\theta_0) & C_2(\theta_0) & O & -I_{G^2} \\ J_{\phi,B}(\theta_0) & J_{\phi,\Gamma}(\theta_0) & J_{\phi,\mu}(\theta_0) & J_{\phi,\Sigma}(\theta_0) \end{bmatrix}
\end{aligned} \tag{4.6.14}$$

Similarly, it follows that the rank conditions for $\text{vec}(\Gamma)$, μ and $\text{vec}(\Sigma)$ are respectively

$$\begin{aligned}
&\text{rank}\begin{bmatrix} I_G \bigotimes \mu_{Y_t}' & I_G \bigotimes \mu_{X_t}' & -I_G & O \\ C_1(\theta_0) & C_2(\theta_0) & O & -I_{G^2} \\ J_{\phi,B}(\theta_0) & J_{\phi,\Gamma}(\theta_0) & J_{\phi,\mu}(\theta_0) & J_{\phi,\Sigma}(\theta_0) \\ O & I_{GK} & O & O \end{bmatrix} \\
&= \text{rank}\begin{bmatrix} I_G \bigotimes \mu_{Y_t}' & I_G \bigotimes \mu_{X_t}' & -I_G & O \\ C_1(\theta_0) & C_2(\theta_0) & O & -I_{G^2} \\ J_{\phi,B}(\theta_0) & J_{\phi,\Gamma}(\theta_0) & J_{\phi,\mu}(\theta_0) & J_{\phi,\Sigma}(\theta_0) \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
&\text{rank}\begin{bmatrix} I_G \bigotimes \mu_{Y_t}' & I_G \bigotimes \mu_{X_t}' & -I_G & O \\ C_1(\theta_0) & C_2(\theta_0) & O & -I_{G^2} \\ J_{\phi,B}(\theta_0) & J_{\phi,\Gamma}(\theta_0) & J_{\phi,\mu}(\theta_0) & J_{\phi,\Sigma}(\theta_0) \\ O & O & I_G & O \end{bmatrix} \\
&= \text{rank}\begin{bmatrix} I_G \bigotimes \mu_{Y_t}' & I_G \bigotimes \mu_{X_t}' & -I_G & O \\ C_1(\theta_0) & C_2(\theta_0) & O & -I_{G^2} \\ J_{\phi,B}(\theta_0) & J_{\phi,\Gamma}(\theta_0) & J_{\phi,\mu}(\theta_0) & J_{\phi,\Sigma}(\theta_0) \end{bmatrix}
\end{aligned}$$

and

$$\operatorname{rank}\begin{bmatrix} I_G \bigotimes \mu_{Y_t}' & I_G \bigotimes \mu_{X_t}' & -I_G & O \\ C_1(\theta_0) & C_2(\theta_0) & O & -I_{G^2} \\ J_{\phi,B}(\theta_0) & J_{\phi,\Gamma}(\theta_0) & J_{\phi,\mu}(\theta_0) & J_{\phi,\Sigma}(\theta_0) \\ O & O & O & I_{G^2} \end{bmatrix}$$
$$= \operatorname{rank}\begin{bmatrix} I_G \bigotimes \mu_{Y_t}' & I_G \bigotimes \mu_{X_t}' & -I_G & O \\ C_1(\theta_0) & C_2(\theta_0) & O & -I_{G^2} \\ J_{\phi,B}(\theta_0) & J_{\phi,\Gamma}(\theta_0) & J_{\phi,\mu}(\theta_0) & J_{\phi,\Sigma}(\theta_0) \end{bmatrix}.$$

Third, we can identify any scalar parameter θ_i by setting

$$\beta(\theta) = \theta_i, \quad i = 1,2,\ldots,2G^2 + GK$$

and the rank condition is

$$\operatorname{rank}\begin{bmatrix} J_m(\theta_0) \\ J_\phi(\theta_0) \\ e_i' \end{bmatrix} = \operatorname{rank}\begin{bmatrix} J_m(\theta_0) \\ J_\phi(\theta_0) \end{bmatrix} \tag{4.6.15}$$

We next generalize some of the well-known rank conditions in the classical literature as the special cases of our general rank condition and its equivalent statements in subsection 4.4.2 ; see Fisher (1966), Rothenberg (1971) and Bekker and Wansbeek (2001). For comparison purpose, let's further make the standard assumptions that X has full column rank and B is nonsingular. Let's assume $\mathbb{E}(U_t) = 0$ for simplicity. Then we obtain the reduced form of (4.2.11)

$$\Pi = -\Gamma B^{-1} \tag{4.6.16}$$

and $V = UB^{-1}$. Accordingly, the reduced form of (4.6.8) is

$$Y_t = \Pi' X_t + V_t, \quad t = 1,2,\ldots,T, \tag{4.6.17}$$

where $V_t = \left(B'\right)^{-1} U_t$ and is the t-th row of V. Since the random vector U_t is assumed to have a zero mean and a positive definite covariance matrix Σ, we have

$$\mathbb{E}(V_t) = 0$$

and

$$\operatorname{Var}(V_t) \equiv \Omega = \left(B^{-1}\right)' \Sigma \left(B^{-1}\right). \tag{4.6.18}$$

Let's denote the $G^2 + GK$reduced parameters as

$$\psi(\theta) = \begin{pmatrix} \text{vec}(\Pi(\theta)) \\ \text{vec}(\Omega(\theta)) \end{pmatrix}. \tag{4.6.19}$$

and its Jacobian matrix as $J_\psi(\theta)$. Also assume $\Pi(\theta_0)$ and $\Omega(\theta_0)$ satisfy (4.6.16) and (4.6.18).

Remark 4.6.2 Both $\Pi(\theta)$ and $\Omega(\theta)$ are functions of the structural parameters. More importantly, they are locally identifiable at θ_0 since they both can be expressed as the moments of functions of observations X and Y.

First, we consider identification of θ. Following Corollary **4.4.9** let's set

$$\gamma(\theta) = \psi(\theta), \quad \xi(\theta) = \phi(\theta), \quad \beta(\theta) = \theta.$$

Applying the rank condition (4.4.8), we achieve the necessary and sufficient condition for the local identification of θ at θ_0

$$\text{rank} \begin{bmatrix} J_\psi(\theta_0) \\ J_\phi(\theta_0) \\ I_{2G^2+GK} \end{bmatrix} = \text{rank} \begin{bmatrix} J_\psi(\theta_0) \\ J_\phi(\theta_0) \end{bmatrix} \tag{4.6.20}$$

which implies that

$$\text{rank} \begin{bmatrix} J_\psi(\theta_0) \\ J_\phi(\theta_0) \end{bmatrix} = 2G^2 + GK. \tag{4.6.21}$$

If we further assume that the distribution is normal and all M restrictions $\phi(\theta)$ are linear functions of only B and Γ, (4.6.20) becomes the condition for global identification of θ at θ_0; see Rothenberg (1971, Corollary 2) and Bekker and Wansbeek (2001, Corollary 2). Moreover, (4.6.21) also includes the identification condition in the linear SEMs by Richmond (1974, Theorem 5) as a special case. On the other hand, (4.6.21) is an extension to Rothenberg (1971, Theorem 9) since our sufficient result does not assume continuous differentiability of $\psi(\theta)$ and $\phi(\theta)$. Obviously, (4.6.21) is a special case of (4.6.13) by imposing restrictions on the ranks of both X and B. Additionally, (4.6.21) is also a direct consequence of Theorem **4.2.15** which is in turn a special case of Theorem **4.4.3** or Corollary **4.4.9**. Thus we generalize the classical condition for local identification in Rothenberg (1971) for SEMs.

Second, we can check identification for scalar parameters in SEMs using our established conclusions. Without loss of generality, we choose to identify the first structural

parameter θ_1 since we can always conduct the elementary operations on the columns of

$$\begin{bmatrix} J_\psi(\theta_0) \\ J_\phi(\theta_0) \\ e_i', \end{bmatrix}$$

such that the column with its first element equal to 1 becomes the first column without changing the linear relationships among the original columns. Since both $\psi(\theta)$ and $\phi(\theta)$ are locally identifiable at θ_0 and continuously differentiable, it follows from (4.4.8) or (4.6.15) that the necessary and sufficient condition for local identification of θ_1 is

$$\text{rank} \begin{bmatrix} J_\psi(\theta_0) \\ J_\phi(\theta_0) \\ e_1' \end{bmatrix} = \text{rank} \begin{bmatrix} J_\psi(\theta_0) \\ J_\phi(\theta_0) \end{bmatrix}. \tag{4.6.22}$$

Remark 4.6.3 θ_0 is assumed to be a regular point of both $J_\psi(\theta)$ and $J_\phi(\theta)$ which is essential for the establishment of necessity.

Let's compare the rank condition (4.6.22) with other well-known results in the classical literature. To begin with, (4.6.22) resembles Theorem **4.2.10** except that we do not assume continuous differentiability of both the reduced parameters and the constraints to establish sufficiency. Furthermore, Fisher (1966, Theorem 6.4.1) proposes a similar argument with a stronger assumption that θ_0 is a normal point rather than a regular point [3]. However, since the rank of e_1 is constant, we believe that θ_0 is normal if and only if it is regular. Besides, we state that the regularity assumption plays a role only in the establishment of necessity and can be dropped off for sufficiency in comparison to Fisher (1966) who argues the reverse. In addition, (4.6.22) also generalizes the conclusion by Richmond (1974, Theorem 3) in a linear SEMs setup; see Dufour and Liang (2013*a*). Moreover, the conclusion of Bekker and Wansbeek (2001, Corollary 1) can also be derived from (4.6.22). Suppose

$$D(\theta_0)\mathscr{K}(\theta_0) = 0,$$

where

$$D(\theta_0) \equiv \begin{bmatrix} J_\psi(\theta_0) \\ J_\phi(\theta_0) \end{bmatrix}$$

and $\mathscr{K}(\theta_0)$ is a basis of $\ker(D(\theta_0))$. For notation ease, drop off the argument θ_0 from

[3] Recall that θ_0 is a normal point of matrix $M(\theta)$ if and only if it is a regular point of both $M(\theta)$ and $M_{(1)}(\theta)$ which consists of the columns of $M(\theta)$ except the first one.

$D(\theta_0)$. If $e_1'\mathscr{K}(\theta_0) = 0$, then $\mathscr{K}(\theta_0)$ is also the basis of $\ker \begin{bmatrix} D \\ e_1' \end{bmatrix}$ which implies that

$$\ker \begin{bmatrix} D \\ e_1' \end{bmatrix} = \ker(D).$$

From the fundamental link between the columns and rows

$$\dim\left(\operatorname{col}\begin{bmatrix} D \\ e_1' \end{bmatrix}\right) + \dim\left(\ker\begin{bmatrix} D \\ e_1' \end{bmatrix}\right) = 2G^2 + GK = \dim(\operatorname{col}(D)) + \dim(\ker(D)),$$

it follows that

$$\operatorname{rank}\begin{bmatrix} D \\ e_1' \end{bmatrix} = \operatorname{rank}(D)$$

which is (4.6.22). The reverse holds trivially since

$$\operatorname{rank}\begin{bmatrix} D \\ e_1' \end{bmatrix} = \operatorname{rank}(D)$$

implies that

$$\dim\left(\ker\begin{bmatrix} D \\ e_1' \end{bmatrix}\right) = \dim(\ker(D)).$$

Due to the fact

$$\ker\begin{bmatrix} D \\ e_1' \end{bmatrix} \subseteq \ker(D),$$

it follows that

$$\ker\begin{bmatrix} D \\ e_1' \end{bmatrix} = \ker(D).$$

Hence

$$D\mathscr{K}(\theta_0) = 0$$

entails

$$e_1'\mathscr{K}(\theta_0) = 0.$$

Note that we can show that (4.6.22) makes the result of Theorem **4.2.12** both necessary and sufficient. Denote

$$D_1(\theta_0) = \begin{bmatrix} J_\psi(\theta_0) \\ J_\phi(\theta_0) \\ e_1' \end{bmatrix}.$$

Proposition 4.6.4 NECESSARY AND SUFFICIENT CONDITION FOR LOCAL IDENTIFICATION OF A SCALAR PARAMETER. *Suppose the assumptions of Theorem* **4.2.10** *are satisfied. Let $D_{(i)}(\theta)$ be the Jacobian matrix $D(\theta)$ without the i-th column. Let θ_0 be a regular point of of both $D_{(i)}(\theta)$ and $D(\theta)$.Then θ_i is locally identifiable at $\theta_{0,i}$ if and only if* $\operatorname{rank}\left(D_{(i)}(\theta)\right) < \operatorname{rank}\left(D(\theta)\right)$.

As shown above, all the classical conditions of either global (local) identification and function identification in the SEMs under the standard assumptions are special cases of our general results in Section 4.2 and Section 4.4, particularly the equivalent statements related to the rank conditions in subsection 4.4.2. As long as $\gamma(\theta)$ is identifiable and the continuous Jacobian matrix satisfies (4.4.8) or any other equivalent condition, we achieve sufficient conditions for identification (either local identification or global identification based on stronger assumptions) of an arbitrary parametric function $\beta(\theta)$. If θ_0 is further assumed to be regular, these general conditions become both necessary and sufficient. As discussed hereinabove, if $\beta(\theta)$ is set to be θ, our general identification conditions guarantee identification of θ. Indeed, $J_\beta(\theta_0)$ turns out to be identity matrix and the linear combination $J_\beta(\theta_0)\theta$ is still θ. In addition, if $\beta(\theta)$ is the i-th element of θ, the corresponding Jacobian matrix becomes a unit vector e_i. More often, empirical studies may be interested in the identification of some linear combinations of θ, for instance $\theta_1+\theta_2$. Then the Jacobian matrix is the vector $\begin{bmatrix} 1 & 1 & 0 & \dots & 0 \end{bmatrix}$ which can be plugged into any identification condition as $J_\beta(\theta_0)$ in subsection 4.4.2. Provided that we are interested in a group of nonlinear functions of θ, the identifiability can be easily checked by comparing the ranks of two matrices $J_\gamma(\theta_0)$ and $\begin{bmatrix} J_\gamma(\theta_0) \\ J_\beta(\theta_0) \end{bmatrix}$. If the two ranks are equal, the interested nonlinear functions are identifiable. They cannot be identified otherwise.

4.6.2. The DSGE models

The DSGE models are usually solved numerically without serious verification of parameter identification. Since the establishment of our identification conditions is not limited to the model specifications and the dimension of parameter space, they are applicable to the highly dimensional and highly nonlinear DSGE models. Let's consider the state-space form of the DSGE models [see An and Schorfheide (2007)]:

$$X_{t+1} = A(\theta)X_t + B(\theta)u_{t+1} \tag{4.6.23}$$

and

$$Y_{t+1} = C(\theta)X_t + D(\theta)u_{t+1}, \tag{4.6.24}$$

where θ is a $k \times 1$ vector of deep or structural parameters, u is the structural shocks, A is $n \times n$, B is $n \times m$, C is $p \times n$ and D is $p \times m$. These coefficients are the nonlinear functions of θ. Assume

$$\mathbb{E}(u_t) = 0$$

and

$$\mathbb{E}(u_t u_s) = \sigma^2_{t-s}\Sigma(\theta),$$

where $\Sigma(\theta)$ is symmetric positive definite. The Cholesky decomposition leads to

$$\Sigma(\theta) = L(\theta)L(\theta)'.$$

We call (4.6.23) the state equation and (4.6.24) the measurement equation. The state vector X may be partially observable or totally unobservable while the measurement vector Y can be observed. Such a state space system can be solved by the Kalman filter. Clearly, in the context of linearized DSGE models, we are dealing with the local identification. Therefore, the necessary and sufficient identification conditions hereinafter are only local.

Suppose θ_0 is the true parameter of the state space model and that the statistical properties of the vector of innovations u_t in the state space representation (4.6.23) and (4.6.24) can be fully characterized by its first two moments which are continuously differentiable in the neighborhood of θ_0. We also assume that θ_0 is a regular point. Let's denote the mean of the $p \times 1$ observation Y_t as

$$\mathbb{E}(Y_t) = \mu_Y(\theta)$$

and the covariance matrix of the vectorized $\text{vec}(\bar{Y}_T)$ as

$$\mathbb{E}(\text{vec}\bar{Y}_T(\text{vec}\bar{Y}_T)') = \Sigma_T(\theta),$$

where $\bar{Y}_T = [\ Y_1 \quad Y_2 \quad \dots \quad Y_T\ ]$. Note that both $\mu_Y(\theta)$ and $\Sigma_T(\theta)$ are also functions of the A, B, C, D coefficients. Next let's denote the vector that consists of the distinguished values of all the first and second moments of the observations as $\delta(\theta)$ which is of length $(T-1)p^2 + \frac{1}{2}p(p+1)$. Since the probability distribution of Y_t depends on θ only through $\delta(\theta)$, $J_\delta(\theta_0)\theta$ is locally identifiable at θ_0 by Proposition **4.4.11**.

If we are interested in identification of the entire structural parameter θ, we can set $\gamma(\theta)$ and $\beta(\theta)$ in Theorem **4.4.3** to be $\delta(\theta)$ and θ respectively and the selection matrix $J_\beta(\theta_0)$ becomes the identity matrix. Therefore θ is locally identifiable at θ_0 provided that (4.4.1) holds, i.e.,

$$\operatorname{rank}\begin{bmatrix} J_\delta(\theta_0) \\ I_k \end{bmatrix} = \operatorname{rank}\left[J_\delta(\theta_0)\right], \tag{4.6.25}$$

which implies

$$\operatorname{rank}\left[J_\delta(\theta_0)\right] = k. \tag{4.6.26}$$

If we relax the assumption that the distribution of u_t can be fully characterized by its first two moments, (4.6.25) is only sufficient and not necessary; see Iskrev (2010). On the other hand, the argument by Iskrev (2010, Theorem 2) is different from Theorem **4.2.15** which is generalized by Theorem **4.4.3** or (4.6.25). First, we do not assume a complete separation of the mean of the structural shocks from their variance. Actually they can be dependent upon each other. Second, the assumption that θ_0 is a regular point of the Jacobian matrix is indispensable to establish the necessity of the identification condition. Third, differentiability of $\delta(\theta)$ at θ_0 will suffice to show sufficiency whereas continuous differentiability of $\delta(\theta)$ in the neighborhood of θ_0 is only required for necessity.

Remark 4.6.5 Theorem **4.4.3** does not impose full rank restriction on the Jacobian matrix $J_\delta(\theta_0)$. If $J_\delta(\theta_0)$ has deficient rank, θ is not locally identifiable at θ_0 by Theorem **4.2.15**. But we can reach the same conclusion by checking whether (4.6.25) holds. Plugging the deficient rank matrix $J_\delta(\theta_0)$ and the identity matrix I_k on both sides of (4.6.25) leads to contradiction.

Clearly, if we are interested in identifying any scalar parameter or an arbitrary linear combination of parameters of the DSGE models, (4.4.8) provides a straightforward criterion to do so. For instance, suppose we want to identify the first scalar structural parameter θ_1. Then the necessary and sufficient condition is

$$\operatorname{rank}\begin{bmatrix} J_\delta(\theta_0) \\ e_1' \end{bmatrix} = \operatorname{rank}\left[J_\delta(\theta_0)\right]. \tag{4.6.27}$$

Recently Komunjer and Ng (2011) discuss the necessary and sufficient condition of parametric function identification in the DSGE models by combining the spectral factorization and similarity transformation. We now show that such a condition for function identification is indeed a special case of our general rank condition and its equivalent state-

ments based on their assumptions; see Komunjer and Ng (2011). Let's denote

$$\xi(\theta) = \begin{pmatrix} \text{vec}A(\theta) \\ \text{vec}B(\theta) \\ \text{vec}C(\theta) \\ \text{vec}D(\theta) \\ \text{vec}\Sigma(\theta) \end{pmatrix}.$$

If Y_t is weak stationary, we can solve the state space system of the DSGE models and derive the $VMA(\infty)$ representation of Y_t as

$$Y_t = \left(C(\theta)(aI - A(\theta))^{-1}B(\theta) + D(\theta)\right)u_t.$$

Let's denote the transfer function as g such that

$$g(a;\theta) = C(\theta)(aI - A(\theta))^{-1}B(\theta) + D(\theta).$$

Then we can write the spectral density Y as

$$f_Y(a;\theta) = g(a;\theta)\Sigma(\theta)g(a^{-1};\theta)^{-1}.$$

Also denote $\Lambda(\theta)$ as the Jacobian of the vector

$$\lambda(\theta,\, S,\, T) = \begin{pmatrix} \text{vec}\left(SA(\theta)S^{-1}\right) \\ \text{vec}(SB(\theta)T) \\ \text{vec}\left(C(\theta)S^{-1}\right) \\ \text{vec}(D(\theta)T) \\ \text{vech}\left(T^{-1}\Sigma(\theta)T^{-1'}\right) \end{pmatrix}.$$

S is a $n \times n$ non-singular matrix that links two sets of coefficients A, B, C, D in a similar transformation so that the transfer function $g(a;\theta)$ has equal values at θ_1 and θ_2. That is to say,

$$(A(\theta_1), B(\theta_1), C(\theta_1), D(\theta_1)) = \left(SA(\theta_2)S^{-1}, SB(\theta_2), C(\theta_2)S^{-1}, D(\theta_2)\right) \text{ for some } S$$

if and only if

$$g(a;\theta_1) = g(a;\theta_2).$$

T is a $m \times m$ non-singular matrix which is equal to $L(\theta_2)VL(\theta_1)^{-1}$, where $L(\theta)$ is the Cholesky decomposition of $\Sigma(\theta)$ and V is an orthogonal matrix such that the following holds

$$g(a;\theta_1)L(\theta_1) = g(a;\theta_2)L(\theta_2)V.$$

Accordingly, denote $\Lambda(\theta_0)$ as the Jacobian matrix of $\lambda(\theta,\ S,\ T)$ evaluated at a triplet $(\theta_0,\ I_n,\ I_m)$. We simplify the notation by combining the elements of S and T as

$$\widetilde{W} = \begin{bmatrix} S & O \\ O & T \end{bmatrix}$$

and

$$\mathrm{vec}W = \begin{bmatrix} \mathrm{vec}S \\ \mathrm{vec}T \end{bmatrix},$$

where vec$W$ is a $j \times 1$ vector, $j = n^2 + m^2$ and denote the Jacobian matrix with respect to vec$W$ as $J_{\lambda,W}(\theta)$. So the first order partial derivatives with respect to θ and vec$W$ evaluated at $(\theta_0,\ I_j)$ can be written as

$$\Lambda(\theta_0) = \begin{bmatrix} J_{\lambda,\theta}(\theta_0,\ I_j) & J_{\lambda,W}(\theta_0,\ I_j) \end{bmatrix},$$

where $j = n+m$. Suppose want to identify all deep parameters θ. Due to Komunjer and Ng (2011, Proposition 1-S, Lemma 2-S), local identification of θ at θ_0 is equivalent to the existence of a unique solution

$$[\ \theta \quad \widetilde{W}\ ] = [\ \theta_0 \quad I_l\ ],\ l = n+m$$

to

$$\begin{pmatrix} \mathrm{vec}\left(SA(\theta)S^{-1}\right) \\ \mathrm{vec}(SB(\theta)T) \\ \mathrm{vec}\left(C(\theta)S^{-1}\right) \\ \mathrm{vec}(D(\theta)T) \\ \mathrm{vech}\left(T^{-1}\Sigma(\theta)T^{-1'}\right) \end{pmatrix} = \begin{pmatrix} \mathrm{vec}A(\theta_0) \\ \mathrm{vec}B(\theta_0) \\ \mathrm{vec}C(\theta_0) \\ \mathrm{vec}D(\theta_0) \\ \mathrm{vec}\Sigma(\theta_0) \end{pmatrix}.$$

Hence if we set

$$\beta(\theta) = \begin{bmatrix} \theta \\ \mathrm{vec}W \end{bmatrix},$$

it follows from Proposition **4.4.11** and the general rank condition (4.4.8) that the necessary

and sufficient condition for local identification of θ is

$$\text{rank}\begin{bmatrix} J_{\lambda,\theta}(\theta_0, I_j) & J_{\lambda,W}(\theta_0, I_j) \\ I_k & O \\ O & I_j \end{bmatrix} = \text{rank}\begin{bmatrix} J_{\lambda,\theta}(\theta_0, I_j) & J_{\lambda,W}(\theta_0, I_j) \end{bmatrix}, \quad (4.6.28)$$

which entails that

$$\text{rank}\begin{bmatrix} J_{\lambda,\theta}(\theta_0, I_j) & J_{\lambda,W}(\theta_0, I_j) \end{bmatrix} = k + j. \quad (4.6.29)$$

On the other hand, our rank condition (4.4.8) and its equivalent arguments also make it easy to check the identification of any part of deep parameters. Let's denote θ_q as a q dimension subvector of θ that need to be identified and θ_{k-q} the rest of the elements. Without loss of generality, we can always rearrange the components of θ such that the first q elements of θ is θ_q. We write the partial derivatives with respect to θ_q and θ_{k-q} as $J_{\lambda,\theta_q}(\theta)$ and $J_{\lambda,\theta_{k-q}}(\theta)$ respectively. If we set

$$\beta(\theta) = \begin{bmatrix} \theta_q \\ \text{vec}W \end{bmatrix},$$

the necessary and sufficient condition for identification of θ_q becomes

$$\begin{aligned} &\text{rank}\begin{bmatrix} J_{\lambda,\theta_q}(\theta_0, I_j) & J_{\lambda,\theta_{k-q}}(\theta_0, I_j) & J_{\lambda,W}(\theta_0, I_j) \\ I_q & O & O \\ O & O & I_j \end{bmatrix} \\ &= \text{rank}\begin{bmatrix} J_{\lambda,\theta_q}(\theta_0, I_j) & J_{\lambda,\theta_{k-q}}(\theta_0, I_j) & J_{\lambda,W}(\theta_0, I_j) \end{bmatrix} \end{aligned} \quad (4.6.30)$$

and this leads to

$$\text{rank}\begin{bmatrix} J_{\lambda,\theta_q}(\theta_0, I_j) & J_{\lambda,\theta_{k-q}}(\theta_0, I_j) & J_{\lambda,W}(\theta_0, I_j) \end{bmatrix} = (q + j) + \text{rank}\left(J_{\lambda,\theta_{k-q}}(\theta_0, I_j)\right). \quad (4.6.31)$$

The rank conditions (4.6.29) and (4.6.31) are just special cases of (4.4.8) when we either choose $\beta(\theta)$ to be $\begin{bmatrix} \theta' & (\text{vec}W)' \end{bmatrix}'$ to identify θ or $\begin{bmatrix} \theta_q' & (\text{vec}W)' \end{bmatrix}'$ to identify θ_q. These results are similar to Komunjer and Ng (2011, Proposition 2-S, Proposition 4). However, we only assume continuous differentiability of $\xi(\theta)$ in a neighborhood of θ_0 rather than the entire parameter space since we are concerned about local identification in the DSGE models. Besides, the validity of sufficiency of the above identification conditions does not depend on continuous differentiability of $\xi(\theta)$ in the neighborhood of θ. Differentiability at θ_0 will suffice. Furthermore, Komunjer and Ng (2011) also discusses

identification conditions for a nonsingular system when the left invertibility assumption fails. Our results can also generalize their conclusions in this setup; see Komunjer and Ng (2011, Proposition 2-NS, Proposition 3). Nevertheless, identification through the spectral density employs information only about the second moments of the second order stationary observations. Thus compared to the classical approach that relies on the first two moments for identification of deep parameters in the DSGE models, such a method comes with a cost of imposing a series of much stronger assumptions, such as left-invertibility and/or minimality of the system. Clearly if the probability distribution of the underlying process does involve the first moment, the identification conditions established from the spectral density could become only sufficient and not necessary.

It is trivial that we can choose to identify any scalar parameter of θ or any linear combinations of θ as long as the selection matrix $J_\lambda(\theta_0)$ can satisfy our general rank condition (4.4.8) or other relevant identification conditions established hereinabove. For instance, the condition for identification of the first element θ_1 should be

$$\begin{aligned}
&\operatorname{rank}\begin{bmatrix} J_{\lambda,\theta_1}(\theta_0,\, I_j) & J_{\lambda,\theta_{k-1}}(\theta_0,\, I_j) & J_{\lambda,W}(\theta_0,\, I_j) \\ 1 & O & O \\ O & O & I_j \end{bmatrix} \\
&= \operatorname{rank}\begin{bmatrix} J_{\lambda,\theta_1}(\theta_0,\, I_j) & J_{\lambda,\theta_{k-1}}(\theta_0,\, I_j) & J_{\lambda,W}(\theta_0,\, I_j) \end{bmatrix} \\
&= (1+j) + \operatorname{rank}\left(J_{\lambda,\theta_{k-1}}(\theta_0,\, I_j)\right).
\end{aligned}$$

Meanwhile, we can also identify any linear combinations $l'\theta$ using (4.4.8)

$$\operatorname{rank}\begin{bmatrix} J_{\lambda,\theta}(\theta_0,\, I_j) & J_{\lambda,W}(\theta_0,\, I_j) \\ l' & O \\ O & I_j \end{bmatrix} = \operatorname{rank}\begin{bmatrix} J_{\lambda,\theta}(\theta_0,\, I_j) & J_{\lambda,W}(\theta_0,\, I_j) \end{bmatrix}. \tag{4.6.32}$$

4.7. Three examples

We look at three macroeconomic models and show that it is convenient to apply the results on parametric function identification in Section 4.4. The first two examples help demonstrate the usefulness of the proposed identification condition through identification failure in two macroeconomic models. The third one is about both weak identification and identification failure.

4.7.1. A two-equation SEMs example

We now demonstrate the results on parametric function identification through a two-equation SEMs [see Wooldridge (2006)]. Romer (1993) proposes and tests the theory that inflation rates will be higher for less "open" countries using 114 countries data. The openness denoted as "Open" is measured by the average share of imports in gross domestic or national product since 1973 and inflation denoted as "Inf" is measured as the average annual change in the log of GDP or log of GNP deflator since 1973. Besides Romer (1993) also uses instrumental variables for estimation, some of which are defined as follows:

$$\text{Rincpc} = \text{real income per capita for1980; Land} = \text{a country's land area (square miles);}$$

$$\text{Oil} = \begin{cases} 1 & \text{a country is a major oil producer} \\ 0 & \text{for woman} \end{cases};$$

$$\text{Good} = \begin{cases} 1 & \text{a country whose national accounts data receive a rating of C or better} \\ 0 & \text{otherwise} \end{cases}.$$

The identification failure of linear models caused by multicollinearity in the dummy variable case has been discussed in detail by Dufour and Liang (2012). Thus we hereby focus on identification failure of SEMs. We specify the following two-equation system:

$$Inf = \alpha_0 + \alpha_1 Open + \alpha_2 \log(Rincpc) + u_1 \tag{4.7.1}$$

$$Open = \beta_0 + \beta_1 Inf + \beta_2 \log(Rincpc) + \beta_3 \log(Land) + u_2. \tag{4.7.2}$$

Assume that $\alpha_1\beta_1 \neq 1$ and we have the reduced form equations:

$$Inf = \eta_0 + \eta_1 \log(Rincpc) + \eta_2 \log(Land) + w_1 \tag{4.7.3}$$

$$Open = \lambda_0 + \lambda_1 \log(Rincpc) + \lambda_2 \log(Land) + w_2, \tag{4.7.4}$$

where

$$\eta_0 = \frac{\alpha_0 + \alpha_1\beta_0}{1-\alpha_1\beta_1},\ \eta_1 = \frac{\alpha_1\beta_2 + \alpha_2}{1-\alpha_1\beta_1},\ \eta_2 = \frac{\alpha_1\beta_3}{1-\alpha_1\beta_1},\ w_1 = \frac{u_1 + \alpha_1 u_2}{1-\alpha_1\beta_1}$$

and

$$\lambda_0 = \frac{\alpha_0\beta_1 + \beta_0}{1-\alpha_1\beta_1},\ \lambda_1 = \frac{\alpha_2\beta_1 + \beta_2}{1-\alpha_1\beta_1},\ \lambda_2 = \frac{\beta_3}{1-\alpha_1\beta_1},\ w_2 = \frac{\beta_1 u_1 + u_2}{1-\alpha_1\beta_1}.$$

Now let θ, $\gamma(\theta)$, $\beta_{Inf}(\theta)$ and $\beta_{Open}(\theta)$ be the structural parameters, reduced form param-

eters and parameters of interest respectively and set

$$\theta = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \gamma(\theta) = \begin{bmatrix} \eta_0 \\ \eta_1 \\ \eta_2 \\ \lambda_0 \\ \lambda_1 \\ \lambda_2 \end{bmatrix}, \beta_{Inf}(\theta) = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}, \beta_{Open}(\theta) = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}.$$

The Jacobian matrices can be written as

$$J_\gamma(\theta) = \frac{1}{1-\alpha_1\beta_1} \begin{bmatrix} 1 & \frac{\alpha_0\beta_1+\beta_0}{1-\alpha_1\beta_1} & 0 & \alpha_1 & \frac{(\alpha_0+\alpha_1\beta_0)\alpha_1}{1-\alpha_1\beta_1} & 0 & 0 \\ 0 & \frac{\alpha_2\beta_1+\beta_2}{1-\alpha_1\beta_1} & 1 & 0 & \frac{(\alpha_1\beta_2+\alpha_2)\alpha_1}{1-\alpha_1\beta_1} & \alpha_1 & 0 \\ 0 & \frac{\beta_3}{1-\alpha_1\beta_1} & 0 & 0 & \frac{{\alpha_1}^2\beta_3}{1-\alpha_1\beta_1} & 0 & \alpha_1 \\ \beta_1 & \frac{(\alpha_0\beta_1+\beta_0)\beta_1}{1-\alpha_1\beta_1} & 0 & 1 & \frac{\alpha_0+\alpha_1\beta_0}{1-\alpha_1\beta_1} & 0 & 0 \\ 0 & \frac{(\alpha_2\beta_1+\beta_2)\beta_1}{1-\alpha_1\beta_1} & \beta_1 & 0 & \frac{\alpha_2+\alpha_1\beta_2}{1-\alpha_1\beta_1} & 1 & 0 \\ 0 & \frac{\beta_1\beta_3}{1-\alpha_1\beta_1} & 0 & 0 & \frac{\alpha_1\beta_3}{1-\alpha_1\beta_1} & 0 & 1 \end{bmatrix},$$

$$J_{\beta_{Inf}}(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}, J_{\beta_{Open}}(\theta) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ & & & & & & \end{bmatrix}.$$

Assume $\beta_3 \neq 0$. We calculate the ranks of the following matrices

$$\operatorname{rank}\big(J_\gamma(\theta)\big) = 6, \operatorname{rank}\begin{bmatrix} J_\gamma(\theta) \\ J_{\beta_{Inf}}(\theta) \end{bmatrix} = 6, \operatorname{rank}\begin{bmatrix} J_\gamma(\theta) \\ J_{\beta_{Open}}(\theta) \end{bmatrix} = 7.$$

From (4.4.8), we immediately know that (4.7.1) is identifiable and (4.7.2) is unidentifiable. In fact, none of the parameters in (4.7.2) can be identified.

Next assume $\beta_3 = 0$. Then $\theta_1 = \begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2 & \beta_0 & \beta_1 & \beta_2 & 0 \end{bmatrix}'$ is an irregular

point of $J_\gamma(\theta)$ since

$$\operatorname{rank}\left(J_\gamma(\theta_1)\right)=\frac{1}{1-\alpha_1\beta_1}\begin{bmatrix} 1 & \frac{\alpha_0\beta_1+\beta_0}{1-\alpha_1\beta_1} & 0 & \alpha_1 & \frac{(\alpha_0+\alpha_1\beta_0)\alpha_1}{1-\alpha_1\beta_1} & 0 & 0 \\ 0 & \frac{\alpha_2\beta_1+\beta_2}{1-\alpha_1\beta_1} & 1 & 0 & \frac{(\alpha_1\beta_2+\alpha_2)\alpha_1}{1-\alpha_1\beta_1} & \alpha_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \alpha_1 \\ \beta_1 & \frac{(\beta_0+\alpha_0\beta_1)\beta_1}{1-\alpha_1\beta_1} & 0 & 1 & \frac{\alpha_0+\alpha_1\beta_0}{1-\alpha_1\beta_1} & 0 & 0 \\ 0 & \frac{(\alpha_2\beta_1+\beta_2)\beta_1}{1-\alpha_1\beta_1} & \beta_1 & 0 & \frac{\alpha_2+\alpha_1\beta_2}{1-\alpha_1\beta_1} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}=5.$$

Meanwhile, we have

$$\operatorname{rank}\begin{bmatrix} J_\gamma(\theta_1) \\ J_{\beta_{Inf}}(\theta_1) \end{bmatrix}=6$$

and

$$\operatorname{rank}\begin{bmatrix} J_\gamma(\theta_1) \\ J_{\beta_{Open}}(\theta_1) \end{bmatrix}=6.$$

This leads to the conclusion that both (4.7.1) and (4.7.2) are not identifiable since the general rank condition is not satisfied. However, we should bear in mind that when the Jacobian matrix is evaluated at irregular points, our general rank condition is only sufficient and not necessary. Thus, it is possible that certain parameters of interest can still be identified even though the general rank condition is violated when Jacobian matrices are evaluated at irregular points. We will give an example at the end of next subsection.

4.7.2. A three-equation New Keynesian model

Let y, π and R be the output gap, inflation rate and nominal interest rate set by the central bank, we write the standard three-equation New Keynesian model in the rational expectations system after log-linearization as follows [see also Smets and Wouters (2003), An and Schorfheide (2007), Nason and Smith (2008) and Canova (2009)]:

$$y_t=\frac{h}{1+h}y_{t-1}+\frac{1}{1+h}\mathbb{E}_t y_{t+1}+\frac{1}{\tau}(R_t-\mathbb{E}_t\pi_{t+1})+\varepsilon_{yt} \tag{4.7.5}$$

$$\pi_t=\frac{\psi}{1+\psi\beta}\pi_{t-1}+\frac{\beta}{1+\psi\beta}\mathbb{E}_t\pi_{t+1}+\frac{(\tau+\nu)(1-\xi\beta)(1-\xi)}{(1+\psi\beta)\xi}y_t+\varepsilon_{\pi t} \tag{4.7.6}$$

$$R_t=\phi_R R_{t-1}+(1-\phi_R)(\phi_\pi\pi_{t-1}+\phi_y y_{t-1})+\varepsilon_{Rt}, \tag{4.7.7}$$

where h is the degree of habit persistence and measured by a portion of consumption of previous period [4], τ is the coefficient of relative risk aversion of households or the inverse of the intertemporal elasticity of substitution, β is the discount factor, ψ is the degree of price indexation [5], ξ is the degree of price stickiness [6], ν is the elasticity of labor supply, ϕ_R, ϕ_π and ϕ_y are the coefficients of the Taylor rule. Also we assume that ε_{yt} and $\varepsilon_{\pi t}$ follow an AR(1) process with coefficients ρ_π and ρ_y and the monetary policy shock is serially uncorrelated. Denote the variances of the shocks as σ_y^2, σ_π^2 and σ_R^2. Note that (4.7.5) is the dynamic IS schedule and (4.7.6) is the New Keynesian Phillips curve.

We can write out the structural parameters as

$$\theta_{strc} = \begin{bmatrix} h & \tau & \psi & \beta & \nu & \xi & \phi_R & \phi_\pi & \phi_y & \rho_\pi & \rho_y & \sigma_y^2 & \sigma_\pi^2 & \sigma_R^2 \end{bmatrix}'.$$

For simplicity, we focus on identification of only economic parameters θ^1 and ignore the auxiliary parameters θ^2, where

$$\theta^1 = \begin{bmatrix} h & \tau & \psi & \beta & \nu & \xi & \phi_R & \phi_\pi & \phi_y \end{bmatrix}'$$
$$\theta^2 = \begin{bmatrix} \rho_\pi & \rho_y & \sigma_y^2 & \sigma_\pi^2 & \sigma_R^2 \end{bmatrix}'.$$

Now the semi-structural version of the above model can be written as:

$$y_t = \zeta_1 y_{t-1} + \zeta_2 \mathbb{E}_t y_{t+1} + \zeta_3 (R_t - \mathbb{E}_t \pi_{t+1}) + \varepsilon_{yt} \tag{4.7.8}$$
$$\pi_t = \delta_1 \pi_{t-1} + \delta_2 \mathbb{E}_t \pi_{t+1} + \delta_3 y_t + \varepsilon_{\pi t} \tag{4.7.9}$$
$$R_t = \kappa_1 R_{t-1} + \kappa_2 \pi_{t-1} + \kappa_3 y_{t-1} + \varepsilon_{Rt}, \tag{4.7.10}$$

where

$$\zeta_1 = \frac{h}{1+h},\ \zeta_2 = \frac{1}{1+h},\ \zeta_3 = \frac{1}{\tau},$$
$$\delta_1 = \frac{\psi}{1+\psi\beta},\ \delta_2 = \frac{\beta}{1+\psi\beta},\ \delta_3 = \frac{(\tau+\nu)(1-\xi\beta)(1-\xi)}{(1+\psi\beta)\xi},$$
$$\kappa_1 = \phi_R,\ \kappa_2 = (1-\phi_R)\phi_\pi,\ \kappa_3 = (1-\phi_R)\phi_y.$$

Suppose that the system of (4.7.8) through (4.7.10) satisfies all three conditions mentioned by Canova (2009), then the semistructural parameters are locally identifiable and we

[4] The habit feature is often omitted from the basic setups of New Keynesian models but is included in large scale DSGE models [see Canova (2009)].

[5] When $\psi = 0$, the inflation equation becomes the standard purely forward-looking Phillips curve.

[6] When all prices are flexible, $\xi = 0$.

denote them as θ_s

$$\theta_{semi} = \begin{bmatrix} \zeta_1 & \zeta_2 & \zeta_3 & \delta_1 & \delta_2 & \delta_3 & \kappa_1 & \kappa_2 & \kappa_3 & \rho_\pi & \rho_y & \sigma_y^2 & \sigma_\pi^2 & \sigma_R^2 \end{bmatrix}'.$$

We are interested in identifying the structural parameter θ, particularly the economic structural parameters θ^1. Set

$$\gamma_{semi}(\theta^1) = \begin{bmatrix} \zeta_1 & \zeta_2 & \zeta_3 & \delta_1 & \delta_2 & \delta_3 & \kappa_1 & \kappa_2 & \kappa_3 \end{bmatrix}'$$
$$\beta_{strc}(\theta^1) = \theta^1 = \begin{bmatrix} h & \tau & \psi & \beta & \nu & \xi & \phi_R & \phi_\pi & \phi_y \end{bmatrix}'.$$

Although both $\gamma_{semi}(\theta^1)$ and $\beta_{strc}(\theta^1)$ have the same number of parameters, it is not clear whether we can retrieve all the information about structural parameter $\beta_{strc}(\theta^1)$ from the semi-structural parameter $\gamma_{semi}(\theta^1)$. More interestingly, we would like to know exactly which structural parameters are identifiable and which are not, i.e., parametric function identification. We list the following Jacobian matrices for convenience.

$$J_{\gamma_{semi}}(\theta^1) = \begin{bmatrix} B_1 & O \\ O & B_2 \end{bmatrix},$$

where

$$B_1 = \begin{bmatrix} \frac{1}{(1+h)^2} & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{(1+h)^2} & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{\tau^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{(1+\psi\beta)^2} & -\frac{\psi^2}{(1+\psi\beta)^2} & 0 & 0 \\ 0 & 0 & -\frac{\beta^2}{(1+\psi\beta)^2} & \frac{1}{(1+\psi\beta)^2} & 0 & 0 \\ 0 & \frac{(1-\xi\beta)(1-\xi)}{(1+\psi\beta)\xi} & -\frac{\beta(\tau+\nu)(1-\xi\beta)(1-\xi)}{(1+\psi\beta)^2\xi} & -\frac{(\tau+\nu)(1-\xi)(\psi+\xi)}{(1+\psi\beta)^2\xi} & \frac{(1-\xi\beta)(1-\xi)}{(1+\psi\beta)\xi} & -\frac{(\tau+\nu)(1-\xi^2\beta)}{(1+\psi\beta)\xi^2} \end{bmatrix}$$

and

$$B_2 = \begin{bmatrix} 1 & 0 & 0 \\ -\phi_\pi & 1-\phi_R & 0 \\ -\phi_y & 0 & 1-\phi_R \end{bmatrix}.$$

Clearly

$$\operatorname{rank}\left(J_{\gamma_{semi}}(\theta^1)\right) = \operatorname{rank}(B_1) + \operatorname{rank}(B_2) = 8.$$

Now we check identifiability of each of the economic parameters that have interesting

economic interpretations. Denote e_i, $i = 1,2,\ldots,9$, be a unit vector whose i-th element is 1 and 0 otherwise. The matrix operations lead to the following results:

$$\text{rank}\begin{bmatrix} \left(J_{\gamma_{semi}}(\theta^1)\right) \\ {e_i}' \end{bmatrix} = 8 = \text{rank}\left(J_{\gamma_{semi}}(\theta^1)\right),\ i = 1,2,3,4,7,8,9$$

and

$$\text{rank}\begin{bmatrix} \left(J_{\gamma_{semi}}(\theta^1)\right) \\ {e_i}' \end{bmatrix} = 9 \neq \text{rank}\left(J_{\gamma_{semi}}(\theta^1)\right),\ i = 5,6.$$

Since ${e_i}'$ is the Jacobian matrix of the ith element of θ^1 with respect to ${\theta^1}'$, we immediately conclude that the structural parameters h, τ, ψ, β, ϕ_R, ϕ_π and ϕ_y are identifiable. Both the elasticity of labor supply ν and the degree of price stickiness ξ are not identifiable.

We are the first to study identification failure in the New Keynesian model using our general rank conditions. Although Canova (2009) focuses on the same simplified three-equation New Keynesian model and points out that ν and ξ are not locally identifiable, no explanations are provided on why only these two parameters are not identifiable. We hereby provide a closed-form analysis and demonstrate that identification problem with this type of dynamic New Keynesian model is a simple application of the rank condition. Furthermore, despite the fact that ν and ξ enter the slope of the Phillips curve in a multiplicative way, identification of one does not necessarily guarantee identification of the other. For instance, we impose additional assumption that there is no price flexibility in the system, i.e., $\xi = 1$ and we want to identify ν without further relative information. Note that $\theta_0^1 = \begin{bmatrix} h & \tau & \psi & \beta & \nu & 1 & \phi_R & \phi_\pi & \phi_y \end{bmatrix}'$ is a regular point of $J_{\gamma_{semi}}(\theta^1)$ since $\text{rank}\left(J_{\gamma_{semi}}(\theta_0^1)\right) = 8$ around the neighborhood of θ_0^1. Then we have

$$\text{rank}\begin{bmatrix} \left(J_{\gamma_{semi}}(\theta_0^1)\right) \\ {e_5}' \end{bmatrix} = 9 \neq \text{rank}\left(J_{\gamma_{semi}}(\theta_0^1)\right)$$

and conclude that ν is not identifiable.

Moreover, if the previous period of interest rate has an one-to-one effect on current interest rate, i.e., when $\phi_R = 1$ we have

$$\text{rank}\left(J_{\gamma_{semi}}(\theta_0^1)\right) = 6$$

and

$$\text{rank}\begin{bmatrix} \left(J_{\gamma_{semi}}(\theta_0^1)\right) \\ {e_i}' \end{bmatrix} = 7,\ i = 8,9.$$

Therefore, we may reach the conclusion that both ϕ_π and ϕ_y are not identifiable. This is obvious from the expressions of κ_2 and κ_3. However, we have to bear in mind that the necessary and sufficient conditions hold conditional on the fact that the rank of Jacobian matrices must be evaluated at regular points. Otherwise, the general rank condition is only sufficient and not necessary. Take the following case as an example. Suppose $\xi = 1$ and $\beta = 1$. Then $\theta_0^1 = \begin{bmatrix} h & \tau & \psi & 1 & \nu & 1 & \phi_R & \phi_\pi & \phi_y \end{bmatrix}'$ is an irregular point of $J_{\gamma_{semi}}(\theta^1)$. Then we get

$$\text{rank}\left(J_{\gamma_{semi}}(\theta_0^1)\right) = 7$$

and

$$\text{rank}\begin{bmatrix} \left(J_{\gamma_{semi}}(\theta_0^1)\right) \\ e_6{}' \end{bmatrix} = 8.$$

Clearly

$$\text{rank}\left(J_{\gamma_{semi}}(\theta_0^1)\right) \neq \text{rank}\begin{bmatrix} \left(J_{\gamma_{semi}}(\theta_0^1)\right) \\ e_6{}' \end{bmatrix}.$$

But we know that ξ is equal to 1 and thus identifiable. Hence the rank condition is only sufficient and not necessary.

Besides, the Jacobian matrix $J_\gamma(\theta^1)$ makes it possible to draw other important conclusions related to weak identification. Suppose ϕ_R is close to 1. Theoretically we can still identify ϕ_π and ϕ_y using the rank condition but they are now both weakly identified. Note that additional information might be helpful in identifying structural parameters, but we have to be cautious about imposing restrictions on structure parameters since it is difficult to justify these subjective assumptions as demonstrated by Carlstrom, Fuerst and Paustian (2009) in a basic Dynamic New Keynesian model.

4.7.3. New Keynesian Phillips Curves (NKPC)

The NKPC has been an important part of the standard macroeconomic policy models and different variants of the baseline NKPC have been studied by the literature. Since the proposition of the hybrid form of New Keynesian Philipps curve (NKPC) by Galí and Gertler (1999), which considers both conventionally forward-looking and backward-looking behavior, it has become increasingly popular in empirical study for modeling the evolution of inflation. However, there has been empirical evidence that both the basic NKPC and the hybrid NKPC models suffer from identifiability problems; see Dufour et al. (2006) and the references therein. We show that how the identification conditions in this paper can be used to check identification failure and weak identification in the NKPC models.

4.7.3.1. A baseline NKPC

Due to the formulation by Calvo (1983), in any given period each firm has a probability of $1-\eta$ to adjust its price, a portion η of firms will not be able to change prices each period. The baseline NKPC specifies the current inflation rate as a function of current marginal cost and expected inflation rate of the next period. It takes the following econometric form:

$$\pi_t = \lambda s_t + \beta \mathbb{E}_t \pi_{t+1} + u_{t+1}, \tag{4.7.11}$$

where s_t is the real marginal cost and usually measured by the output gap, β is the subjective discount factor and $\lambda = (1-\eta)(1-\beta\eta)/\eta$. Clearly we can denote the deep parameters and semi-structural parameters as

$$\theta = \begin{bmatrix} \eta \\ \beta \end{bmatrix}, \ \gamma(\theta) = \begin{bmatrix} \lambda \\ \beta \end{bmatrix}.$$

To establish identification condition for η, we calculate the Jacobian matrices of $\gamma(\theta)$ and η with respect to θ as follows;

$$J_\gamma(\theta) = \begin{bmatrix} \frac{\beta\eta^2-1}{\eta^2} & \eta-1 \\ 0 & 1 \end{bmatrix}, \ J_\eta(\theta) = \begin{bmatrix} 1 & 0 \end{bmatrix}.$$

From Proposition **4.4.12**, one necessary and sufficient condition for identification of η provided that $\gamma(\theta)$ is identifiable is

$$\text{Im} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \subseteq \text{Im} \begin{bmatrix} \frac{\beta\eta^2-1}{\eta^2} & 0 \\ \eta-1 & 1 \end{bmatrix}. \tag{4.7.12}$$

Obviously (4.7.12) holds if and only if

$$\beta\eta^2 - 1 \neq 0. \tag{4.7.13}$$

Note that (4.7.13) is also a necessary and sufficient for $\begin{bmatrix} \eta \\ \beta \end{bmatrix}$ to be a regular point of $J_\gamma(\theta)$. Hence η is identifiable if and only if

$$\beta\eta^2 \neq 1. \tag{4.7.14}$$

Condition (4.7.14) is new to the literature and it demonstrates that imposing restrictions on deep parameters does not guarantee identification. In the baseline NKPC example, if

$\beta\eta^2 = 1$ given that β is identifiable and $\eta > 0$, η is still unidentified. Obviously, identification failure of the baseline NKPC can also occur when we retrieve information of the semi-structural parameters $\gamma(\theta)$ from the reduced formed parameters using instrumental variable regression. We will discuss this type of identification issue in the following more complicated hybrid NKPC which nests the baseline NKPC.

4.7.3.2. The Hybrid NKPC [Galí and Gertler (1999)]

Galí and Gertler (1999) assume Calvo formulation. Meanwhile they assume that a fraction ω of firms that are allowed to change prices choose to use a rule of thumb rather than to optimize. This group of firms is said to be backward-looking. For convenience, the hybrid NKPC by Galí and Gertler (1999) in an econometric setup is restated as follows:

$$\pi_t = \lambda s_t + \gamma_f \mathbb{E}_t \pi_{t+1} + \gamma_b \pi_{t-1} + u_{t+1}, \tag{4.7.15}$$

where

$$\begin{aligned}
\lambda &= \frac{(1-\omega)(1-\eta)(1-\beta\eta)}{\eta + \omega[1-\eta(1-\beta)]}, \\
\gamma_f &= \frac{\beta\eta}{\eta + \omega[1-\eta(1-\beta)]}, \\
\gamma_b &= \frac{\omega}{\eta + \omega[1-\eta(1-\beta)]}.
\end{aligned}$$

Denote the deep parameters as $\theta = \begin{bmatrix} \omega & \eta & \beta \end{bmatrix}'$ and the semi-structural parameters as $\gamma(\theta) = \begin{bmatrix} \lambda & \gamma_f & \gamma_b \end{bmatrix}'$. Since the semi-structural parameters are nonlinear functions of the deep parameters, identification of the former does not necessarily guarantee identification of the latter which we are interested in [see Ma (2002)]. If there does not exist an inverse mapping from $\gamma(\theta)$ to θ as we have shown in the three-equation New Keynesian model and the baseline NKPC, we still fail to identify structural parameters $\begin{bmatrix} \omega & \eta & \beta \end{bmatrix}'$ of the hybrid NKPC. We check if the hybrid NKPC also has such an identification problem. For simplicity , suppose we can identify all the semi-structural parameters $\begin{bmatrix} \lambda & \gamma_f & \gamma_b \end{bmatrix}'$. Then applying the general rank condition we have

$$J_\gamma(\theta) = \frac{1}{(\eta + \omega - \omega\eta + \omega\beta\eta)^2} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ -\beta\eta(1-\eta+\beta\eta) & \beta\omega & \eta(\eta+\omega-\omega\eta) \\ \eta & -\omega(1-\omega+\omega\beta) & -\omega^2\eta \end{bmatrix}$$

where

$$
\begin{aligned}
a_{11} &= -(1-\eta)(1-\beta\eta),\\
a_{12} &= (1-\omega)(-1-2\omega\beta+2\omega\beta\eta+\beta\eta^2-\omega\beta\eta^2+\omega\beta^2\eta^2),\\
a_{13} &= (1-\omega)(1-\eta)(-\eta^2-2\omega\eta+\omega\eta^2).
\end{aligned}
$$

Note that $\begin{bmatrix} \omega & 0 & 1 \end{bmatrix}'$, $\begin{bmatrix} \omega & 1 & 1 \end{bmatrix}'$ and $\begin{bmatrix} 0 & 0 & \beta \end{bmatrix}'$ are irregular points of $J_\gamma(\theta)$, $\forall\omega,\ \beta$. Thus except for irregular points,

$$
\operatorname{rank}\left(J_\gamma(\theta)\right) = \operatorname{rank}\begin{bmatrix} J_\gamma(\theta) \\ e_i' \end{bmatrix} = 3,\ i = 1,2,3.
$$

Hence there exists a inverse mapping from the semi-structural parameters to the structural parameters and θ can be identified given that $\gamma(\theta)$ is identifiable.

Nevertheless, the semi-structural parameters themselves are not always identifiable and some of them are prone to weak identification as we will show hereinafter, it is not surprising that we have a serious identification issue with the deep parameters of the hybrid NKPC in practice. Specifically, using π_{t+1} as a proxy for $\mathbb{E}_t\pi_{t+1}$, we notice that both s_t and π_{t+1} on the right hand side of (4.7.15) are endogenous variables. We can choose to use the IV regression method to deal with endogeneity issue which is intrinsic to the hybrid NKPC and achieve consistent estimators of the above semi-structural parameters. We follow Dufour et al. (2006) and use the quarterly data for the U.S. over the period of $1970:1-1997:4$. The choice of instrumental variables varies in the literature. To make the results comparable, we choose four lags of inflation, labour share, commodity-price inflation, wage inflation, the long-short interest rate and output gap as instruments (24 in total [7]) [see also Galí and Gertler (1999)]. For convenience, we express the IV equation and rewrite (4.7.15) in the matrix form with T observations as follows:

$$
y = Y\vartheta + X_1\gamma_b + u \tag{4.7.16}
$$

$$
Y = X_1\Pi_1 + X_2\Pi_2 + v, \tag{4.7.17}
$$

where y is the $T\times 1$ endogenous vector of π_t, Y is the $T\times 2$ matrix of endogenous variables $\begin{bmatrix} s_t & \pi_{t+1} \end{bmatrix}$, X_1 is the $T\times 1$ exogenous vector of π_{t-1}, X_2 is a $T\times 24$ matrix of instruments.

[7] To estimate the reduced form equations we add the constant term which can be treated as an additional instrument.

We also define the following partitioned matrices:

$$\vartheta = \begin{bmatrix} \lambda \\ \gamma_f \end{bmatrix}, \Pi_1 = \begin{bmatrix} \Pi_{11} & \Pi_{12} \end{bmatrix}, \Pi_2 = \begin{bmatrix} \Pi_{21} & \Pi_{22} \end{bmatrix}, \Pi_s = \begin{bmatrix} \Pi_{11} \\ \Pi_{21} \end{bmatrix}, \Pi_{inf} = \begin{bmatrix} \Pi_{12} \\ \Pi_{22} \end{bmatrix}.$$

Clearly Π_s and Π_{inf} are the coefficients of output gap equation and inflation equation respectively in the IV equation (4.7.17)

The reduced form equation of (4.7.16) is

$$y = X_1\varphi_1 + X_2\varphi_2 + w, \tag{4.7.18}$$

where

$$\varphi_1 = \Pi_1\vartheta + \gamma_b = \Pi_{11}\lambda + \Pi_{12}\gamma_f + \gamma_b, \tag{4.7.19}$$

$$\varphi_2 = \Pi_2\vartheta = \Pi_{21}\lambda + \Pi_{22}\gamma_f, \tag{4.7.20}$$

$$w = v\vartheta + u. \tag{4.7.21}$$

Based on our IV construction, φ_1 and φ_2 are functions of semi-structural parameters $\begin{bmatrix} \lambda & \gamma_f & \gamma_b \end{bmatrix}'$ and are identifiable. It is easy to calculate the Jacobian matrix of $\varphi = \begin{bmatrix} \varphi_1 & \varphi_2 \end{bmatrix}'$ with respect to the semi-structural parameters as

$$J_\varphi(\lambda, \gamma_f, \gamma_b) = \begin{bmatrix} \Pi_1 & 1 \\ \Pi_2 & O \end{bmatrix} = \begin{bmatrix} \Pi_{11} & \Pi_{12} & 1 \\ \Pi_{21} & \Pi_{22} & O \end{bmatrix}.$$

Thus

$$\operatorname{rank}\Big(J_\varphi(\lambda, \gamma_f, \gamma_b)\Big) = 1 + \operatorname{rank}(\Pi_2). \tag{4.7.22}$$

As can be seen, to get the rank information from (4.7.22) we can plug in the estimated the reduced form parameters $\hat{\Pi}$. We list the estimated parameter matrix, the t-statistics and p-values in Table 4.1, where π, $winf$, $cinf$, rd, $lshare$ and s stand for inflation rate, wage inflation, commodity-price inflation, the difference between long and short interest rates, labour share and output gap respectively.

TABLE 4.1
Estimation for reduced form parameters, 1970:1-1997:4

IV	$\hat{\Pi}_s$	t-stat	p-value	$\hat{\Pi}_{inf}$	t-stat	p-value
π_{t-1}	-0.1459	-0.4500	0.6538	0.0597	0.5329	0.5955
π_{t-2}	0.4369	1.3315	0.1865	-0.0379	-0.3342	0.7391
π_{t-3}	-0.3228	-0.9859	0.3270	0.1687	1.4895	0.1400
π_{t-4}	0.2917	0.9060	0.3675	0.2819	2.5313	0.0132
π_{t-5}	-0.2416	-0.8078	0.4214	-0.0063	-0.0611	0.9514
$winf_{t-1}$	-0.0783	-1.6249	0.1078	0.0692	4.1488	7.8e-05
$winf_{t-2}$	-0.0450	-0.9290	0.3555	0.0682	4.0692	1.0e-04
$winf_{t-3}$	0.0772	1.5334	0.1288	-0.0186	-1.0662	0.2893
$winf_{t-4}$	0.0476	1.0575	0.2932	-0.0264	-1.6941	0.0939
$cinf_{t-1}$	0.0028	0.5735	0.5678	0.0052	3.0881	0.0027
$cinf_{t-2}$	-0.0088	-1.8011	0.0752	0.0051	3.0076	0.0035
$cinf_{t-3}$	-7.9e-04	-0.1561	0.8764	0.0050	2.8434	0.0056
$cinf_{t-4}$	0.0025	0.4816	0.6313	9.6e-04	0.5444	0.5876
rd_{t-1}	0.1026	0.8631	0.3905	-0.0396	-0.9643	0.3376
rd_{t-2}	0.1865	1.1605	0.2490	0.0730	1.3130	0.1927
rd_{t-3}	-0.2156	-1.3258	0.1884	-0.0248	-0.4405	0.6607
rd_{t-4}	0.0725	0.5982	0.5513	0.0154	0.3684	0.7135
$lshare_{t-1}$	2.7647	2.3692	0.0201	-0.6325	-1.5671	0.1208
$lshare_{t-2}$	-1.6450	-1.0280	0.3068	-0.1065	-0.1923	0.8479
$lshare_{t-3}$	-1.8844	-1.1128	0.2689	0.9819	1.6765	0.0973
$lshare_{t-4}$	1.7304	1.4434	0.1525	-0.0609	-0.1470	0.8835
s_{t-1}	1.3590	8.8223	1.1e-13	-0.0906	-1.7012	0.0925
s_{t-2}	-0.1990	-0.8323	0.4075	-0.0100	-0.1206	0.9043
s_{t-3}	-0.2936	-1.2132	0.2284	0.1579	1.8868	0.0626
s_{t-4}	0.1038	0.6722	0.5032	-0.0322	-0.6026	0.5483

$$\hat{\Pi}_{11} = -0.1459,\ \hat{\Pi}_{12} = 0.0597,\ \hat{\Pi}_{21} = \begin{bmatrix} 0.4369 \\ -0.3228 \\ 0.2917 \\ -0.2416 \\ -0.0783 \\ -0.0450 \\ 0.0722 \\ 0.0476 \\ 0.0028 \\ -0.0088 \\ -7.9\text{e-}04 \\ 0.0025 \\ 0.1026 \\ 0.1865 \\ -0.2156 \\ 0.0725 \\ 2.7647 \\ -1.6450 \\ -1.8844 \\ 1.7304 \\ 1.3590 \\ -0.1990 \\ -0.2936 \\ 0.1038 \end{bmatrix},\ \hat{\Pi}_{22} = \begin{bmatrix} -0.0379 \\ 0.1687 \\ 0.2819 \\ -0.0063 \\ 0.0692 \\ 0.0682 \\ -0.0186 \\ -0.0264 \\ 0.0052 \\ 0.0051 \\ 0.0050 \\ 9.6\text{e-}04 \\ -0.0396 \\ 0.0730 \\ -0.0248 \\ 0.0154 \\ -0.6325 \\ -0.1065 \\ 0.9819 \\ -0.0609 \\ -0.0906 \\ -0.0100 \\ 0.1579 \\ -0.0322 \end{bmatrix}.$$

To simplify the notation, we will use Π rather than $\hat{\Pi}$ for the following analysis when there is no confusion. According to the above results, we can substitute 0 for the estimated coefficients that are statistically insignificant at 0.05 and 0.01 significance levels. Then the

corresponding estimated matrices become

$$\hat{\Pi}_{(0.05)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.2819 \\ 0 & 0 \\ 0 & 0.0692 \\ 0 & 0.0682 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.0052 \\ 0 & 0.0051 \\ 0 & 0.0050 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 2.7647 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1.3590 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \hat{\Pi}_{(0.01)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.0692 \\ 0 & 0.0682 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.0052 \\ 0 & 0.0051 \\ 0 & 0.0050 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1.3590 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

We can check identification of λ, γ_f and γ_b in detail by applying the general rank condition (4.4.8). First, we look at identification of λ.

$$\begin{aligned} \text{rank} \begin{bmatrix} J_\varphi(\lambda, \gamma_f, \gamma_b) \\ {e_1}' \end{bmatrix} &= \text{rank} \begin{bmatrix} \Pi_{11} & \Pi_{12} & 1 \\ \Pi_{21} & \Pi_{22} & O \\ 1 & 0 & 0 \end{bmatrix} \\ &= 1 + \text{rank} \begin{bmatrix} \Pi_{12} & 1 \\ \Pi_{22} & O \end{bmatrix} \end{aligned}$$

$$= 2 + \text{rank}(\Pi_{22}). \tag{4.7.23}$$

The necessary and sufficient condition for identification of λ is

$$\text{rank}\left(J_\varphi(\lambda, \gamma_f, \gamma_b)\right) = \text{rank}\begin{bmatrix} J_\varphi(\lambda, \gamma_f, \gamma_b) \\ {e_1}' \end{bmatrix}$$

and equating (4.7.22) and (4.7.23) we have

$$\text{rank}(\Pi_2) = 1 + \text{rank}(\Pi_{22}). \tag{4.7.24}$$

That means either Π_{22} is zero or Π_{21} and Π_{22} are linearly independent if Π_{22} is nonzero [see Harville (2008, Theorem 17.2.4)]. It can be easily verified that the validity of (4.7.24) entails

$$\Pi_{21} \neq O.$$

This is consistent with the expression (4.7.20) since λ is unidentifiable if Π_{21} is zero. Furthermore, when Π_{21} is not equal to zero but close to be zero, we can only have weak identification of λ. Besides, when Π_{21} and Π_{22} are close to be linearly dependent, λ is also weakly identified. We can now use the estimated parameters to verify the above analysis. It is trivial that Π_{22} is nonzero and Π_{21} and Π_{22} are linearly independent. (4.7.24) is satisfied

$$\text{rank}(\Pi_2) = 2 = 1 + \text{rank}(\Pi_{22}).$$

So λ is indeed identifiable.

Second, we want to identify the forward-looking coefficient γ_f and calculate the rank as follows

$$\begin{aligned} \text{rank}\begin{bmatrix} J_\varphi(\lambda, \gamma_f, \gamma_b) \\ {e_2}' \end{bmatrix} &= \text{rank}\begin{bmatrix} \Pi_{11} & \Pi_{12} & 1 \\ \Pi_{21} & \Pi_{22} & O \\ 0 & 1 & 0 \end{bmatrix} \\ &= 1 + \text{rank}\begin{bmatrix} \Pi_{21} & \Pi_{22} \\ 0 & 1 \end{bmatrix} \\ &= 2 + \text{rank}(\Pi_{21}). \end{aligned} \tag{4.7.25}$$

Similarly, we have necessary and sufficient condition for identification from γ_f by equating

(4.7.22) and (4.7.25) as follows

$$\text{rank}(\Pi_2) = 1 + \text{rank}(\Pi_{21}). \tag{4.7.26}$$

That is to say the condition for identification of the forward-looking behaviour of inflation is either Π_{21} is zero or Π_{21} and Π_{22} are linearly independent if Π_{21} is nonzero. We cannot identify γ_f when Π_{22} is zero, which is evidenced by (4.7.20). Again weak identification of γ_f occurs when Π_{22} is close to be zero or there exists a close collinearity between Π_{21} and Π_{22}. Similarly, using the data we can easily calculate that

$$\text{rank}(\Pi_2) = 2 = 1 + \text{rank}(\Pi_{21}).$$

Therefore the forward-looking behaviour of inflation γ_f is also identifiable.

Third, we check identification of the backward-looking component γ_b of inflation. The rank calculation leads to

$$\begin{aligned} \text{rank}\begin{bmatrix} J_\varphi(\lambda, \gamma_f, \gamma_b) \\ {e_3}' \end{bmatrix} &= \text{rank}\begin{bmatrix} \Pi_{11} & \Pi_{12} & 1 \\ \Pi_{21} & \Pi_{22} & O \\ 0 & 0 & 1 \end{bmatrix} \\ &= 1 + \text{rank}(\Pi). \end{aligned} \tag{4.7.27}$$

Hence the necessary and sufficient condition for identification of γ_b is

$$\text{rank}(\Pi_2) = \text{rank}(\Pi), \tag{4.7.28}$$

which means that the information contained by Π_1 is already included in the rows of Π_2. In the extreme case where Π_1 is zero, γ_b is obviously identifiable from (4.7.19). If all the columns of Π_2 are linearly independent, (4.7.28) holds. Given the specific data we use, we find that this is exactly the case and (4.7.28) is satisfied trivially. Thus we conclude that the backword-looking behaviour of inflation γ_b is identifiable.

Fourth, we want to identify both λ and γ_f. Then the rank of the augmented matrix becomes

$$\begin{aligned} \text{rank}\begin{bmatrix} J_\varphi(\lambda, \gamma_f, \gamma_b) \\ {e_1}' \\ {e_2}' \end{bmatrix} &= \text{rank}\begin{bmatrix} \Pi_{11} & \Pi_{12} & 1 \\ \Pi_{21} & \Pi_{22} & O \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\ &= 3. \end{aligned} \tag{4.7.29}$$

From (4.7.22) and (4.7.29), we have the identification condition for both λ and γ_f is

$$\text{rank}(\Pi_2) = 2, \tag{4.7.30}$$

which means that Π_2 has full column rank and it is the same as the conventional rank condition for IV regression models. It is trivial that the condition for identification of the entire semi-structural parameters $\begin{bmatrix} \lambda & \gamma_f & \gamma_b \end{bmatrix}'$ is the same as (4.7.30), which is obvious from (4.7.19) since identification of λ and γ_f implies that of γ_b. Likewise, the chosen data show that (4.7.30) holds. Thus both λ and γ_f are identified.

Considering the small estimated values in Π_{22} and fact that the Federal Reserve System changed its monetary policy from pursuit of multiple goals during the pre-1980 period to a sole goal of price after-1980 due to the impact of the revision in economic theory by Sargent and Wallace (1975), Lucas (1976), Friedman (1977), Barro (1977) and Kydland and Prescott (1977) [see Handa (2009)], which has a far-reaching influence on inflation rates, we check the robustness of the estimated parameters using data over the period of $1980:1-1997:4$. The estimated coefficients with t-statistics and p-values are as follows:

Substituting 0 for the estimated coefficients that are statistically insignificant at 0.05

TABLE 4.2
Estimation for reduced form parameters, 1980:1-1997:4

IV	$\hat{\Pi}_s$	t-stat	p-value	$\hat{\Pi}_{inf}$	t-stat	p-value
π_{t-1}	-0.1697	-0.3612	0.7196	0.2421	1.4979	0.1410
π_{t-2}	0.2485	0.4954	0.6227	0.1741	1.0086	0.3184
π_{t-3}	-0.4678	-0.9439	0.3502	0.0639	0.3748	0.7095
π_{t-4}	0.2460	0.4794	0.6339	0.0107	0.0607	0.9518
π_{t-5}	-0.6260	-1.2547	0.2159	-0.0291	-0.1697	0.8660
$winf_{t-1}$	-0.0477	-0.7498	0.4572	0.0413	1.8855	0.0657
$winf_{t-2}$	-0.0057	-0.0894	0.9292	0.0569	2.6055	0.0123
$winf_{t-3}$	0.0383	0.5823	0.5632	-0.0014	-0.0609	0.9517
$winf_{t-4}$	-0.0044	-0.0812	0.9357	-0.0253	-1.3575	0.1813
$cinf_{t-1}$	-0.0029	-0.4048	0.6875	0.0044	1.7690	0.0835
$cinf_{t-2}$	-0.0113	-1.6528	0.1052	0.0040	1.6830	0.0991
$cinf_{t-3}$	0.0035	0.5131	0.6103	0.0043	1.8233	0.0748
$cinf_{t-4}$	0.0069	1.0080	0.3187	2.3e-04	0.0981	0.9223
rd_{t-1}	-0.1405	-1.1017	0.2763	-0.0256	-0.5837	0.5623
rd_{t-2}	0.4776	2.8982	0.0057	-0.0118	-0.2086	0.8357
rd_{t-3}	-0.3914	-2.1232	0.0391	0.0305	0.4813	0.6326
rd_{t-4}	-0.0059	-0.0428	0.9660	0.0022	0.0473	0.9625
$lshare_{t-1}$	2.0562	1.1230	0.2313	-0.1610	-0.2760	0.7837
$lshare_{t-2}$	-2.3632	-1.1230	0.2673	-0.2385	-0.3293	0.7434
$lshare_{t-3}$	1.3774	0.6703	0.5060	0.0699	0.0988	0.9217
$lshare_{t-4}$	0.8957	0.6014	0.5506	0.8357	1.6305	0.1098
s_{t-1}	1.4086	6.6600	3.0e-08	-0.1140	-1.5659	0.1242
s_{t-2}	-0.3048	-0.9248	0.3599	0.0172	0.1520	0.8798
s_{t-3}	-0.1977	-0.6131	0.5428	0.1079	0.9726	0.3358
s_{t-4}	0.0202	0.0983	0.9221	0.0042	0.0598	0.9526

and 0.01 significance levels, we have

$$\hat{\Pi}_{(0.05)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0.0569 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0.4776 & 0 \\ -0.3914 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1.4086 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \hat{\Pi}_{(0.01)} = \begin{bmatrix} 0 & 0 \\ 1.4086 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

We notice that the rank of estimated Π is equal to 1 if the significance level is chosen to be 0.01, which casts light on some interesting identification issues. First, we check identifiability of λ. Clearly (4.7.24) holds and

$$\text{rank}(\Pi_2) = 1 + \text{rank}(\Pi_{22}) = 1.$$

λ is identifiable.

Second, we look at identification of the forward-looking component of inflation γ_f.

Simple rank calculations lead to

$$1 + \text{rank}(\Pi_{21}) = 2 > \text{rank}(\Pi_2) = 1.$$

Thus (4.7.26) is violated. Since such violation is based on the substitution of 0 for estimated parameters which equal to 0 with a null probability in practice, it is more accurate to argue that γ_f is weakly identified. Meanwhile, it is worth mentioning that weak identification of γ_f will not have any effect on identification of λ since (4.7.24) holds even if Π_{22} is equal to 0.

Third, we check identification of the backward-looking behaviour of inflation γ_b. From (4.7.28), it follows that if some columns of Π_2 are linearly dependent, then the added elements in Π_1 will determine whether or not (4.7.28) is true. Given the specific data we use, since the estimated parameters in Π_{22} are all statistically insignificant at 0.01 level and relative small in magnitude Π_{22} approximates to 0. Therefore, the columns of Π_2 are close to be linearly dependent. Moreover, the estimated Π_{12} equals 0.24 which is statistically insignificant even at 0.14 level. Thus we conclude that (4.7.28) is satisfied and the backward-looking behaviour of inflation γ_b is identifiable whether or not γ_f is weakly identified.

Fourth, we check identification of both λ and γ_f. Likewise, the chosen data show that the rank of Π_2 is 1 if we replace insignificant estimated parameters in Π with 0. Therefore (4.7.30) is not satisfied due to the fact that $\gamma_f$can only be weakly identified.

Fifth, it is interesting to study whether $\gamma_{inf} = \gamma_f + \gamma_b$ is identifiable [8]. Plugging in the derivative of γ_{inf} into the Jacobian matrix and calculating the rank, we have

$$\begin{aligned}
\text{rank}\begin{bmatrix} J_\varphi(\lambda, \gamma_f, \gamma_b) \\ J_{\gamma_{inf}} \end{bmatrix} &= \text{rank}\begin{bmatrix} \Pi_{11} & \Pi_{12} & 1 \\ \Pi_{21} & \Pi_{22} & O \\ 0 & 1 & 1 \end{bmatrix} \\
&= \text{rank}\begin{bmatrix} \Pi_{11} & \Pi_{12} - 1 & 0 \\ \Pi_{21} & \Pi_{22} & O \\ 0 & 1 & 1 \end{bmatrix} \\
&= 1 + \text{rank}\begin{bmatrix} \Pi_{11} & \Pi_{12} - 1 \\ \Pi_{21} & \Pi_{22} \end{bmatrix}. \qquad (4.7.31)
\end{aligned}$$

Note that the second equality sign is due to Gaussian elimination. Equating (4.7.22) and

[8] Galí and Gertler (1999) consider the case where the discount factor $\beta = 1$ so that $\gamma_f + \gamma_b = 1$.

(4.7.31) leads to

$$\text{rank}(\Pi_2) = \text{rank}\begin{bmatrix} \Pi_{11} & \Pi_{12}-1 \\ \Pi_{21} & \Pi_{22} \end{bmatrix}. \tag{4.7.32}$$

Again if the columns of Π_2 are linearly independent, (4.7.32) holds and $\gamma_f + \gamma_b$ is identifiable. However, the estimated Π_{22} is statistically insignificant at 0.01 significance level and $(\Pi_{12}-1)$ is -0.76 which is statistically significant with a p-value very close to 0. Thus when γ_f is weakly identified,

$$\text{rank}(\Pi_2) = 1 < \text{rank}\begin{bmatrix} \Pi_{11} & \Pi_{12}-1 \\ \Pi_{21} & \Pi_{22} \end{bmatrix} = 2$$

and (4.7.32) fails. So we can only weakly identify the linear function $\gamma_f + \gamma_b$. Hence we should cast doubt on the assumption of $\beta = 1$ which implies that $\gamma_f + \gamma_b = 1$ in the related literature on NKPC [see Galí and Gertler (1999), Lindé (2005), Rudd and Whelan (2005) and Nason and Smith (2008), etc.] when forward-looking behaviour of inflation γ_f is weakly identified.

Our analysis leads to the similar conclusion of Dufour et al. (2006) that hybrid NKPC of Galí and Gertler (1999) is somehow supported by the U.S. data over $1970:1-1997:4$. However, this paper also checks the robustness of the NKPC setup using the U.S. data $1980:1-1997:4$. We find out that Galí and Gertler (1999) NKPC involves weak identification issues. Although all the semi-structure parameters are individually identifiable, the forward-looking component of inflation is only weakly identified. Specifically, our paper makes the following two major contributions to NKPC identification research. On one hand, the existing literature focuses on estimation of NKPC and obtains reliable statistical inference when some parameters are weakly identified or unidentified [see Mavroeidis, Plagborg-Møller and Stock (2014) and the references therein]. In contrast, we are the first to give explicit necessary and sufficient conditions for identification of semi-structural parameters of the hybrid NKPC. We provide a more general framework which unifies all types of identification issues, including exact identification, weak identification and identification failure. It is very convenient to study identification of each of these semi-structural parameters. Empirical researchers will now have a clear idea about how likely the weak identification of certain parameter can occur and how to measure the weakness of identification. Take output gap coefficient λ and the forward-looking bahaviour γ_f as an example. Given none of the columns of Π_2 is 0, the more stronger is the linear relationship between the columns of Π_2, the more likely λ and γ_f are weakly identified. Furthermore, the stronger the linear dependence of the columns of Π_2, the weaker is the identification.

Meanwhile, if one column of Π_2 is close to 0 as demonstrated in the above case, we can identify one parameter Λ and weakly identify another γ_f. On the other hand, the generality of our proposed identification conditions also allows for checking identification of an arbitrary function of parameters of interest, which has been scarcely studied in the literature. For instance, we point out that $\gamma_f + \gamma_b$ can be very well weakly identified as proven by our data. Therefore the popular analysis based on the assumption $\beta = 1$ should be treated with caution. Note that identification failure (or weak identification) of γ_f does not necessarily lead to identification failure (or weak identification)of a function of it. To summarize, it is likely that we cannot identify all the semi-structural parameters unless all the relevant rank conditions are satisfied. In fact it might well be the case that all parameters are identifiable but some are only weakly identified, such as the coefficient of forward-looking component of inflation γ_f.

4.8. Conclusion

We have been focusing on several important issues related to identification in this paper. First of all, we examine the conditions for parametric function identification with and without linear or nonlinear restrictions. Specifically, we are interested in identifying a parametric function $\beta(\theta)$ in terms of another identifiable parametric function $\gamma(\theta)$. Since there is no constraint on the function form of $\beta(\theta)$, it can be linear or nonlinear in θ. For instance, it can take the form of θ itself, an arbitrary subvector or a scalar of θ, any linear combination of θ, the score function and the Kullback-Leibler divergence. Therefore, the necessary and sufficient condition for parametric function identification includes the classical statement of identification of θ as a special case. Besides, for the convenience of empirical researchers who usually study statistical models through the explanatory variables, we also provide some equivalent conditions that are intuitive and easy to check based on the properties of linear subspaces.

Second, due to the importance and yet difficulty of globally identifying the parameter θ [see Rothenberg (1971)], we define a stronger version of local identification, i.e., local identification of θ around a particular parameter value θ_0 in the sense that it entails the standard definition of local identification at θ_0. The purpose of introducing this new concept is that we can still identify the parameters of interest within a neighborhood of some given point although achieving global identification over the entire parameter space is challenging in most cases without further useful information. Thus this arbitrarily small "neighborhood" can be treated as a reduced parameter space where any two distinguished parameter values imply different probability distributions characterized by parameters. We

have established a necessary and sufficient condition for local identification around θ_0 under weaker assumptions in comparison to some well-known results on local identification at θ_0, such as these by Rothenberg (1971), Bekker and Wansbeek (2001) and Chen et al. (2011).

Third, we establish sufficient conditions (and sometimes both necessary and sufficient conditions) for identification through the Kullback-Leibler divergence and the first order partial derivative of the Kullback-Leibler divergence. It can be easily verified that the assumptions about the standard conditions for local identification are more restrictive than needed and can be replaced with weaker ones. Furthermore, we can reach the same conclusions as those by Rothenberg (1971) and Bowden (1973) using the rank information about the Hessian matrix of the Kullback-Leibler divergence evaluated at the true θ_0 which is equivalent to the Fisher information matrix at θ_0. Specifically, we give the necessary and sufficient rank condition for local identification of θ at θ_0 through the Kullback-Leibler divergence. Besides, it is straightforward to establish the condition for parametric function identification within the likelihood model setup.

Fourth, we show that our identification conditions can be applicable to some important statistical and macroeconomic models, namely the SEMs and the DSGE models. In comparison to the classical literature, we relax some of the classical regularity assumptions about the simultaneous models, such as normality of the structural shocks, nonsingular coefficient matrix of the endogenous variables and linearity of the restrictions, etc; see Fisher (1966), Rothenberg (1971), Richmond (1974) and Bekker and Wansbeek (2001). It follows that the general rank condition, which is just one of a group of equivalent identification conditions, holds whether or not the design matrix has full column rank. Furthermore, the validity of our rank condition is not affected by whether we can have a unique expression of the reduced form of the simultaneous equation system since it is derived from the constructed system of moment conditions. Consequently, all the standard conditions in the simultaneous models are special cases of the general rank condition. On the other hand, it is easy to apply our results for function identification to the DSGE models with high dimension and high nonlinearity properties. Similarly, the statements of identification in the literature for the DSGE models can be generalized as special cases.

Fifth, we look into some real macroeconomic models and extend our identification results to analysis of weak identification. We consider a two-equation SEMs of Romer (1993) followed by a three-equation dynamic New Keynesian model in the format of Canova (2009). Since the hybrid NKPC proposed by Galí and Gertler (1999) gains popularity in the study of inflation behaviour, we examine this particular type of model and derive some interesting results on identification of the semi-structural parameters, which are new to the

literature. More importantly, our general framework provides the flexibility of checking both identification failure and weak identification, which has been regarded as a big challenge for getting reliable statistical inference. As shown in the paper, the measurement of weak identification can be reduced to verifying the linear dependent relationship among certain rows or columns of known matrices, which provides a convenient tool for empirical study on parameter identification and estimation.

There are some possible extensions to the main conclusions of this paper. Recently, Cho and White (2012) study the asymptotic distribution of the extreme estimator by generalizing some of the statistical inference results from the standard differentiable econometric models to the models where the functions are directionally differentiable. Thus it is promising to investigate identification conditions when the partial derivatives are not well defined along certain coordinates whereas the directional derivatives exist. Another interesting topic related to this paper is how to achieve global (parametric function) identification for nonlinear regression models.

Chapter 5

Conclusion

Since distributional theory and reliable statistical inference depend critically on which parameters or transformations of parameters are identifiable, this thesis characterizes sufficient conditions as well as necessary and sufficient conditions for different types of identifications – global identification, local identification and parametric function identification – in both linear and nonlinear models. It covers the following important topics: relationship between conditions for estimability and identifiability and global identification of linear functions of parameters, identification of parametric functions in IV regressions and SEMs, local identification of nonlinear parametric function identification and global identification in nonlinear models.

In view of the fact that the necessary and sufficient conditions for estimability coincide with those for identifiability when we have a separable mean-variance structure in general linear models [see Reiersøl (1963), Rao and Mitra (1971) and Kounias and Chalikias (2008)], it is a natural way to study identification from familiar results on estimability in a linear setup in the literature. The first essay looks at identification conditions for partially linear models which includes most of the frequently used models in statistics and econometrics, such as general linear models, generalized linear models, linear mixed models, linear median regressions and quantile regressions, single index models, etc. We propose a general necessary and sufficient condition for global identification of a general transformation of parameters under separability assumption where parameters of interest can be completely separated from nuisance parameters. This separability assumption guarantees both necessity and sufficiency of the proposed condition which will otherwise lose necessity. Furthermore, this condition is applied to partially linear models and the model parameters can be fully characterized despite the fact that partially linear models usually relax assumptions of additivity of the error term and separability of mean-variance structure. Besides, since restrictions are closely related to statistical inference, such as test and

hypothesis, we take into account the effect of adding linear restrictions on identification and the corresponding conditions are provided. Additionally, a group of intuitive equivalent conditions are proposed based on properties of column subspaces and generalized inverse methods.

As the current literature on weak instruments concentrate on linear IV regression and SEMs, the second essay studies global identification of parametric functions in such models. The framework of this paper is quite general in that a series of standard assumptions regarding model specification are relaxed [see Bekker and Wansbeek (2001)]. First, we do not impose full column rank assumption on the design matrix of the reduced form equations because it is irrelevant to our interest in identification of structural parameters. Second, we can still characterize identification of a linear transformation of deep parameters although none of the original model parameters can be identified. Third, the design matrix does not have to be exogeneous, which provides flexibility in dealing with some crucial issues on instruments. Indeed, identification of parameters depend on realized values of instruments. Fourth, we allows for the presence of nuisance parameters in the model setup so that the proposed general condition for identification in the first essay is easily extended to linear IV regressions.

Given the growing popularity of specifying nonlinear models to tackle economic problems, such as nonlinear models estimated by GMM and the DSGE models, the third essay studies identification of nonlinear parametric functions in nonlinear models. Parametric function identification is a new concept introduced to literature which generalizes partial identification of a scalar parameter or a subvector of parameters by Phillips (1989) and Bekker and Wansbeek (2001). It also distinguishes itself from partial identification by Manski (1995) who uses this terminology to refer to set-valued identification. We propose both necessary and sufficient conditions for nonlinear parametric function identification with and without nonlinear restrictions. Meanwhile, considering the difficulty of achieving global identification in nonlinear models, this paper focuses on local identification around a point which lies between global identification and classical local identification at a point. The reason of introducing local identification around a point is that this stronger version of local identification will secure a consistent estimator whereas local identification at a point is very restrictive for statistical inference and we might lose consistency under certain circumstances. To make our general condition easier to apply, we provide a set of equivalent conditions based on the conclusions in the first essay. Moreover, we demonstrate the generality of the proposed condition by applying it to nonlinear SEMs, likelihood models and DSGE models.

There are some promising extensions to the topics covered in this thesis. First, it is

interesting to obtain identification conditions through properties of a monotone operator. Employing monotonicity to achieve identification conditions has some gains. For instance, the parametric function does not have to be smooth and we do not need to calculate rank of a Jacobian matrix. Besides, it is possible to achieve global identification even if the function is not continuous. Second, identification is feasible through higher order moments when information given by the first two moments is inadequate to obtain identification. For example, in the likelihood models, suppose the Hessian matrix of the Kullback-Leibler divergence is singular, a sufficient condition for local identification can be derived based on the information on the third order derivatives of the Kullback-Leibler divergence. Third, it is interesting to study identification in a multi-dimensional space where certain partial derivatives are not well defined along specific coordinates whereas the directional derivatives exist. Thus directional derivatives rather than partial derivatives can help identify parameters of interest.

Appendix

PROOF OF PROPOSITION **2.3.2** Suppose (2.3.2) holds. Then, the identifiability of $\delta(\theta)$ on Θ_0 simply follows by applying Definition **2.2.1** with $\psi(\theta) = \delta(\theta)$. Conversely, suppose $\delta(\theta)$ is identifiable on Θ_0, but condition (2.3.2) does not hold. In this case, we can find θ_1, $\theta_2 \in \Theta_0$ such that $\delta(\theta_1) \neq \delta(\theta_2)$ and $\gamma(\theta_1) = \gamma(\theta_2)$. Since $\delta(\theta)$ is identifiable on Θ_0, we have $\mathsf{P}_{\theta_1} \neq \mathsf{P}_{\theta_2}$. On the other hand, since $\gamma(\theta_1) = \gamma(\theta_2)$,

$$\mathsf{P}_{\theta_1} = \bar{\mathsf{P}}_{\gamma(\theta_1)} = \bar{\mathsf{P}}_{\gamma(\theta_2)} = \mathsf{P}_{\theta_2}.$$

This leads to a contradiction, so condition (2.3.2) is necessary for the identifiability of $\delta(\theta)$ on Θ_0. □

PROOF OF THEOREM **2.3.4** Suppose (2.3.4) holds. Then, the identifiability of $\delta(\theta)$ on Θ_0 simply follows by applying Definition **2.2.1** with $\psi(\theta) = \gamma_1(\theta)$. Conversely, suppose $\delta(\theta)$ is identifiable on Θ_0, but condition (2.3.4) does not hold. Then we can find θ_1, $\theta_2 \in \Theta_0$ such that $\delta(\theta_1) \neq \delta(\theta_2)$ and $\gamma_1(\theta_1) = \gamma_1(\theta_2)$. Using Assumption **2.3.3** with $\theta = \theta_1$ and $\bar{\gamma}_2 = \gamma_2(\theta_2)$, we can also find θ_1^* such that $\delta(\theta_1^*) = \delta(\theta_1)$, $\gamma_1(\theta_1^*) = \gamma_1(\theta_1)$ and $\gamma_2(\theta_1^*) = \gamma_2(\theta_2)$, so that

$$\gamma(\theta_1^*) = (\gamma_1(\theta_1^*), \gamma_2(\theta_1^*)) = (\gamma_1(\theta_1), \gamma_2(\theta_2)) = (\gamma_1(\theta_2), \gamma_2(\theta_2)) = \gamma(\theta_2). \qquad \text{(A.1)}$$

Since $\delta(\theta_1^*) = \delta(\theta_1) \neq \delta(\theta_2)$ and $\delta(\theta)$ is identifiable on Θ_0, we must have $\mathsf{P}_{\theta_1^*} \neq \mathsf{P}_{\theta_2}$. On the other hand, since $\gamma(\theta)$ is sufficient for $\mathscr{P}$ on Θ_0, (A.1) entails $\mathsf{P}_{\theta_1^*} = \mathsf{P}_{\theta_2}$. This leads to a contradiction, so condition (2.3.4) is necessary for the identifiability of $\delta(\theta)$ on Θ_0. □

PROOF OF PROPOSITION **2.3.7** The fact that Assumption **2.3.3** entails Assumptions **2.3.5** and **2.3.6** follows by taking $\phi(\theta) = \big(\delta(\theta), \gamma_1(\theta)\big)$. Conversely, suppose Assumptions **2.3.5** and **2.3.6** hold. By Assumption **2.3.5**, we have

$$\delta(\theta) = g[\phi(\theta)] \text{ and } \gamma_1(\theta) = h_1[\phi(\theta)], \quad \forall \theta \in \Theta_0,$$

for some functions g and h_1. Further, by Assumption **2.3.6**, for any $\theta \in \Theta$ and $\bar{\gamma}_2 \in \gamma_2(\Theta_0)$, we can find $\theta^* \in \Theta_0$ such that $\phi(\theta^*) = \phi(\theta)$ and $\gamma_2(\theta^*) = \bar{\gamma}_2$, hence

$$\delta(\theta^*) = g[\phi(\theta^*)] = g[\phi(\theta)] = \delta(\theta), \ \gamma_1(\theta^*) = h_1[\phi(\theta^*)] = h_1[\phi(\theta)] = \gamma_1(\theta)$$

and (2.3.3) holds. Assumption **2.3.3** is thus satisfied. □

PROOF OF PROPOSITION **2.3.8** The fact that (2.3.10) entails Assumption **2.3.6** follows directly from the definition of the Cartesian product. Suppose now Assumption **2.3.6** holds. It is easy to see that $\nu(\Theta_0) \subseteq \phi(\Theta_0) \times \gamma_2(\Theta_0)$. Let $\bar{\nu} = (\bar{\phi}, \bar{\gamma}_2) \in \phi(\Theta_0) \times \gamma_2(\Theta_0)$. We can then find $\theta_1, \theta_2 \in \Theta_0$ such that $\phi(\theta_1) = \bar{\phi}$ and $\gamma_2(\theta_2) = \bar{\gamma}_2$. By Assumption **2.3.6**, we can find $\theta_1^* \in \Theta_0$ such that $\phi(\theta_1^*) = \bar{\phi}$ and $\gamma_2(\theta_1^*) = \bar{\gamma}_2$, hence $\nu(\theta_1^*) = \big(\phi(\theta_1^*), \gamma_2(\theta_1^*)\big) \in \nu(\Theta_0)$. Thus $\phi(\Theta_0) \times \gamma_2(\Theta_0) \subseteq \nu(\Theta_0)$, which along with $\nu(\Theta_0) \subseteq \phi(\Theta_0) \times \gamma_2(\Theta_0)$, entails $\nu(\Theta_0) = \phi(\Theta_0) \times \gamma_2(\Theta_0)$. □

PROOF OF LEMMA **2.4.1** Consider the block triangular matrices:

$$T_L = \begin{bmatrix} I_p & O \\ V_1 & I_q \end{bmatrix}, \quad T_U = \begin{bmatrix} I_p & V_2 \\ O & I_q \end{bmatrix}.$$

Clearly, T_L and T_U are invertible, hence

$$\operatorname{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \operatorname{rank}\left\{ \begin{bmatrix} I_p & O \\ V_1 & I_q \end{bmatrix} \begin{bmatrix} Z \\ Q \end{bmatrix} \right\} = \operatorname{rank}\begin{bmatrix} Z \\ Q + V_1 Z \end{bmatrix}, \tag{A.2}$$

$$\operatorname{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \operatorname{rank}\left\{ \begin{bmatrix} I_p & V_2 \\ O & I_q \end{bmatrix} \begin{bmatrix} Z \\ Q \end{bmatrix} \right\} = \operatorname{rank}\begin{bmatrix} Z + V_2 Q \\ Q \end{bmatrix}. \tag{A.3}$$

This establishes (2.4.1). We can now look at the rank additivity properties.

(a) If (2.4.2) holds, each row of Z is linearly independent of the row space of $Q + V_1 Z$ are essentially disjoint. Thus

$$\operatorname{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \operatorname{rank}\begin{bmatrix} Z \\ Q + V_1 Z \end{bmatrix} = \operatorname{rank}(Z) + \operatorname{rank}(Q + V_1 Z)$$

and (2.4.3) holds.

(b) If (2.4.4) holds, each row of $Z + V_2 Q$ is linearly independent of the row space of Q. Thus

$$\operatorname{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \operatorname{rank}\begin{bmatrix} Z + V_2 Q \\ Q \end{bmatrix} = \operatorname{rank}(Z + V_2 Q) + \operatorname{rank}(Q)$$

and (2.4.5) holds.

(c) Suppose $V_1 = -QZ^-$ but (2.4.2) does not hold, *i.e.* $V_1 = -QZ^-$ and $\operatorname{Im}\big(Z'\big) \cap \operatorname{Im}\big(Q' + Z'V_1{}'\big) \neq \{0\}$. We can then find non-zero vectors x_1 and x_2 such that

$$Z'x_1 = \big(Q' + Z'V_1{}'\big)x_2 \neq 0,$$

hence

$$x_1{}'Z = x_2{}'(Q - V_1 Z) = x_2{}'Q(I_k - Z^- Z) \neq 0 \tag{A.4}$$

and, on post-multiplying by $(I_k - Z^- Z)$,

$$x_1{}'Z(I_k - Z^- Z) = x_2{}'Q(I_k - Z^- Z)(I_k - Z^- Z) = x_2{}'Q(I_k - Z^- Z) \neq 0\,.$$

However,

$$x_1{}'Z(I_k - Z^- Z) = x_1{}'(Z - ZZ^- Z) = 0 \tag{A.5}$$

and we have a contradiction. Consequently, (2.4.2) must hold. (2.4.3) follows from part (a) of this Lemma.

(d) Suppose $V_2 = -ZQ^-$ but (2.4.4) does not hold: *i.e.* $V_2 = -ZQ^-$ and $\mathrm{Im}\big(Q'\big) \cap \mathrm{Im}\big(Z' + Q'V_2{}'\big) \neq \{0\}$. We can then find non-zero vectors x_1 and x_2 such that

$$Q'x_1 = \big(Z' + Q'V_2{}'\big)x_2 \neq 0\,. \tag{A.6}$$

hence

$$x_1'Q = x_2'\big(Z + V_2 Q\big) = x_2'Z\big(I_k - Q^- Q\big) \neq 0 \tag{A.7}$$

and, on post-multiplying by $(I_k - Q^- Q)$,

$$x_1'Q(I_k - Q^- Q) = x_2{}'Z(I_k - Q^- Q)(I_k - Q^- Q) = x_2{}'Z(I_k - Q^- Q) \neq 0\,. \tag{A.8}$$

However,

$$x_1'Q(I_k - Q^- Q) = x_1'(Q - QQ^- Q) = 0 \tag{A.9}$$

and we have a contradiction. Consequently, (2.4.4) must hold. (2.4.5) follows from part (b) of this Lemma. □

PROOF OF PROPOSITION **2.4.2** Let $d \equiv \beta_1 - \beta_2$. The equivalence between (2.4.6) and (2.4.7) simply follows from the definition of a function. Further,

$$\begin{aligned}
&\big[(Z\beta_1 = Z\beta_2) \Rightarrow (Q\beta_1 = Q\beta_2), \forall \beta_1, \beta_2\big] \Leftrightarrow \big[(Z\delta = 0) \Rightarrow (Q\delta = 0), \forall d \in \mathbb{R}^k\big] \\
&\quad \Leftrightarrow \Big[\big(d \in \ker(Z)\big) \Rightarrow \big(d \in \ker(Q)\big), \forall d\Big] \\
&\quad \Leftrightarrow \big[\ker(Z) \subseteq \ker(Q)\big] \Leftrightarrow \Big[\big[\mathrm{Im}(Z')\big]^{\perp} \subseteq \big[\mathrm{Im}(Q')\big]^{\perp}\Big] \\
&\quad \Leftrightarrow \big[\mathrm{Im}(Q') \subseteq \mathrm{Im}(Z')\big] \\
&\quad \Leftrightarrow \big[Q' = Z'B', \text{for some matrix } B\big] \Leftrightarrow \mathrm{rank}\begin{bmatrix} Z' & Q' \end{bmatrix} = \mathrm{rank}\big(Z'\big)
\end{aligned}$$

$$\Leftrightarrow \text{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \text{rank}(Z)$$

$$\Leftrightarrow \text{rank}\begin{bmatrix} Z \\ Q+V_1Z \end{bmatrix} = \text{rank}(Z)$$

$$\Leftrightarrow \text{rank}\begin{bmatrix} Z+V_2Q \\ Q \end{bmatrix} = \text{rank}(Z)$$

where we use the well known relation between the kernel of a matrix C and and the image of of its transpose C':

$$\ker(C) = [\text{Im}(C')]^{\perp}, \text{ for any matrix } C;$$

see Gouriéroux and Monfort (1995). The last two equivalences follow from (2.4.1). This establishes the equivalences between the statements (2.4.7) to (2.4.13).

We now show that (2.4.14) and (2.4.9) are equivalent. Suppose (2.4.9) holds: $\text{Im}(Q') \subseteq \text{Im}(Z')$. For any $q_1 \times q$ matrix S, we thus have $\text{Im}\big(Q'S'\big) \subseteq \text{Im}\big(Q'\big) \subseteq \text{Im}\big(Z'\big)$, hence $\text{rank}[Z' \quad Q'S'] = \text{rank}(Z')$. Since the latter identity holds for any $q_1 \times q$ matrix (without rank restriction), we have

$$\text{rank}\begin{bmatrix} Z \\ SQ \end{bmatrix} = \text{rank}(Z)$$

for any S matrix such that $\text{rank}(SQ) = \text{rank}(Q)$. *i.e.* (2.4.14) holds. Conversely, suppose (2.4.14) holds. When $\text{rank}(SQ) = \text{rank}(Q)$, we must have $\text{Im}(Q'S') = \text{Im}(Q')$; see Harville (2008, Theorem 4.4.7). Further,

$$\Big[\text{rank}\begin{bmatrix} Z \\ SQ \end{bmatrix} = \text{rank}(Z)\Big] \Rightarrow \Big[\text{rank}[Z' \quad Q'S'] = \text{rank}(Z')\Big] \Rightarrow \Big[\text{Im}(Q'S') \subseteq \text{Im}(Z')\Big]$$
$$\Rightarrow \Big[\text{Im}(Q') \subseteq \text{Im}(Z')\Big]$$

and (2.4.9) holds.

To get (2.4.15), suppose (2.4.11) holds. Using Lemma **2.4.1** with $V_1 = -QZ^-$, where Z^- is a g-inverse of Z, we see that

$$\text{rank}(Z) = \text{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \text{rank}(Z) + \text{rank}(Q - QZ^-Z)$$

hence $\text{rank}\Big\{Q - QZ^-Z\Big\} = 0$ and $Q = QZ^-Z$. Thus (4.4.13) holds. Conversely, suppose

$Q = QZ^{-}Z$, for some g-inverse Z^{-} of Z. Then $Q' = Z'(QZ^{-})'$

$$Q' = Z'(QZ^{-})'$$

which entails $\text{rank}\begin{bmatrix} Z' & Q' \end{bmatrix} = \text{rank}(Z')$ and

$$\text{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \text{rank}(Z)\,.$$

This establishes the equivalence between (2.4.15) and (2.4.11). each one of the conditions (2.4.7) to (2.4.14).
Finally, to get (2.4.16), suppose (2.4.11) holds. Using Lemma **2.4.1** with $V_2 = -ZQ^{-}$, where Q^{-} is a g-inverse of Q, we have:

$$\text{rank}(Z) = \text{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \text{rank}\Big\{Z - ZQ^{-}Q\Big\} + \text{rank}(Q) = \text{rank}\Big\{Z(I_k - Q^{-}Q)\Big\} + \text{rank}(Q) \tag{A.10}$$

and (2.4.16) holds. Conversely, suppose (2.4.16) holds, *i.e.*,

$$\text{rank}(Z) = \text{rank}\Big\{Z(I_k - Q^{-}Q)\Big\} + \text{rank}(Q) \tag{A.11}$$

for some g-inverse Q^{-}. It is easy to see that:

$$\text{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} \geq \text{rank}(Z)\,. \tag{A.12}$$

Further, using (2.4.1) with $V_2 = -ZQ^{-}$, we see that

$$\begin{aligned} \text{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} &= \text{rank}\begin{bmatrix} Z - ZQ^{-}Q \\ Q \end{bmatrix} = \text{rank}\begin{bmatrix} Z(I_k - Q^{-}Q) \\ Q \end{bmatrix} \\ &\leq \text{rank}\Big\{Z(I_k - Q^{-}Q)\Big\} + \text{rank}(Q) = \text{rank}(Z)\,. \end{aligned} \tag{A.13}$$

On combining (A.12) with (A.13), we get

$$\text{rank}\begin{bmatrix} Z \\ Q \end{bmatrix} = \text{rank}(Z) \tag{A.14}$$

and (2.4.11) holds. The equivalence between (2.4.16) and (2.4.11) follows.
All the required equivalences have been established. □

PROOF OF LEMMA **2.4.3** The equivalence (2.4.24) follows from standard results on solutions of linear equations: namely, if $c_0 \in \mathrm{Im}(R)$, the set of all solutions of equation (2.4.22) is given by:

$$\beta = R^- c_0 + (I_k - R^- R)d,\ d \in \mathbb{R}^k, \tag{A.15}$$

where R is any generalized inverse of R; see Rao and Mitra (1971) or Harville (1997, Section 11.2). To get (2.4.25), suppose first that: $\beta = b_0 + Ce$ for some $e \in \mathbb{R}^m$. Multiplying both sides of this equation by $(I_k - CC^-)$, we get:

$$(I_k - CC^-)\beta = (I_k - CC^-)b_0 + (C - CC^-C)e = (I_k - CC^-)b_0\,. \tag{A.16}$$

Conversely, suppose the latter identity holds. This is a linear equation system, and its general solution can be written:

$$\beta = D^- D b_0 + (I_k - D^- D)d = b_0 + (I_k - D^- D)(d - b_0), \quad d \in \mathbb{R}^k, \tag{A.17}$$

where $D = I_k - CC^-$. We can take $D^- = I_k + CC^-$, for D^- verifies the definition of a g-inverse of D:

$$\begin{aligned} DD^-D &= (I_k - CC^-)(I_k + CC^-)(I_k - CC^-) = (I_k - CC^-)(I_k - CC^-CC^-) \\ &= (I_k - CC^-)(I_k - CC^-) = (I_k - CC^-) = D\,. \end{aligned} \tag{A.18}$$

Further, $D^-D = D$ and $I_k - D = CC^-$, hence

$$\beta = b_0 + (I_k - D)(d - b_0) = b_0 + CC^-(d - b_0) = b_0 + Ce \tag{A.19}$$

where $e = C^-(d - b_0) \in \mathbb{R}^m$. This completes the proof of the equivalence (2.4.25). □

PROOF OF PROPOSITION **2.4.4** Suppose (2.4.27) holds, and let $c_0 \in \mathrm{Im}(R)$. Then we can find a function $g_{c_0} : \mathrm{Im}(Z) \mapsto \mathrm{Im}(Q)$ such that

$$Q\beta = g_{c_0}(Z\beta),\ \forall \beta \in \mathscr{L}_I(R, c_0)\,. \tag{A.20}$$

Set

$$g(Z\beta, c_0) = g_{c_0}(Z\beta) \text{ for } \beta \in \mathscr{L}_I(R, c_0)\,, \tag{A.21}$$

$$\bar{g}(\bar{Z}\beta) = g(Z\beta, R\beta) \text{ for } \beta \in \mathbb{R}^k\,. \tag{A.22}$$

Clearly,

$$\bar{g}(\bar{Z}\beta) = g(Z\beta, R\beta) = g(Z\beta, c_0) = Q\beta \text{ for } \beta \in \mathscr{L}_I(R, c_0) \tag{A.23}$$

and $\bar{g}$ satisfies (2.4.28). Conversely, if we have (2.4.28), it is clear (2.4.27) holds upon considering restricted-domain functions of the form: $g_{c_0}(Z\beta) = \bar{g}(\bar{Z}\beta) = g(Z\beta, c_0)$ for $\beta \in \mathscr{L}_I(R, c_0)$. Thus (2.4.27) and (2.4.28) are equivalent. The equivalences between (2.4.28) and the other statements (2.4.29) - (2.4.38) of the Proposition follow as in Proposition **2.4.2** after replacing Z by $\bar{Z}$. □

PROOF OF PROPOSITION **2.4.5** Suppose (2.4.44) holds, and let $\beta_1, \beta_2 \in \mathscr{L}_E(C, b_0)$. We thus have

$$\left(Z\beta_1 = Z\beta_2\right) \Rightarrow \left(Q\beta_1 = Q\beta_2\right) \tag{A.24}$$

and we can find $d_1, d_2 \in \mathbb{R}^m$ such that $\beta_1 = b_0 + Cd_1$ and $\beta_2 = b_0 + Cd_2$. Further, it is clear that:

$$\left(Z\beta_1 = Z\beta_2\right) \Leftrightarrow \left(ZCd_1 = ZCd_2\right), \tag{A.25}$$

$$\left(Q\beta_1 = Q\beta_2\right) \Leftrightarrow \left(QCd_1 = QCd_2\right), \tag{A.26}$$

hence

$$\left(ZCd_1 = ZCd_2\right) \Rightarrow \left(QCd_1 = QCd_2\right). \tag{A.27}$$

Since this holds for any $\beta_1, \beta_2 \in \mathscr{L}_E(C, b_0)$, (2.4.45) follows. Conversely, suppose (2.4.45) holds. Then (2.4.44) follows from (A.25) - (A.26) and the definition of a function. (2.4.44) and (2.4.45) are thus equivalent. The equivalences between the statements (2.4.45) - (2.4.54) can be proved as in Proposition **2.4.2** on replacing Z by ZC and Q by QC. □

PROOF OF PROPOSITION **2.4.6** The equivalence between (2.4.56) and (2.4.57) follows from the equivalence between (2.4.44) and (2.4.47) in Proposition **2.4.5** with $Q = l'$: in this case, $C'Q' = C'l = C'_{j\cdot} l_j + C'_{(j\cdot)} l_{(j)} \neq 0$ a non-zero vector, and

$$[\text{Im}(C'Q') \subseteq \text{Im}(C'Z')] \Leftrightarrow C'l \in \text{Im}(C'Z'). \tag{A.28}$$

To get (2.4.58), we use condition (2.4.49) with $Q = l'$:

$$\text{rank}(ZC) = \text{rank}\begin{bmatrix} ZC \\ l'C \end{bmatrix} = \text{rank}\,[\bar{Z}C] \quad \text{where } \bar{Z} = \begin{bmatrix} Z \\ l' \end{bmatrix}. \tag{A.29}$$

Further, on noting that $ZC_{\cdot j}$ is the j-th column of ZC while $ZC_{(\cdot j)}$ gathers the other columns

of ZC, we see that

$$\text{rank}(ZC) = \text{rank}[ZC_{\cdot j} \;\; ZC_{(\cdot j)}] = \text{rank}[ZC_{\cdot j} \;\; ZN_jC_{(\cdot j)}] \tag{A.30}$$

where the second identity is based on on observing that the matrix

$$ZN_jC_{(\cdot j)} = Z\left[I_k - \frac{1}{C'_{\cdot j}l}C_{\cdot j}l'\right]C_{(\cdot j)} = ZC_{(\cdot j)} - (ZC_{\cdot j})\frac{1}{C'_{\cdot j}l}(l'C_{(\cdot j)}) \tag{A.31}$$

is obtained by subtracting from each column of $ZC_{(\cdot j)}$ a multiple of the first column $ZC_{\cdot j}$, so the ranks of $[ZC_{\cdot j} \;\; ZC_{(\cdot j)}]$ and $[ZC_{\cdot j} \;\; ZN_jC_{(\cdot j)}]$ are identical. Similarly,

$$\begin{aligned}
\text{rank}(\bar{Z}C) &= \text{rank}[\bar{Z}C_{\cdot j} \;\; \bar{Z}C_{(\cdot j)}] = \text{rank}[\bar{Z}C_{\cdot j} \;\; \bar{Z}N_jC_{(\cdot j)}] \\
&= \text{rank}\begin{bmatrix} ZC_{\cdot j} & ZN_jC_{(\cdot j)} \\ l'C_{\cdot j} & l'N_jC_{(\cdot j)} \end{bmatrix} = \text{rank}\begin{bmatrix} ZC_{\cdot j} & ZN_jC_{(\cdot j)} \\ l'C_{\cdot j} & 0' \end{bmatrix} \\
&= 1 + \text{rank}\left[ZN_jC_{(\cdot j)}\right]
\end{aligned} \tag{A.32}$$

where we use the facts that $l'N_jC_{(\cdot j)} = 0'$ and $C'_{\cdot j}l \neq 0$. Combining (A.30) and (A.32), we see that (2.4.58) is equivalent to (A.29). Finally, the identity

$$\text{rank}[ZC_{\cdot j} \;\; ZN_jC_{(\cdot j)}] = 1 + \text{rank}\left[ZN_jC_{(\cdot j)}\right] \tag{A.33}$$

means that $ZC_{\cdot j} \notin \text{Im}\left[ZN_jC_{(\cdot j)}\right]$. □

PROOF OF THEOREM **2.5.3** Set $\delta(\theta) = Q\beta$, $\gamma_1(\theta) = X\beta$, $\gamma_2(\theta) = \mathsf{P}_u(\theta, X)$, and $\Gamma_2 = \{\mathsf{P}_u(\theta, X) : \theta \in \Theta_0\}$ the set of all error distributions determined by $\theta \in \Theta_0$, where X is taken as a known matrix of constants. In this model, the distribution of y is completely determined by $X\beta$ (a fixed vector) and the distribution of u. Clearly, $\gamma(\theta) = \big(\gamma_1(\theta),\ \gamma_2(\theta)\big)$ is sufficient on Θ_0. Further, $\gamma_1(\theta) = \mathsf{E}_\theta(y|X)$ is identifiable on Θ_0, and Assumption **2.3.3** is satisfied [by Assumption **2.5.2**]. The result then follows on applying Theorem **2.3.4**. □

PROOF OF THEOREM **2.5.4** By Proposition **2.4.2** with $Z = X$, each one of the conditions (2.5.17) - (2.5.25) is equivalent to

$$[(X\beta_1 = X\beta_2) \Rightarrow (Q\beta_1 = Q\beta_2)]\ (\forall \beta_1, \beta_2 \in \mathbb{R}^k) \tag{A.34}$$

hence, on noting that $\beta_1 = \beta(\theta_1)$ and $\beta_2 = \beta(\theta_2)$ [see (2.5.6)],

$$\left[\left(X\beta(\theta_1) = X\beta(\theta_2)\right) \Rightarrow \left(Q\beta(\theta_1) = Q\beta(\theta_2)\right)\right] (\forall \theta_1, \theta_2 \in \Theta_0). \tag{A.35}$$

This means that $Q\beta$ is $(X\beta)$-identifiable on Θ_0.
Conversely, suppose $Q\beta$ is $(X\beta)$-identifiable on Θ_0, along with Assumption **2.5.2** and $\beta(\Theta_0) = \mathbb{R}^k$. By Theorem **2.5.3**, this entails (A.35). Further, since $\beta(\Theta_0) = \mathbb{R}^k$, (A.35) implies (A.34), hence [by Proposition **2.4.2**] each one of the conditions (2.5.17) - (2.5.25). This completes the proof. □

PROOF OF THEOREM **2.5.5** By Proposition **2.4.4** with $\bar{Z} = \bar{X}$, each one of the conditions (2.5.28) - (2.5.36) is equivalent to

$$[(\bar{X}\beta_1 = \bar{X}\beta_2) \Rightarrow (Q\beta_1 = Q\beta_2)] \left(\forall \beta_1, \beta_2 \in \mathbb{R}^k\right) \tag{A.36}$$

hence, for any $c_0 \in \mathrm{Im}(R)$,

$$\left[\left(X\beta(\theta_1) = X\beta(\theta_2)\right) \Rightarrow \left(Q\beta(\theta_1) = Q\beta(\theta_2)\right)\right] \left(\forall \theta_1, \theta_2 \in \bar{\Theta}_{0I}(R, c_0)\right). \tag{A.37}$$

This means that $Q\beta$ is $(X\beta)$-identifiable on $\bar{\Theta}_{0I}(R, c_0)$, for any $c_0 \in \mathrm{Im}(R)$.
Conversely, suppose $Q\beta$ is $(X\beta)$-identifiable on $\bar{\Theta}_{0I}(R, c_0)$, for any $c_0 \in \mathrm{Im}(R)$, along with Assumption **2.5.2** and $\bar{\beta}_I(R, c_0) = \{\beta \in \mathbb{R}^k : R\beta = c_0\}$. Under Assumption **2.5.2**, Theorem **2.5.3** entails that (A.37) holds for any $c_0 \in \mathrm{Im}(R)$. Further, $\bar{\beta}_I(R, c_0) = \beta\left(\bar{\Theta}_{0I}(R, c_0)\right) \subseteq \{\beta \in \mathbb{R}^k : R\beta = c_0\} = \mathscr{L}_I(R, c_0)$: so, when $\bar{\beta}_I(R, c_0) = \{\beta \in \mathbb{R}^k : R\beta = c_0\}$, (A.37) entails

$$[(X\beta_1 = X\beta_2) \Rightarrow (Q\beta_1 = Q\beta_2)] \left(\forall \beta_1, \beta_2 \in \mathscr{L}_I(R, c_0)\right) \tag{A.38}$$

for any $c_0 \in \mathrm{Im}(R)$. Further, since (A.38) holds for any $c_0 \in \mathrm{Im}(R)$, we must have (A.36), hence [by Proposition **2.4.4**] each one of the conditions (2.5.28) - (2.5.36). This completes the proof. □

PROOF OF THEOREM **2.5.6** By Proposition **2.4.5** with $Z = X$, each one of the conditions (2.5.38) - (2.5.46) is equivalent to

$$[(XCd_1 = XCd_2) \Rightarrow (QCd_1 = QCd_2)](\forall d_1, d_2 \in \mathbb{R}^m) \tag{A.39}$$

hence, for any $b_0 \in \mathbb{R}^k$,

$$\left[\left(X\beta(\theta_1) = X\beta(\theta_2)\right) \Rightarrow \left(Q\beta(\theta_1) = Q\beta(\theta_2)\right)\right] \left(\forall \theta_1, \theta_2 \in \bar{\Theta}_{0E}(C, b_0)\right). \tag{A.40}$$

This means that $Q\beta$ is $(X\beta)$-identifiable on $\bar{\Theta}_{0E}(C, b_0)$, for any $b_0 \in \mathbb{R}^k$.
Conversely, suppose $Q\beta$ is $(X\beta)$-identifiable on $\bar{\Theta}_{0E}(C, b_0)$, for any $b_0 \in \mathbb{R}^k$. Under Assumption **2.5.2**, Theorem **2.5.3** then entails that (A.40) holds for any $b_0 \in \mathbb{R}^k$. Further, $\bar{\beta}_E(C, b_0) = \beta\left(\bar{\Theta}_{0E}(C, b_0)\right) \subseteq \{\beta \in \mathbb{R}^k : \beta = b_0 + Cy,\ y \in \mathbb{R}^m\} = \mathscr{L}_E(C, b_0)$: so, when $\bar{\beta}_E(C, b_0) = \{\beta \in \mathbb{R}^k : \beta = b_0 + Cy,\ y \in \mathbb{R}^m\}$, (A.35) entails

$$\left[(X\beta_1 = X\beta_2) \Rightarrow (Q\beta_1 = Q\beta_2)\right] \left(\forall \beta_1, \beta_2 \in \mathscr{L}_E(C, b_0)\right) \tag{A.41}$$

for any $b_0 \in \mathbb{R}^k$. Further, since (A.41) holds for any $b_0 \in \mathbb{R}^k$, we must have (A.39), hence [by Proposition **2.4.5**] each one of the conditions (2.5.38) - (2.5.46). This completes the proof. □

PROOF OF PROPOSITION **2.5.7** This proposition follows on applying Theorem **2.5.6** along with Proposition **2.4.6** with $Z = X$. □

Corollary A.1 *Given a linear parametric model*

$$y = X\beta + u,$$

where $u \sim \mathrm{N}[0, \sigma^2\mathrm{I}]$ *and* X *is fixed, the mean is sufficient to identify the parameters* β *in terms of the family of a normal probability distribution.*

PROOF. We want to show that we can identify the parameter β as long as the mean can be identifiable in a normal distribution setup. To be more specific, let's define a parameter vector as follows,

$$\theta = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}.$$

By the definition of identification, we know that if $\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2} \Longrightarrow \theta_1 = \theta_2$, then θ is identifiable. Here we want to show that the above identification condition is equivalent to

$$\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2} \Longrightarrow \beta_1 = \beta_2 \iff \theta_1 = \theta_2.$$

Consider the classical linear model $y = X\beta + \mu$, where $\mu \sim \mathrm{N}[0, \sigma^2\mathrm{I}]$ and X is fixed.

$$
\begin{aligned}
\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2} &\iff F(y,\theta_1) = F(y,\theta_2) \\
&\iff \frac{1}{\left(\sqrt{2\pi\sigma_1^2}\right)^n} \exp\left[-\frac{1}{2\sigma_1^2}\|y - X\beta_1\|^2\right] \\
&= \frac{1}{\left(\sqrt{2\pi\sigma_2^2}\right)^n} \exp\left[-\frac{1}{2\sigma_2^2}\|y - X\beta_2\|^2\right].
\end{aligned}
$$

Taking logarithm on both sides, we have

$$
\begin{aligned}
\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2} &\iff -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma_1^2 - \frac{1}{2\sigma_1^2}\|y - X\beta_1\|^2 \\
&= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma_2^2 - \frac{1}{2\sigma_2^2}\|y - X\beta_2\|^2 \\
&\iff \frac{n}{2}\log\frac{\sigma_1^2}{\sigma_2^2} = \frac{1}{2\sigma_1^2}\|y - X\beta_1\|^2 - \frac{1}{2\sigma_2^2}\|y - X\beta_2\|^2.
\end{aligned}
$$

As can be seen,

$$
\begin{aligned}
\sigma_1^2 = \sigma_2^2 = \sigma^2 &\Longrightarrow \frac{1}{2\sigma^2}\left[\|y - X\beta_1\|^2 - \|y - X\beta_2\|^2\right] = 0 && \text{(A.42)} \\
&\Longrightarrow y'y - 2y'X\beta_1 + \beta_1'X'X\beta_1 - y'y + 2y'X\beta_2 - \beta_2'X'X\beta_2 = 0 && \text{(A.43)} \\
&\Longrightarrow 2y'X(\beta_2 - \beta_1) + (\beta_1'X'X\beta_1 - \beta_2'X'X\beta_2) = 0, \forall y. && \text{(A.44)}
\end{aligned}
$$

Suppose $\beta_1 \neq \beta_2$, then we can always take y large enough so that equation(A.44) does not hold. Thus we must require that $\beta_1 = \beta_2$ to make equation(A.44) hold for any y. That is to say $\sigma_1^2 = \sigma_2^2 \Longrightarrow \beta_1 = \beta_2$.
On the other hand,

$$
\beta_1 = \beta_2 = \beta \Longrightarrow \frac{1}{2}\|y - X\beta\|^2\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) = \frac{n}{2}\log\frac{\sigma_1^2}{\sigma_2^2}, \forall y. \tag{A.45}
$$

Similarly by contradiction, suppose $\sigma_1^2 \neq \sigma_2^2$. Since σ_1^2, σ_2^2 and β are parameters and n is given, the right hand side of equation(A.45) is actually a constant but the left hand side of equation(A.45) is changing with y. Therefore, equation(A.45) cannot hold for any y if $\sigma_1^2 \neq \sigma_2^2$. That is to say $\beta_1 = \beta_2 \Longrightarrow \sigma_1^2 = \sigma_2^2$. To summarize, under the normality assumption, the necessary and sufficient condition for the identification of θ, i.e., $\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2} \Longrightarrow \theta_1 = \theta_2$ is equivalent to $\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2} \iff \beta_1 = \beta_2 \iff \theta_1 = \theta_2$. This completes the proof. □

PROOF OF THEOREM **3.2.7** Let's denote all the deep parameters in the linear system as θ and set

$$\gamma(\theta) = \mathbb{E}(\overline{Y})A$$

and

$$\beta(\theta) = A.$$

Since $\mathbb{E}(Z)$ is the first moment of observable random variable Z, it is identifiable when it exists. Meanwhile it is obvious

$$\left(\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2}\right) \Rightarrow \left(\mathbb{E}(\overline{Y})A_1 = \mathbb{E}(\overline{Y})A_2\right).$$

From Definition **3.2.1** and **3.2.2**, A is identifiable if

$$\left(\mathbb{E}(\overline{Y})A_1 = \mathbb{E}(\overline{Y})A_2\right) \Rightarrow (A_1 = A_2), \tag{A.46}$$

which is equivalent to stating that the linear system (3.2.3) has a unique solution. This completes the proof. □

PROOF OF THEOREM **3.2.9** Sufficiency is shown in the proof of Theorem **3.2.7**. The establishment of necessity is a direct consequence of Dufour and Liang (2012, Theorem 2.3.4) if we replace $\gamma_1(\theta)$, $\gamma_2(\theta)$ and $\beta(\theta)$ with $\mathbb{E}(Z)$, the nuisance parameter and A respectively. This completes the proof. □

PROOF OF PROPOSITION **3.2.10** The proof is the same as that of Theorem **3.2.9** except by replacing $\mathbb{E}(\overline{Y})$ with $\mathbb{E}(\overline{Y}|X)$. This completes the proof. □

PROOF OF PROPOSITION **3.2.11** Substituting $\mathbf{P}_L(\overline{Y}|X)$ for $\mathbb{E}(\overline{Y}|X)$ in the proof of Proposition **3.2.10** will lead to the result. This completes the proof. □

PROOF OF LEMMA **3.2.12** To ease notation, let's denote the j-th and i-th column of $\overline{Y}$ and X as $\overline{Y}_j$ and X_i, where $j = 1, 2, \ldots, G$ and $i = 1, 2, \ldots, K$. Suppose

$$\mathbb{E}(\overline{Y}|X) = 0$$

which means

$$\mathbb{E}(\overline{Y}_j|X) = 0.$$

Then

$$\mathbb{E}(\overline{Y}_j) = 0$$

and

$$\mathbb{E}(X_i \overline{Y}_j{}') = 0.$$

Hence

$$\mathbb{C}(X_i, \overline{Y}_j) = \mathbb{E}(X_i \overline{Y}_j{}') - \mathbb{E}(X_i)\mathbb{E}(\overline{Y}_j{}') = 0.$$

Since the above arguments are true for any $\overline{Y}_j$ and X_i, where $j = 1, 2, \ldots, G$ and $i = 1, 2, \ldots, K$, it follows from (3.2.7) that

$$\mathbf{P}_L(\overline{Y}|X) = 0.$$

This completes the proof. □

PROOF OF PROPOSITION **3.3.2** From Rao and Mitra (1971, Theorem 2.3.2), the general solution of A to the linear system

$$\mathbb{E}(\overline{Y}|X)A = \mathbb{E}(Z|X)$$

is

$$A^* = \big(\mathbb{E}(\overline{Y}|X)\big)^- \mathbb{E}(Z|X) + M - \big(\mathbb{E}(\overline{Y}|X)\big)^- \big(\mathbb{E}(\overline{Y}|X)\big) M,$$

where M is an arbitrary $G \times H$ matrix. It follows that

$$\begin{aligned}
&\Big[Q(X)\Big(\big(\mathbb{E}(\overline{Y}|X)\big)^- \mathbb{E}(Z|X) + M - \big(\mathbb{E}(\overline{Y}|X)\big)^- \big(\mathbb{E}(\overline{Y}|X)\big) M\Big) = Q(X)\big(\mathbb{E}(\overline{Y}|X)\big)^- \mathbb{E}(Z|X),\ \forall M\Big] \\
&\iff \Big[Q(X)M = Q(X)\big(\mathbb{E}(\overline{Y}|X)\big)^- \mathbb{E}(\overline{Y}|X)M,\ \forall M\Big] \\
&\iff \Big[Q(X) = Q(X)\big(\mathbb{E}(\overline{Y}|X)\big)^- \mathbb{E}(\overline{Y}|X)\Big] \\
&\iff \Big[\mathrm{Im}\Big(\big(Q(X)\big)'\Big) \subseteq \mathrm{Im}\Big(\big(\mathbb{E}(\overline{Y}|X)\big)'\Big)\Big].
\end{aligned}$$

The last equivalent sign is due to Rao and Mitra (1971, Lemma 2.2.4). This completes the proof. □

PROOF OF PROPOSITION **3.3.3** Let's define $\Delta \equiv B_1 - B_2$ and let δ be any given column of Δ.

First, we show the equivalence among the statements from (3.3.1) through (3.3.6). Obviously, the equivalence between (3.3.1) and (3.3.2) simply follows from the definition of

identification and that of a function.

$$
\begin{aligned}
&\Big[\big(\mathbb{E}(Y|X)B_1 = \mathbb{E}(Y|X)B_2\big) \Rightarrow \big(Q(X)B_1 = Q(X)B_2\big), \forall B_1, B_2 \in \mathbb{R}^{G\times H}\Big] \\
&\iff \Big[\big(\mathbb{E}(Y|X)\Delta = \mathbf{0}\big) \Rightarrow \big(Q(X)\Delta = \mathbf{0}\big), \forall \Delta \in \mathbb{R}^{G\times H}\Big] \\
&\iff \Big[\Big(\delta \in \ker\big(\mathbb{E}(Y|X)\big)\Big) \Rightarrow \Big(\delta \in \ker\big(Q(X)\big)\Big), \forall \delta \in \Delta\Big] \\
&\iff \Big[\ker\big(\mathbb{E}(Y|X)\big) \subseteq \ker\big(Q(X)\big)\Big] \\
&\iff \Big[\Big[\mathrm{Im}\big((\mathbb{E}(Y|X))'\big)\Big]^{\perp} \subseteq \Big[\mathrm{Im}\big((Q(X))'\big)\Big]^{\perp}\Big] \\
&\iff \Big[\mathrm{Im}\Big(\big(Q(X)\big)'\Big) \subseteq \mathrm{Im}\Big(\big(\mathbb{E}(Y|X)\big)'\Big)\Big] \\
&\iff \Big[\big(Q(X)\big)' = \big(\mathbb{E}(Y|X)\big)'D', \text{ for some matrix } D\Big] \\
&\iff \mathrm{rank}\begin{bmatrix} \big(\mathbb{E}(Y|X)\big)' & \big(Q(X)\big)' \end{bmatrix} = \mathrm{rank}\Big(\big(\mathbb{E}(Y|X)\big)'\Big) \\
&\iff \mathrm{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} = \mathrm{rank}\big(\mathbb{E}(Y|X)\big).
\end{aligned}
$$

Second, we show (3.3.6), (3.3.7) and (3.3.8) are equivalent. Let's construct a lower triangular matrix T_L and an upper triangular matrix T_U with conformable dimensions

$$
T_U = \begin{bmatrix} I & V_1 \\ O & I \end{bmatrix} \text{ and } T_L = \begin{bmatrix} I & O \\ V_2 & I \end{bmatrix},
$$

where V_1 and V_2 are arbitrary. Since both triangular matrices are nonsingular we can express their inverses as

$$
T_U^{-1} = \begin{bmatrix} I & -V_1 \\ O & I \end{bmatrix} \text{ and } T_L^{-1} = \begin{bmatrix} I & O \\ -V_2 & I \end{bmatrix}.
$$

Then

$$
\mathrm{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} = \mathrm{rank}\left\{\begin{bmatrix} I & V_1 \\ O & I \end{bmatrix}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix}\right\} = \mathrm{rank}\begin{bmatrix} \mathbb{E}(Y|X) + V_1 Q(X) \\ Q(X) \end{bmatrix}.
$$

and

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} = \operatorname{rank}\left\{ \begin{bmatrix} I & O \\ V_2 & I \end{bmatrix} \begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} \right\} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) + V_2\mathbb{E}(Y|X) \end{bmatrix}$$

Since both V_1 and V_2 are arbitrarily chosen it is always possible to construct some upper and lower triangular matrices such that

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) + V_1 Q(X) \\ Q(X) \end{bmatrix}$$

and

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} = \operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) + V_2\mathbb{E}(Y|X) \end{bmatrix}.$$

Clearly, the statements (3.3.6), (3.3.7) and (3.3.8) are equivalent.

Third, we prove the equivalence between (3.3.6) and (3.3.9). Let's start by demonstrating that condition (3.3.6) implies (3.3.9). We have shown that

$$\left[\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(Y|X)\big)\right] \iff \left[\operatorname{Im}\Big(\big(Q(X)\big)'\Big) \subseteq \operatorname{Im}\Big(\big(\mathbb{E}(Y|X)\big)'\Big)\right].$$

Since

$$\operatorname{Im}\Big(\big(Q(X)\big)'S'\Big) \subseteq \operatorname{Im}\Big(\big(Q(X)\big)'\Big),$$

we have

$$\operatorname{Im}\Big(\big(Q(X)\big)'S'\Big) \subseteq \operatorname{Im}\Big(\big(\mathbb{E}(Y|X)\big)'\Big),$$

which implies

$$\operatorname{rank}\begin{bmatrix} \big(\mathbb{E}(Y|X)\big)' & \big(Q(X)\big)'S' \end{bmatrix} = \operatorname{rank}\Big(\big(\mathbb{E}(Y|X)\big)'\Big).$$

In other words, we obtain

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ SQ(X) \end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(Y|X)\big), \text{ for some matrix } S \text{ with a dimension } s \times q.$$

To show that (3.3.9) implies (3.3.6), we notice

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ SQ(X) \end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(Y|X)\big) \iff \operatorname{rank}\begin{bmatrix} \big(\mathbb{E}(Y|X)\big)' & \big(Q(X)\big)'S' \end{bmatrix} = \operatorname{rank}\Big(\big(\mathbb{E}(Y|X)\big)'\Big)$$
$$\iff \operatorname{Im}\Big(\big(Q(X)\big)'S'\Big) \subseteq \operatorname{Im}\Big(\big(\mathbb{E}(Y|X)\big)'\Big).$$

Since

$$\operatorname{rank}\big(SQ(X)\big) = \operatorname{rank}\big(Q(X)\big) = s$$

and

$$\operatorname{Im}\Big(\big(Q(X)\big)'S'\Big) \subseteq \operatorname{Im}\Big(\big(Q(X)\big)'\Big),$$

we have

$$\operatorname{Im}\Big(\big(Q(X)\big)'S'\Big) = \operatorname{Im}\Big(\big(Q(X)\big)'\Big).$$

Therefore

$$\operatorname{Im}\Big(\big(Q(X)\big)'\Big) \subseteq \operatorname{Im}\Big(\big(\mathbb{E}(Y|X)\big)'\Big)$$

and this is equivalent to

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(Y|X)\big).$$

Thus (3.3.6) and (3.3.9) are equivalent.

Finally, we establish the equivalence among conditions (3.3.6), (3.3.10) and (3.3.11). The crucial step here is to find some conformable matrices V_1 and V_2 such that the row spaces of $\mathbb{E}(Y|X)+V_1Q(X)$ and $Q(X)$ are essentially disjoint and the row spaces of $\mathbb{E}(Y|X)$ and $Q(X)+V_2\mathbb{E}(Y|X)$ are essentially disjoint as well. That is to say, we need to show

$$\operatorname{Im}\Big(\big(\mathbb{E}(Y|X)\big)' + \big(Q(X)\big)'V_1'\Big) \cap \operatorname{Im}\Big(\big(Q(X)\big)'\Big) = \{0\}$$

and

$$\operatorname{Im}\Big(\big(\mathbb{E}(Y|X)\big)'\Big) \cap \operatorname{Im}\Big(\big(Q(X)'\big) + \big(\mathbb{E}(Y|X)\big)'V_2'\Big) = \{0\}.$$

Since V_1 is arbitrary by construction we can take $V_1 = -\mathbb{E}(Y|X)\big(Q(X)\big)^{-}$. Then

$$\mathbb{E}(Y|X) + V_1 Q(X) = \mathbb{E}(Y|X) - \mathbb{E}(Y|X)\big(Q(X)\big)^{-} Q(X) = \mathbb{E}(Y|X)\Big(I - \big(Q(X)\big)^{-} Q(X)\Big).$$

Suppose

$$\mathrm{Im}\Big(\big(\mathbb{E}(Y|X)\big)' + \big(Q(X)\big)' V_1'\Big) \cap \mathrm{Im}\Big(\big(Q(X)\big)'\Big) \neq \{0\},$$

then there exist non-zero vectors x_1 and x_2 such that

$$\Big(\big(\mathbb{E}(Y|X)\big)' + \big(Q(X)\big)' V_1'\Big) x_1 = \big(Q(X)\big)' x_2 \neq 0.$$

Hence

$$\Big(x_1'\big(\mathbb{E}(Y|X) + V_1 Q(X)\big) = x_2' Q(X) \neq 0\Big) \Rightarrow \Big(x_1' \mathbb{E}(Y|X)\Big(I - \big(Q(X)\big)^{-} Q(X)\Big) = x_2' Q(X) \neq 0\Big).$$

If we multiply both sides by $\Big(I - \big(Q(X)\big)^{-} Q(X)\Big)$, we have

$$x_1' \mathbb{E}(Y|X)\Big(I - \big(Q(X)\big)^{-} Q(X)\Big)\Big(I - \big(Q(X)\big)^{-} Q(X)\Big) = x_1' \mathbb{E}(Y|X)\Big(I - \big(Q(X)\big)^{-} Q(X)\Big) \neq 0,$$

and

$$x_2' Q(X)\Big(I - \big(Q(X)\big)^{-} Q(X)\Big) = 0.$$

This is a contradiction and thus

$$\mathrm{Im}\Big(\big(\mathbb{E}(Y|X)\big)' + \big(Q(X)\big)' V_1'\Big) \cap \mathrm{Im}\Big(\big(Q(X)\big)'\Big) = \{0\}.$$

Therefore,

$$\begin{aligned} \mathrm{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} &= \mathrm{rank}\begin{bmatrix} \mathbb{E}(Y|X) - \mathbb{E}(Y|X)\big(Q(X)\big)^{-} Q(X) \\ Q(X) \end{bmatrix} \\ &= \mathrm{rank}\Big(\mathbb{E}(Y|X) - \mathbb{E}(Y|X)\big(Q(X)\big)^{-} Q(X)\Big) + \mathrm{rank}\big(Q(X)\big). \end{aligned}$$

This is equivalent to

$$\operatorname{rank}\begin{bmatrix} \mathbb{E}(Y|X) \\ Q(X) \end{bmatrix} = \operatorname{rank}\Big(\mathbb{E}(Y|X)\Big(I - \big(Q(X)\big)^{-}Q(X)\Big)\Big) + \operatorname{rank}\big(Q(X)\big). \tag{A.47}$$

Similarly, we can take $V_2 = -Q(X)\big(\mathbb{E}(Y|X)\big)^{-}$. Then

$$Q(X) + V_2\mathbb{E}(Y|X) = Q(X) - Q(X)\big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X) = Q(X)\Big(I - \big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big).$$

Suppose

$$\operatorname{Im}\Big(\big(\mathbb{E}(Y|X)\big)'\Big) \cap \operatorname{Im}\Big(\big(Q(X)'\big) + \big(\mathbb{E}(Y|X)\big)'V_2'\Big) \neq \{0\}.$$

Then we can always find some non-zero vectors x_1 and x_2 such that

$$\big(\mathbb{E}(Y|X)\big)'x_1 = \Big(\big(Q(X)'\big) + \big(\mathbb{E}(Y|X)\big)'V_2'\Big)x_2 \neq 0.$$

Thus

$$x_1{}'\mathbb{E}(Y|X) = x_2{}'\big(Q(X) + V_2\mathbb{E}(Y|X)\big) = x_2{}'Q(X)\Big(I - \big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big) \neq 0.$$

If we multiply both sides by $\Big(I - \big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big)$, we get

$$x_1{}'\mathbb{E}(Y|X)\Big(I - \big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big) = x_1{}'\mathbb{E}(Y|X) - x_1{}'\mathbb{E}(Y|X) = 0$$

and

$$x_2{}'Q(X)\Big(I - \big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big)\Big(I - \big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big) = x_2{}'Q(X)\Big(I - \big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big) \neq 0.$$

Again, this leads to the contradiction and thus

$$\operatorname{Im}\Big(\big(\mathbb{E}(Y|X)\big)'\Big) \cap \operatorname{Im}\Big(\big(Q(X)'\big) + \big(\mathbb{E}(Y|X)\big)'V_2'\Big) = \{0\}.$$

Likely, we have

$$\begin{aligned}\operatorname{rank}\begin{bmatrix}\mathbb{E}(Y|X)\\ Q(X)\end{bmatrix} &= \operatorname{rank}\begin{bmatrix}\mathbb{E}(Y|X)\\ Q(X)-Q(X)\big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\end{bmatrix}\\ &= \operatorname{rank}\big(\mathbb{E}(Y|X)\big)+\operatorname{rank}\Big(Q(X)-Q(X)\big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big).\end{aligned}$$

This is equivalent to

$$\operatorname{rank}\begin{bmatrix}\mathbb{E}(Y|X)\\ Q(X)\end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(Y|X)\big)+\operatorname{rank}\Big(Q(X)\Big(I-\big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big)\Big). \quad \text{(A.48)}$$

Combining the above results, we have

$$\begin{aligned}\operatorname{rank}\begin{bmatrix}\mathbb{E}(Y|X)\\ Q(X)\end{bmatrix} &= \operatorname{rank}\Big(\mathbb{E}(Y|X)\Big(I-\big(Q(X)\big)^{-}Q(X)\Big)\Big)+\operatorname{rank}\big(Q(X)\big)\\ &= \operatorname{rank}\big(\mathbb{E}(Y|X)\big)+\operatorname{rank}\Big(Q(X)\Big(I-\big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big)\Big).\end{aligned}$$

Given this equation holds, it is obvious that

$$\begin{aligned}&\operatorname{rank}\begin{bmatrix}\mathbb{E}(Y|X)\\ Q(X)\end{bmatrix} = \operatorname{rank}\big(\mathbb{E}(Y|X)\big)\\ \iff\ &\operatorname{rank}\Big(Q(X)\Big(I-\big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big)\Big)=0\\ \iff\ &\operatorname{rank}\big(\mathbb{E}(Y|X)\big)=\operatorname{rank}\Big(\mathbb{E}(Y|X)\Big(I-\big(Q(X)\big)^{-}Q(X)\Big)\Big)+\operatorname{rank}\big(Q(X)\big)\\ \iff\ &\Big[Q(X)\Big(I-\big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X)\Big)=0\Big]\\ \iff\ &\Big[Q(X)=Q(X)\big(\mathbb{E}(Y|X)\big)^{-}\mathbb{E}(Y|X),\ \text{for some g-inverse } \big(\mathbb{E}(Y|X)\big)^{-}\Big].\end{aligned}$$

This completes the proof. □

PROOF OF THEOREM **3.3.4** The proof is the same as that of Dufour and Liang (2012, Theorem 2.5.5) except by replacing $\bar{X}$, R and β with $\widetilde{X}$, $R(X)$ and A. This completes the proof. □

PROOF OF THEOREM **3.3.5** The proof is the same as that of Dufour and Liang (2012,

Theorem 2.5.6) except by replacing X, Q and β with $\mathbb{E}(\overline{Y}|X)$, $Q(X)$ and A. This completes the proof. □

PROOF OF THEOREM **4.2.14** Suppose (4.2.12) holds but θ is not locally identifiable at θ_0. Then within any open neighborhood of θ_0, we can always construct an infinite sequence of parameters $\{\theta_m\}_{m=1}^{\infty}$ such that

$$\theta_m \neq \theta_0, \quad \mathsf{P}_{\theta_m} = \mathsf{P}_{\theta_0} \quad \text{and} \quad \theta_m \to \theta_0 \text{ as } m \to \infty.$$

Since $\gamma(\theta)$ is locally identifiable at θ_0, we must also have

$$\gamma(\theta_m) = \gamma(\theta_0), \quad \text{for } m \geq m_0, \tag{A.49}$$

where $m_0 < \infty$. By the differentiability assumption [see Rudin (1976)], it follows that

$$\lim_{m\to\infty} \frac{\|J_\gamma(\theta_0)(\theta_m - \theta_0)\|}{\|\theta_m - \theta_0\|} = \lim_{m\to\infty} \frac{\|\gamma(\theta_m) - \gamma(\theta_0) - J_\gamma(\theta_0)(\theta_m - \theta_0)\|}{\|\theta_m - \theta_0\|} = 0. \tag{A.50}$$

Let

$$d_m = \frac{\theta_m - \theta_0}{\|\theta_m - \theta_0\|}, \quad m = 1, 2, \ldots.$$

Clearly,

$$\|d_m\| = 1, \quad \forall m \geq 1.$$

Since the infinite sequence $\{d_m\}_{m=1}^{\infty}$ is on the unit sphere, it is bounded. Thus there exists a convergent subsequence

$$\{d_{m_j}\}_{j=1}^{\infty} \subseteq \{d_m\}_{m=1}^{\infty}$$

such that

$$d_{m_j} \to d_0 \text{ as } j \to \infty,$$

where $d_0 \in \mathbb{R}^k$ and $\|d_0\| = 1$. Since

$$d_{m_j} = \frac{\theta_{m_j} - \theta_0}{\|\theta_{m_j} - \theta_0\|}$$

we have $\|d_{m_j}\| = 1$ and

$$\theta_{m_j} - \theta_0 = d_{m_j}\|\theta_{m_j} - \theta_0\|.$$

By (A.50), it follows that

$$\begin{aligned}\lim_{j\to\infty}\frac{\|J_\gamma(\theta_0)(\theta_{m_j}-\theta_0)\|}{\|\theta_{m_j}-\theta_0\|} &= \lim_{j\to\infty}\frac{\|J_\gamma(\theta_0)d_{m_j}\|\theta_{m_j}-\theta_0\|\|}{\|\theta_{m_j}-\theta_0\|}\\ &= \lim_{j\to\infty}\|J_\gamma(\theta_0)d_{m_j}\|\\ &= 0.\end{aligned}$$

Since

$$d_{m_j}\to d_0\neq 0 \text{ as } j\to\infty,$$

we have

$$J_\gamma(\theta_0)d_{m_j}\to J_\gamma(\theta_0)d_0 \text{ as } j\to\infty.$$

In other words,

$$\|J_\gamma(\theta_0)d_0\|=0.$$

which is equivalent to

$$J_\gamma(\theta_0)d_0=0,$$

where $d_0\neq 0$. It follows that

$$\text{rank}[J_\gamma(\theta_0)]<k.$$

This contradicts (4.2.12). Consequently, the rank condition

$$\text{rank}[J_\gamma(\theta_0)]=k$$

entails that θ is locally identifiable at θ_0. This completes the proof. □

PROOF OF THEOREM **4.2.15** The establishment of sufficiency is the same as that of Theorem **4.2.14**. We next show that the necessity holds as well. Suppose that $J_\gamma(\theta)$ does not have a full column rank in an open neighborhood of θ_0 which we denote as $\mathscr{V}(\theta_0)$. In other words,

$$\text{rank}[J_\gamma(\theta)'J_\gamma(\theta)]=\text{rank}[J_\gamma(\theta)]=\bar{k}<k,\quad \forall\theta\in\mathscr{V}(\theta_0).$$

This means $J_\gamma(\theta)'J_\gamma(\theta)$ has at least one zero eigenvalue with possible algebraic multiplicity larger than one. Let's denote the distinct eigenvalues of $J_\gamma(\theta)'J_\gamma(\theta)$ as $\lambda_1(\theta)$, $\lambda_2(\theta)$, …, $\lambda_j(\theta)$ and their associated eigenprojections as $\mathbf{P}_{\lambda_1}(\theta)$, $\mathbf{P}_{\lambda_2}(\theta)$, …, $\mathbf{P}_{\lambda_j}(\theta)$. Denote the geometric multiplicity of λ_i as ν_i and the

corresponding orthonormal eigenvectors as $q_{i1}(\theta)$, $q_{i2}(\theta)$, $\ldots$, $q_{iv_i}(\theta)$ which span the eigenspace associated with λ_i. From the spectral decomposition of the real symmetric matrix [see Harville (2008)], it follows that

$$J_\gamma(\theta)' J_\gamma(\theta) = \sum_{m=1}^{j} \lambda_m(\theta) \mathbf{P}_{\lambda_m}(\theta),$$

where

$$\mathbf{P}_{\lambda_i}(\theta) = \sum_{l=1}^{v_i} q_{il}(\theta) q_{il}(\theta)'.$$

Without loss of generality, take

$$\lambda_1 = 0$$

and choose the first nonzero column of $\mathbf{P}_{\lambda_1}(\theta)$ and denote it as $p_{\lambda_1}(\theta)$. Hence

$$p_{\lambda_1}(\theta)' J_\gamma(\theta)' J_\gamma(\theta) p_{\lambda_1}(\theta) = 0, \quad \forall \theta \in \mathscr{V}(\theta_0). \tag{A.51}$$

Since $\gamma(\theta)$ is continuously differentiable in $\mathscr{V}(\theta_0)$, it follows that $J_\gamma(\theta)' J_\gamma(\theta)$ is continuous in $\mathscr{V}(\theta_0)$. According to Tyler (1981), both the eigenvalues and eigenprojection matrices of $J_\gamma(\theta)' J_\gamma(\theta)$ are continuous. Since the composite of two continuous functions is continuous, $p_{\lambda_1}(\theta)$ can be chosen to be continuous in the neighborhood $\mathscr{V}(\theta_0)$. Let's define a mapping $\theta(t)$ which is the solution to the differential equation for $t \in [0, t^*]$

$$\frac{\partial \theta(t)}{\partial t} = p_{\lambda_1}(\theta) \tag{A.52}$$

and

$$\theta(0) = \theta_0.$$

From (A.51), it follows that

$$\frac{\partial \gamma(\theta)}{\partial t} = J_\gamma(\theta) p_{\lambda_1}(\theta) = 0.$$

This implies that $\gamma(\theta)$ is constant along the curve $\theta(t)$ on $[0, t^*]$ and therefore constant in $\theta(t)$ within the neighborhood $\mathscr{V}(\theta_0)$. Since $p_{\lambda_1}(\theta)$ is chosen to be nonzero, (A.52) shows that $\theta(t)$ cannot be constant at least within a small neighborhood of some point $t \in [0, t^*]$. Therefore the following statements hold simultaneously for at least some $\theta \in \mathscr{V}(\theta_0)$

$$\mathsf{P}_\theta = \mathsf{P}_{\theta_0}, \quad \gamma(\theta(t)) = \gamma(\theta_0) \quad \text{and} \quad \theta(t) \neq \theta_0.$$

From Definition **4.2.5**, we conclude that θ is not locally identifiable at θ_0 which contradicts the assumption that θ is locally identifiable at θ_0. Thus

$$(\theta \text{ is locally identifiable at } \theta_0) \Rightarrow (\text{rank}[J(\theta_0)] = k).$$

This completes the proof. □

PROOF OF THEOREM **4.3.5** According to Rudin (1976), if $\gamma(\theta)$ is continuously differentiable at θ_0 and the Jacobian matrix $J_\gamma(\theta_0)$ has a full column rank, we conclude by the inverse function theorem that there exist open subsets $\mathscr{A}(\theta_0) \subset \Theta$ with center at θ_0 and $\mathscr{B} \subset \mathbb{R}^G$ such that $\gamma(\theta)$ is one-to-one on $\mathscr{A}(\theta_0)$ and

$$\gamma(\mathscr{A}(\theta_0)) = \mathscr{B}.$$

If $\gamma(\theta)$ is locally identifiable around θ_0 and one-to-one on some open neighborhood of θ_0, then it follows that θ is locally identifiable around θ_0 by Definition **4.3.1** and Definition **4.3.2**. On the other hand, if θ is locally identifiable around θ_0, it follows that every θ in such an open neighborhood of θ_0 is locally identifiable at θ_0 due to Definition **4.2.5**. Then Theorem **4.2.15** leads to

$$\text{rank}[J_\gamma(\theta_0)] = k.$$

This completes the proof. □

PROOF OF THEOREM **4.3.7** From the stronger version of the inverse function theorem by Krantz and Parks (2002, Theorem 3.3.2), it follows that there exist open subsets $\mathscr{A}(\theta_0) \subset \Theta$ with center at θ_0 and $\mathscr{B} \subset \mathbb{R}^G$ such that $\gamma(\theta)$ is one-to-one on $\mathscr{A}(\theta_0)$. Besides, its inverse function denoted as γ^{-1} is also p-th order continuously differentiable. Thus local identification of $\gamma(\theta)$ around θ_0 implies local identification of θ around θ_0. The necessity holds trivially. This completes the proof. □

PROOF OF PROPOSITION **4.3.8** Since θ is locally identifiable at θ_0, there must exists a transformation $g : \mathbb{R}^G \mapsto \mathbb{R}^k$ such that for some open neighborhood $\mathscr{V}(\theta_0)$

$$\theta = g(\gamma(\theta)), \quad \forall \theta \in \mathscr{V}(\theta_0). \tag{A.53}$$

Suppose there does not exist such a function g. Then we must have

$$\gamma(\theta_1) = \gamma(\theta_2)$$

and

$$\theta_1 \neq \theta_2$$

for some $\theta \in \mathscr{V}(\theta_0)$. However, this contradicts our assumptions that both $\gamma(\theta)$ and θ are locally identifiable at θ_0 (see Definition **4.2.6**). Obviously, g is the inverse function of γ. Thus taking the partial derivative with respect to θ on both sides of (A.53) gives

$$I_k = \frac{\partial g}{\partial \gamma'} J_\gamma(\theta), \quad \forall \theta \in \mathscr{V}(\theta_0). \tag{A.54}$$

Suppose an arbitrary k-vector $\phi(\theta) \in \ker(J_\gamma(\theta))$. From (A.54), we have

$$\phi(\theta) = \frac{\partial g}{\partial \gamma'} J_\gamma(\theta)\phi(\theta) = 0, \quad \forall \theta \in \mathscr{V}(\theta_0).$$

That is to say, the only element in the kernel space of the Jacobian matrix $J_\gamma(\theta)$ in a neighborhood of θ_0 is the zero vector. Thus

$$\dim\big(\mathrm{col}(J_\gamma(\theta))\big) = k - \dim\big(\ker(J_\gamma(\theta))\big) = k, \quad \forall \theta \in \mathscr{V}(\theta_0).$$

This implies that

$$\mathrm{rank}[J_\gamma(\theta)] = k, \quad \forall \theta \in \mathscr{V}(\theta_0).$$

This completes the proof. □

PROOF OF THEOREM **4.4.3** We first prove sufficiency. Let's distinguish between two situations:

$$(a)\ J_\gamma(\theta_0) = 0;$$
$$(b)\ J_\gamma(\theta_0) \neq 0.$$

If $J_\gamma(\theta_0) = 0$, we have $\ker(J_\gamma(\theta_0)) = \mathbb{R}^k$ so that

$$\begin{aligned}\big[\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big)\big] &\iff \big[\ker(J_\gamma(\theta_0)) \subseteq \ker(J_\beta(\theta_0))\big] \\ &\Rightarrow \Big[\ker(J_\beta(\theta_0)) = \mathbb{R}^k\Big] \\ &\iff [J_\beta(\theta_0) = 0].\end{aligned}$$

Meanwhile, since $\beta(\theta)$ is continuously differentiable on $\mathscr{V}(\theta_0)$ and θ_0 is a regular point of $J_\beta(\theta)$, it follows from the rank theorem [see Rudin (1976, Theorem 9.32)] that there

exist an open neighborhood $\mathscr{E} \subset \mathbb{R}^k$ of η_0 and an open neighborhood $\mathscr{F} \subset \mathbb{R}^k$ of θ_0 such that there exists a local diffeomorphism $g : \mathscr{E} \mapsto \mathscr{F}$

$$\theta = g(\eta),\ \theta_0 = g(\eta_0)$$

and

$$\beta(\theta) = \beta(g(g^{-1}(\theta))) = \beta(g(\eta)) = J_\beta(\theta_0)\eta + \phi(J_\beta(\theta_0)\eta),\ \forall \eta \in \mathscr{E}.$$

If $J_\beta(\theta_0) = 0$, it follows that

$$\beta(\theta) = 0 + \phi(0),\ \forall \theta \in \mathscr{F},$$

which is a constant. Since the statement $\beta(\theta) = \beta(\theta_0)$ is true for $\forall \theta \in \mathscr{F}$, we have the following implication

$$(\gamma(\theta) = \gamma(\theta_0)) \Rightarrow (\beta(\theta) = \beta(\theta_0)),$$

for any θ in some open neighborhood of θ_0. Second, $J_\gamma(\theta_0) \neq 0$. Again it follows from the rank theorem that there exists a local diffeomorhpism g such that $\theta = g(\eta)$ and $\theta_0 = g(\eta_0)$. Meanwhile, we can rewrite $\beta(\theta)$ in terms of η within the neighborhood of θ_0 as

$$\beta(\theta) = \beta(g(\eta)) = \tilde{\beta}(\eta).$$

Then the chain rule leads to

$$J_\beta(\theta)J_g(\eta) = J_{\tilde{\beta}}(\eta)$$

which implies

$$J_\beta(\theta_0)J_g(\eta_0) = J_{\tilde{\beta}}(\eta_0).$$

Thus $\ker(J_g(\eta_0)) \subseteq \ker(J_{\tilde{\beta}}(\eta_0))$. Since the mapping g is a local diffeomorphism, we have

$$[\text{rank}(J_g(\eta_0)) = k] \iff \left[\ker(J_g(\eta_0)) = \mathbb{R}^k\right] \Rightarrow \left[\ker(J_{\tilde{\beta}}(\eta_0)) = \mathbb{R}^k\right].$$

On the other hand, for some point $\theta \in \mathscr{V}(\theta_0)$ and $\theta \neq \theta_0$, we can always define a normalized direction from θ_0 to θ as

$$h = \frac{\theta - \theta_0}{\|\theta - \theta_0\|}$$

and an infinite sequence $\{\theta_m\}_{m=1}^{\infty}$ along h such that

$$\theta_m = \theta_0 + t_m h,$$

where $t_m > 0$ and $t_m \to 0$ as $m \to \infty$. Accordingly, we can also define a normalized direction from η_0 to η as

$$h_\eta = \frac{\eta - \eta_0}{\|\eta - \eta_0\|}$$

and an infinite sequence $\{\eta_m\}_{m=1}^{\infty}$ along h_η such that

$$\eta_m = \eta_0 + d_m h_\eta,$$

where $d_m > 0$ and $d_m \to 0$ as $m \to \infty$. Therefore, the rank theorem gives

$$\begin{aligned}
[\beta(\theta_m) \neq \beta(\theta_0)] &\Rightarrow \left[J_\beta(\theta_0)\eta_m \neq J_\beta(\theta_0)\eta_0\right] \\
&\Rightarrow [\eta_m \neq \eta_0] \\
&\Rightarrow [J_g(\eta_0)\eta_m \neq J_g(\eta_0)\eta_0] \\
&\iff [(\eta_m - \eta_0) \notin \ker(J_g(\eta_0))] \\
&\Rightarrow \left[(\eta_m - \eta_0) \notin \ker(J_{\tilde{\beta}}(\eta_0))\right] \\
&\iff \left[J_{\tilde{\beta}}(\eta_0)\eta_m \neq J_{\tilde{\beta}}(\eta_0)\eta_0\right] \\
&\Rightarrow \left[J_{\tilde{\beta}}(\eta_0)\frac{\eta_m - \eta_0}{\|\eta_m - \eta_0\|} \neq 0\right] \\
&\Rightarrow \left[J_{\tilde{\beta}}(\eta_0)h_\eta \neq 0\right].
\end{aligned}$$

Then it follows that

$$\begin{aligned}
J_\beta(\theta_0)(\theta_m - \theta_0) &= J_\beta(\theta_0)(g(\eta_m) - g(\eta_0)) \\
&= J_\beta(\theta_0)\left(J_g(\eta_0)(\eta_m - \eta_0) + r_g(\eta_m - \eta_0)\right) \\
&= J_\beta(\theta_0)\left(\left[J_g(\eta_0)\frac{\eta_m - \eta_0}{\|\eta_m - \eta_0\|} + \frac{r_g(\eta_m - \eta_0)}{\|\eta_m - \eta_0\|}\right]\|\eta_m - \eta_0\|\right) \\
&= J_\beta(\theta_0)\left(\left[J_g(\eta_0)h_\eta + \frac{r_g(\eta_m - \eta_0)}{d_m}\right]d_m\right) \\
&= \left[J_\beta(\theta_0)J_g(\eta_0)h_\eta + J_\beta(\theta_0)\frac{r_g(\eta_m - \eta_0)}{d_m}\right]d_m \\
&= \left[J_{\tilde{\beta}}(\eta_0)h_\eta + J_\beta(\theta_0)\frac{r_g(\eta_m - \eta_0)}{d_m}\right]d_m,
\end{aligned}$$

where

$$\lim_{\eta_m \to \eta_0} \frac{\|r_g(\eta_m - \eta_0)\|}{\|\eta_m - \eta_0\|} = \lim_{m \to \infty} \frac{\|r_g(\eta_m - \eta_0)\|}{d_m} = 0.$$

Since $\|J_\beta(\theta_0)\|$ is bounded and $J_{\tilde{\beta}}(\eta_0)h_\eta \neq 0$, within the square brackets the second term is dominated by the first term. Thus $J_\beta(\theta_0)(\theta_m - \theta_0) \neq 0$. It then follows that

$$[\beta(\theta_m) \neq \beta(\theta_0)] \Rightarrow \left[J_\beta(\theta_0)(\theta_m - \theta_0) \neq 0\right].$$

By assumption $\ker(J_\gamma(\theta_0)) \subseteq \ker(J_\beta(\theta_0))$, we have

$$\begin{aligned}\left[J_\beta(\theta_0)(\theta_m - \theta_0) \neq 0\right] &\Rightarrow \left[J_\gamma(\theta_0)(\theta_m - \theta_0) \neq 0\right] \\ &\Rightarrow \left[J_\gamma(\theta_0)\frac{\theta_m - \theta_0}{\|\theta_m - \theta_0\|} \neq 0\right] \\ &\Rightarrow \left[J_\gamma(\theta_0)h \neq 0\right].\end{aligned}$$

Since $\gamma(\theta)$ is differentiable at θ_0, it follows that

$$\begin{aligned}\gamma(\theta_m) - \gamma(\theta_0) &= J_\gamma(\theta_0)(\theta_m - \theta_0) + r_\gamma(\theta_m - \theta_0) \\ &= \left[J_\gamma(\theta_0)\frac{\theta_m - \theta_0}{\|\theta_m - \theta_0\|} + \frac{r_\gamma(\theta_m - \theta_0)}{\|\theta_m - \theta_0\|}\right]\|\theta_m - \theta_0\| \\ &= \left[J_\gamma(\theta_0)h + \frac{r_\gamma(\theta_m - \theta_0)}{t_m}\right]t_m, \qquad \text{(A.55)}\end{aligned}$$

where

$$\lim_{\theta_m \to \theta_0} \frac{\|r_\gamma(\theta_m - \theta_0)\|}{\|\theta_m - \theta_0\|} = 0.$$

Since on the right hand side (RHS) of (A.55) within the square brackets the first term is nonzero and dominates the second term, it then follows that

$$\gamma(\theta_m) \neq \gamma(\theta_0).$$

That is to say

$$[\beta(\theta_m) \neq \beta(\theta_0)] \Rightarrow [\gamma(\theta_m) \neq \gamma(\theta_0)].$$

Since the direction h is chosen arbitrarily, it follows that for any θ in the neighborhood of θ_0 we have

$$[\beta(\theta) \neq \beta(\theta_0)] \Rightarrow [\gamma(\theta) \neq \gamma(\theta_0)].$$

Therefore $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ at θ_0. Next we establish necessity, i.e., we want to show that if $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ at θ_0, then $\ker(J_\gamma(\theta_0)) \subseteq \ker(J_\beta(\theta_0))$. If we set $\theta_1 = \theta$ and $\theta_2 = \theta_0$, the proof will follow that of Theorem **4.4.5**. This completes the proof. □

PROOF OF COROLLARY **4.4.4** We first prove sufficiency. Let's distinguish between two situations:

$$(a)\ J_\gamma(\theta_0) = 0;$$
$$(b)\ J_\gamma(\theta_0) \neq 0.$$

If $J_\gamma(\theta_0) = 0$, we have $\ker(J_\gamma(\theta_0)) = \mathbb{R}^k$ so that

$$\left[\mathrm{Im}Q' \subseteq \mathrm{Im}(J_\gamma(\theta_0)')\right] \iff \left[\ker(J_\gamma(\theta_0)) \subseteq \ker(Q)\right] \Rightarrow \left[\ker(Q) = \mathbb{R}^k\right] \Rightarrow [Q = 0] \Rightarrow [Q\theta = 0],\ \forall \theta.$$

Then the statement $Q\theta \neq Q\theta_0$ is false and we have the the following implication

$$[Q\theta \neq Q\theta_0] \Rightarrow [\gamma(\theta) \neq \gamma(\theta_0)],\ \forall \theta$$

which is equivalent to

$$[\gamma(\theta) = \gamma(\theta_0)] \Rightarrow [Q\theta = Q\theta_0 = 0],\ \forall \theta \in \mathscr{V}(\theta_0).$$

Hence, $Q\theta$ is locally identifiable at θ_0 in terms of $\gamma(\theta)$. Now let $J_\gamma(\theta_0) \neq 0$. We wish to show that

$$[Q\theta \neq Q\theta_0] \Rightarrow [\gamma(\theta) \neq \gamma(\theta_0)],\ \forall \theta \in \mathscr{V}(\theta_0)$$

where $\mathscr{V}(\theta_0)$ is an open neighborhood of θ_0. Since $\theta \neq \theta_0$, we can define an arbitrary normalized direction from θ_0 to θ as

$$h = \frac{\theta - \theta_0}{\|\theta - \theta_0\|}$$

and an infinite sequence $\{\theta_m\}_{m=1}^{\infty}$ along the direction h as

$$\theta_m = \theta_0 + t_m h, \quad t_m > 0 \text{ and } t_m \to 0 \text{ as } m \to \infty.$$

Since $\ker(J_\gamma(\theta_0)) \subseteq \ker(Q)$, we have

$$\begin{aligned}
[Q\theta_m \neq Q\theta_0] &\iff [Q(\theta_m - \theta_0) \neq 0] \\
&\iff [(\theta_m - \theta_0) \notin \ker(Q)] \\
&\Rightarrow \left[(\theta_m - \theta_0) \notin \ker(J_\gamma(\theta_0))\right] \\
&\iff \left[J_\gamma(\theta_0)(\theta_m - \theta_0) \neq 0\right]
\end{aligned}$$

$$\Rightarrow \left[J_\gamma(\theta_0) \frac{\theta_m - \theta_0}{\|\theta_m - \theta_0\|} \neq 0 \right]$$
$$\Rightarrow \left[J_\gamma(\theta_0) h \neq 0 \right].$$

The rest of the sufficiency proof will be the same as that of Theorem **4.4.3**. Therefore $Q\theta$ is locally identifiable in terms of $\gamma(\theta)$ at θ_0. The proof of necessity follows that of Theorem **4.4.5**. This completes the proof. □

PROOF OF THEOREM **4.4.5** We first establish sufficiency. Suppose

$$\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big)$$

which is equivalent to

$$\ker(J_\gamma(\theta_0)) \subseteq \ker(J_\beta(\theta_0)).$$

Let's denote $\mathbf{P}_\gamma$ as a projection in $\mathbb{R}^G$ with image space $\mathrm{Im}(\mathbf{P}_\gamma) = \mathrm{Im}(J_\gamma(\theta_0))$. Similarly, denote $\mathbf{P}_\beta$ as a projection in $\mathbb{R}^H$ with image space $\mathrm{Im}(\mathbf{P}_\beta) = \mathrm{Im}(J_\beta(\theta_0))$. Also denote the kernel spaces of $\mathbf{P}_\gamma$ and $\mathbf{P}_\beta$ as $\ker(\mathbf{P}_\gamma)$ and $\ker(\mathbf{P}_\beta)$ respectively. Since both $\gamma(\theta)$ and $\beta(\theta)$ are continuously differentiable on an open neighborhood $\mathscr{V}(\theta_0)$ of θ_0 and θ_0 is a regular point of both $J_\gamma(\theta)$ and $J_\beta(\theta)$, it follows from the rank theorem [see Rudin (1976, Theorem 9.32)] that there exist an open neighborhood $\mathscr{E} \subset \mathbb{R}^k$ of η_0 and an open neighborhood $\mathscr{F} \subset \mathbb{R}^k$ of θ_0 such that there exists a local diffeomorphism $g : \mathscr{E} \mapsto \mathscr{F}$

$$\theta = g(\eta),\ \theta_0 = g(\eta_0)$$

and

$$\gamma(\theta) = \gamma(g(g^{-1}(\theta))) = \gamma(g(\eta)) = J_\gamma(\theta_0)\eta + \phi_\gamma(J_\gamma(\theta_0)\eta),\ \forall \eta \in \mathscr{E}$$
$$\beta(\theta) = \beta(g(g^{-1}(\theta))) = \beta(g(\eta)) = J_\beta(\theta_0)\eta + \phi_\beta(J_\beta(\theta_0)\eta),\ \forall \eta \in \mathscr{E}$$

where $\phi_\gamma : J_\gamma(\theta_0)(\mathscr{E}) \mapsto \ker(\mathbf{P}_\gamma)$ and $\phi_\beta : J_\beta(\theta_0)(\mathscr{E}) \mapsto \ker(\mathbf{P}_\beta)$ are both continuously differentiable. Let θ_1 and θ_2 by any two distinct points in $\mathscr{F}$. Then

$$\begin{aligned}
[\beta(\theta_1) \neq \beta(\theta_2)] &\Rightarrow \left[J_\beta(\theta_0)\eta_1 \neq J_\beta(\theta_0)\eta_2 \right] \\
&\iff \left[J_\beta(\theta_0)(\eta_1 - \eta_2) \neq 0 \right] \\
&\iff \left[(\eta_1 - \eta_2) \notin \ker(J_\beta(\theta_0)) \right] \\
&\Rightarrow \left[(\eta_1 - \eta_2) \notin \ker(J_\gamma(\theta_0)) \right] \\
&\iff \left[J_\gamma(\theta_0)\eta_1 \neq J_\gamma(\theta_0)\eta_2 \right]
\end{aligned}$$

$$\Rightarrow [\gamma(\theta_1) \neq \gamma(\theta_2)],$$

where the last implication holds due to the fact that

$$\mathbf{P}_\gamma \gamma(\theta) = J_\gamma(\theta_0)\eta, \forall \gamma(\theta) \in \gamma(\mathscr{F})$$

and $\mathbf{P}_\gamma$ is a one-to-one mapping restricted to $\gamma(\mathscr{F})$. Since $\gamma(\theta)$ is locally identifiable around θ_0, it follows from Definition **4.3.3** that $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ around θ_0.

We next show that necessity holds. Let's denote an arbitrary open neighborhood of θ_0 as $\mathscr{V}(\theta_0)$ and consider two situations:

$$\begin{aligned} &(a) \quad J_\gamma(\theta_0) = 0; \\ &(b) \quad J_\gamma(\theta_0) \neq 0. \end{aligned}$$

First assume $J_\gamma(\theta_0) = 0$. Since $\gamma(\theta)$ is differentiable on $\mathscr{V}(\theta_0)$ and θ_0 is a regular point of $J_\gamma(\theta)$, it follows from Rudin (1976, Theorem 9.19) that $\gamma(\theta)$ is constant in $\mathscr{V}(\theta_0)$. On the other hand, since $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ around θ_0, we can rewrite

$$\beta(\theta) = \bar{\beta}(\gamma(\theta)), \ \forall \theta \in \mathscr{V}(\theta_0).$$

Then it follows that $\beta(\theta)$ is also constant in $\mathscr{V}(\theta_0)$. Since θ_0 is a regular point of $J_\beta(\theta)$, we have $J_\beta(\theta) = 0$, $\forall \theta \in \mathscr{V}(\theta_0)$. Hence the only element in $\text{Im}\left(J_\beta(\theta_0)'\right)$ and $\text{Im}\left(J_\gamma(\theta_0)'\right)$ is the k-dimension zero vector and we have

$$\text{Im}\big(J_\beta(\theta_0)'\big) = \text{Im}\big(J_\gamma(\theta_0)'\big),$$

which implies that

$$\text{Im}\big(J_\beta(\theta_0)'\big) \subseteq \text{Im}\big(J_\gamma(\theta_0)'\big).$$

Now let $J_\gamma(\theta_0) \neq 0$. By assumption there exists a transformation $\bar{\beta} : \mathbb{R}^G \mapsto \mathbb{R}^H$ such that

$$\beta(\theta) = \bar{\beta}(\gamma(\theta)), \quad \forall \theta \in \mathscr{V}(\theta_0). \tag{A.56}$$

We now show that the transformation $\bar{\beta}$ is differentiable in γ. Denote $\mathscr{V}(\theta_0)$ as an open connected neighborhood of θ_0. Let's arbitrarily choose two distinct points $\theta_1, \theta_2 \in \mathscr{V}(\theta_0)$

such that the normalized direction

$$h_\gamma = \frac{\gamma(\theta_1) - \gamma(\theta_2)}{\|\gamma(\theta_1) - \gamma(\theta_2)\|}$$

is nonzero, which can be expressed in the matrix form

$$\begin{bmatrix} h_{\gamma_1} \\ h_{\gamma_2} \\ \vdots \\ h_{\gamma_G} \end{bmatrix} = \left(\begin{bmatrix} \gamma_1(\theta_1) \\ \gamma_2(\theta_1) \\ \vdots \\ \gamma_G(\theta_1) \end{bmatrix} - \begin{bmatrix} \gamma_1(\theta_2) \\ \gamma_2(\theta_2) \\ \vdots \\ \gamma_G(\theta_2) \end{bmatrix} \right) \frac{1}{\|\gamma(\theta_1) - \gamma(\theta_2)\|}.$$

Obviously, $0 \leq \|h_{\gamma_i}\| \leq 1$, $i = 1, 2, \ldots, G$. Since θ_0 is a regular point of $J_\gamma(\theta)$, we have

$$\left[J_\gamma(\theta_0) \neq 0\right] \Rightarrow \left[J_\gamma(\theta_2) \neq 0\right],$$

which means that at least one row or column of $J_\gamma(\theta_2)$ is nonzero. Without loss of generality, let's assume the first row of $J_\gamma(\theta_2)$ is nonzero, i.e., $J_{\gamma_1}(\theta_2) \neq 0$. Since $\gamma(\theta)$ is continuous in $\mathscr{V}(\theta_0)$, for any convergent sequence $\{\theta_m\}_{m=1}^{\infty}$ that approaches θ_2 as $m \to \infty$, there always exists a convergent sequence $\{\gamma(\theta_m)\}_{m=1}^{\infty}$ in $\mathbb{R}^G$ such that

$$\gamma(\theta_m) \to \gamma(\theta_2) \text{ as } m \to \infty$$

and

$$\gamma(\theta_m) = \gamma(\theta_2) + t_m h_\gamma,$$

where $t_m > 0$ and $t_m \to 0$ as $m \to \infty$. That is to say, for any point $\gamma(\theta_m)$ along the direction of h_γ, there exists a point θ_m in the neighborhood of θ_0 such that the direction between $\gamma(\theta_m)$ and $\gamma(\theta_2)$ equals $t_m h_\gamma$. It follows from the mean value theorem for the real-valued function that

$$\gamma_i(\theta_m) - \gamma_i(\theta_2) = J_{\gamma_i}(\hat{\theta}_i)(\theta_m - \theta_2) = t_m h_{\gamma_i},\ i = 1, 2, \ldots, G,$$

where

$$\hat{\theta}_i = a_i \theta_m + (1 - a_i)\theta_2.$$

Then

$$h_{\gamma_1} = \lim_{m\to\infty} \frac{\gamma_1(\theta_m) - \gamma_1(\theta_2)}{t_m} = \lim_{m\to\infty} J_{\gamma_1}(\hat{\theta}_1) \lim_{m\to\infty} \frac{\theta_m - \theta_2}{t_m} = J_{\gamma_1}(\theta_2) \lim_{m\to\infty} \frac{\theta_m - \theta_2}{t_m}.$$

Since $J_{\gamma_1}(\theta_2) \neq 0$ and h_{γ_1} is bounded, $\lim_{m\to\infty} \frac{\theta_m - \theta_0}{t_m}$ exists. Likewise, the mean value theorem for the real-valued function leads to

$$\bar{\beta}_j(\gamma(\theta_2) + t_m h_\gamma) - \bar{\beta}_j(\gamma(\theta_2)) = \beta_j(\theta_m) - \beta_j(\theta_2) = J_{\beta_j}(\theta_j{}^*)(\theta_m - \theta_2),\ j = 1,2,\ldots,H,$$

where

$$\theta_j{}^* = b_j\theta_m + (1-b_j)\theta_2.$$

It follows that

$$\lim_{m\to\infty} \frac{\bar{\beta}_j(\gamma(\theta_2) + t_m h_\gamma) - \bar{\beta}_j(\gamma(\theta_2))}{t_m} = J_{\beta_j}(\theta_2) \lim_{m\to\infty} \frac{\theta_m - \theta_2}{t_m},\ j = 1,2,\ldots,H. \quad \text{(A.57)}$$

Since $\beta(\theta)$ is continuously differentiable on $\mathscr{V}(\theta_0)$, $J_\beta(\theta_2)$ exists. Thus the limit on the left hand side (LHS) of (A.57) exists for any nonzero h_γ. Since the direction h_γ is arbitrarily chosen, $\bar{\beta}$ is differentiable with respect to any component of vector γ. Hence it follows from the chain rule that

$$J_\beta(\theta) = J_{\bar{\beta}}(\gamma) J_\gamma(\theta),\ \forall \theta \in \mathscr{V}(\theta_0)$$

which implies that

$$\ker(J_\gamma(\theta_0)) \subseteq \ker(J_\beta(\theta_0)).$$

Hence

$$\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big).$$

This completes the proof. □

PROOF OF PROPOSITION **4.4.7** Since $\beta(\theta)$ is locally identifiable in terms of $\gamma(\theta)$ around θ_0, there exists a mapping $\bar{\beta}$ such that

$$\beta(\theta) = \bar{\beta}(\gamma(\theta)),\ \forall \theta \in \mathscr{V}(\theta_0).$$

As in the necessity proof of Theorem **4.4.5**, we consider two situations:

$$(a) \quad J_\gamma(\theta_0) = 0;$$
$$(b) \quad J_\gamma(\theta_0) \neq 0.$$

If $J_\gamma(\theta_0) = 0$, $\gamma(\theta)$ is constant in $\mathscr{V}(\theta_0)$ since $\gamma(\theta)$ is differentiable on $\mathscr{V}(\theta_0)$ and θ_0 is a regular point of $J_\gamma(\theta)$. It follows that $\bar{\beta}$ is also a constant and so is $\beta(\theta)$. Thus $\bar{\beta}$ is differentiable in γ. If fact it is C^∞. If $J_\gamma(\theta_0) \neq 0$, then we can apply the mean value theorem

as in the proof of Theorem **4.4.5** to show that $\bar{\beta}$ is differentiable in γ. This completes the proof. □

PROOF OF COROLLARY **4.4.9** The necessity is straightforward. By assumption the R restrictions $\xi(\theta)$ are locally identifiable at θ_0. Since

$$\mathrm{Im}\big(J_\gamma(\theta)'\big) \subseteq \mathrm{Im}\left[\begin{array}{cc} J_\gamma(\theta)' & J_\xi(\theta)' \end{array}\right], \quad \forall \theta \in \mathscr{V}(\theta_0)$$

and θ_0 is the regular point, Theorem **4.4.3** leads to

$$\mathrm{Im}\big(J_\beta(\theta_0)'\big) \subseteq \mathrm{Im}\big(J_\gamma(\theta_0)'\big) \subseteq \mathrm{Im}\left[\begin{array}{cc} J_\gamma(\theta_0)' & J_\xi(\theta_0)' \end{array}\right].$$

To prove that sufficiency holds, we notice that both $\gamma(\theta)$ and $\xi(\theta)$ are locally identifiable at θ_0 and the same arguments for the sufficiency proof of Theorem **4.4.3** will follow. This completes the proof. □

PROOF OF COROLLARY **4.4.10** Since the R restrictions $\xi(\theta)$ are constant and thus locally identifiable around θ_0, given that θ_0 is a regular point, the proof will be similar to that of Corollary **4.4.3** and is a direct consequence of Theorem **4.4.5**. This completes the proof. □

PROOF OF PROPOSITION **4.4.11** This corollary is a direct consequence of Theorem **4.4.3** and Theorem **4.4.5** by treating $J(\theta_0)\theta$ as $\beta(\theta)$. If we take partial derivative of $J(\theta_0)\theta$ with respect to θ and evaluated at θ_0, we obtain the identity

$$\mathrm{Im}\left(J(\theta_0)'\right) = \mathrm{Im}\left(J(\theta_0)'\right).$$

This completes the proof. □

PROOF OF PROPOSITION **4.5.1** From the definition of the Kullback-Leibler divergence, it is straightforward that

$$(\mathsf{P}_{\theta_1} = \mathsf{P}_{\theta_2}) \Rightarrow (D_{KL}(\theta_1|\theta_0) = D_{KL}(\theta_2|\theta_0)), \quad \forall \theta_1,\ \theta_2 \in \Theta.$$

This completes the proof. □

PROOF OF PROPOSITION **4.5.2** We want to show that

$$\left[\mathsf{P}_\theta = \mathsf{P}_{\theta_0}\right] \Rightarrow \left[\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta} = \frac{\partial D_{KL}(\theta_0|\theta_0)}{\partial \theta_0}\right], \quad \forall\, \theta \in \Theta.$$

Let $\theta \in \Theta$ and

$$\mathsf{P}_\theta = \mathsf{P}_{\theta_0}.$$

Then

$$f(y;\theta) = f(y;\theta_0) \ \ a.e.$$

Let's define

$$S_{KL}(\theta|\theta_0) \equiv \frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}.$$

Thus

$$\begin{aligned}
S_{KL}(\theta|\theta_0) &= -\int_{\mathscr{Y}} \frac{\partial \log f(y;\theta)}{\partial \theta} f(y;\theta_0)\mu(\mathrm{dy}) \\
&= -\int_{\mathscr{Y}} \frac{1}{f(y;\theta)} \frac{\partial f(y;\theta)}{\partial \theta} f(y;\theta_0)\mu(\mathrm{dy}) \\
&= -\int_{\mathscr{Y}} \frac{\partial f(y;\theta)}{\partial \theta} \mu(\mathrm{dy}) \\
&= -\frac{\partial}{\partial \theta} \int_{\mathscr{Y}} f(y;\theta)\mu(\mathrm{dy}) \\
&= 0.
\end{aligned}$$

Meanwhile

$$S_{KL}(\theta_0|\theta_0) = \left(-\int_{\mathscr{Y}} \frac{\partial \log f(y;\theta)}{\partial \theta} f(y;\theta_0)\mu(\mathrm{dy})\right)_{\theta=\theta_0} = 0. \tag{A.58}$$

Therefore

$$\left(\mathsf{P}_\theta = \mathsf{P}_{\theta_0}\right) \Rightarrow \left(S_{KL}(\theta|\theta_0) = S_{KL}(\theta_0|\theta_0) = 0\right), \quad \forall\, \theta \in \Theta.$$

This completes the proof. □

PROOF OF THEOREM **4.5.3** From Proposition **4.5.2**, $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ is locally identifiable at θ_0. If the Hessian matrix of $D_{KL}(\theta|\theta_0)$ evaluated at θ_0 has full column rank, applying Theorem **4.2.15** leads to the desired result. On the other hand, given that θ_0 is a regular point of the Hessian matrix of $D_{KL}(\theta|\theta_0)$ and $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ is locally identifiable at θ_0, we conclude that θ is locally identifiable at θ_0 if and only if $\mathrm{rank}[H_{KL}(\theta_0|\theta_0)] = k$ from Theorem **4.2.15**. This completes the proof. □

PROOF OF COROLLARY **4.5.5** Substituting $\frac{\partial D_{KL}(\theta|\theta_0)}{\partial \theta}$ for $\gamma(\theta)$ in Theorem **4.4.3** produces the result. This completes the proof. □

PROOF OF PROPOSITION **4.6.4** Sufficiency is given by Bekker and Wansbeek (2001, Theorem 9). We show that necessity holds. Partition $D_1(\theta_0)$ as

$$Q_1(\theta_0) = \begin{bmatrix} q_1(\theta_0) & D_{(1)}(\theta_0) \\ 1 & O \end{bmatrix}.$$

It follows from (4.6.22) that θ_1 is locally identifiable if and only if

$$\operatorname{rank}(D_1(\theta_0)) = \operatorname{rank}(D(\theta_0)).$$

Conducting elementary operations on $D_1(\theta_0)$ gives

$$\operatorname{rank}(D_1(\theta_0)) = \operatorname{rank}\begin{bmatrix} O & D_{(1)}(\theta_0) \\ 1 & O \end{bmatrix} = 1 + \operatorname{rank}\left(D_{(1)}(\theta_0)\right),$$

which implies that

$$\operatorname{rank}(D(\theta_0)) > \operatorname{rank}\left(D_{(1)}(\theta_0)\right).$$

Since θ_1 is chosen arbitrarily, if θ_i is locally identifiable, we have

$$\operatorname{rank}(D(\theta_0)) > \operatorname{rank}\left(D_{(i)}(\theta_0)\right).$$

This completes the proof. □

Corollary A.2 *The following identity holds*

$$\frac{\partial^2 D_{KL}(\theta_0|\theta_0)}{\partial \theta_0 \partial {\theta_0}'} = -\mathbb{E}_{\theta_0}\frac{\partial^2 \log f(y;\theta_0)}{\partial \theta_0 \partial {\theta_0}'} = \mathbb{E}_{\theta_0}\left(\frac{\partial \log f(y;\theta_0)}{\partial \theta_0}\frac{\partial \log f(y;\theta_0)}{\partial \theta_0'}\right).$$

PROOF. From the definition of the the Kullback-Leibler divergence, it follows

$$\frac{\partial^2 D_{KL}(\theta|\theta_0)}{\partial \theta \partial \theta'} = -\int_{\mathscr{Y}} \frac{\partial^2 \log f(y;\theta)}{\partial \theta \partial \theta'} f(y;\theta_0)\mu(\mathrm{d}y).$$

Replacing θ with θ_0, we have

$$\frac{\partial^2 D_{KL}(\theta_0|\theta_0)}{\partial \theta_0 \partial {\theta_0}'} = -\int_{\mathscr{Y}} \frac{\partial^2 \log f(y;\theta_0)}{\partial \theta_0 \partial {\theta_0}'} f(y;\theta_0)\mu(\mathrm{d}y)$$

$$= -\mathbb{E}_{\theta_0} \frac{\partial^2 \log f(y;\theta_0)}{\partial \theta_0 \partial {\theta_0}'}.$$

On the other hand,

$$\frac{\partial^2 D_{KL}(\theta|\theta_0)}{\partial \theta \partial \theta'} = \int_{\mathscr{Y}} \frac{1}{f^2(y;\theta)} \frac{\partial f(y;\theta)}{\partial \theta} \frac{\partial f(y;\theta)}{\partial \theta'} f(y;\theta_0)\mu(\mathrm{d}y) - \int_{\mathscr{Y}} \frac{1}{f(y;\theta)} \frac{\partial^2 f(y;\theta)}{\partial \theta \partial \theta'} f(y;\theta_0)\mu(\mathrm{d}y).$$

Again, substituting θ_0 for θ leads to

$$\begin{aligned}
\frac{\partial^2 D_{KL}(\theta|\theta_0)}{\partial \theta \partial \theta'} &= \int_{\mathscr{Y}} \frac{1}{f^2(y;\theta_0)} \frac{\partial f(y;\theta_0)}{\partial \theta_0} \frac{\partial f(y;\theta_0)}{\partial {\theta_0}'} f(y;\theta_0)\mu(\mathrm{d}y) - \int_{\mathscr{Y}} \frac{\partial^2 f(y;\theta_0)}{\partial \theta_0 \partial {\theta_0}'} \mu(\mathrm{d}y) \\
&= \int_{\mathscr{Y}} \frac{\partial \log f(y;\theta_0)}{\partial \theta_0} \frac{\partial \log f(y;\theta_0)}{\partial \theta_0'} f(y;\theta_0)\mu(\mathrm{d}y) - \frac{\partial^2}{\partial \theta_0 \partial \theta_0'} \int_{\mathscr{Y}} f(y;\theta_0)\mu(\mathrm{d}y) \\
&= \mathbb{E}_{\theta_0} \left(\frac{\partial \log f(y;\theta_0)}{\partial \theta_0} \frac{\partial \log f(y;\theta_0)}{\partial \theta_0'} \right).
\end{aligned}$$

This completes the proof. □

Bibliography

Abadir, K. M. and Magnus, J. R. (2005), *Matrix Algebra*, Cambridge University Press, Cambridge, U.K.

Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, John Wiley & Sons, New York.

Aitkin, M. and Clayton, D. (1980), 'The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using glim', *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **29**(2), 156–163.

Alalouf, I. S. and Styan, G. P. H. (1979*a*), 'Characterizations of estimability in the general linear model', *The Annals of Statistics* **7**(1), 194–200.

Alalouf, I. S. and Styan, G. P. H. (1979*b*), 'Estimability and testability in restricted linear models', *Statistics (Mathematische Operationsforschung und Statistik)* **10**(2), 189 – 201.

An, S. and Schorfheide, F. (2007), 'Bayesian analysis of DSGE models', *Econometric Reviews* **26**, 113–172.

Arens, R. (1946), 'Topologies for homeomorphism groups', *American Journal of Mathematics* **68**, 593–610.

Baksalary, J. K. and Kala, R. (1976), 'Extensions of Milliken's estimability criterion', *The Annals of Statistics* **4**(3), 639–641.

Barankin, E. W. (1960), 'Sufficient parameters: Solution of the minimal dimensionality problem', *Annals of the Institute of Statistical Mathematics* **12**(2, June), 91–118.

Barro, R. J. (1977), 'Unanticipated money growth and unemployment in the united states', *American Economic Review* **67**, 101–115.

Bekker, P., Merckens, A. and Wansbeek, T. (1994), *Identification, Equivalent Models and Computer Algebra*, Orlando: Academic Press, San Diego, California.

Bekker, P. and Wansbeek, T. (2001), Identification in parametric models, *in* B. Baltagi, ed., 'Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, U.K., chapter 7, pp. 144–161.

Birkhoff, G. (1934), 'The topology of transformation-sets', *Annals of Mathematics* **35**(4), 861–875.

Bose, R. C. (1944), The fundamental problem of linear estimation (abstract), *in* 'Proceedings of the Indian Science Congress', Vol. 4 (Part III), pp. 2–3.

Bowden, R. (1973), 'The theory of parametric identification', *Econometrica* **41**(6), 1069–1074.

Bunke, H. and Bunke, O. (1974), 'Identifiability and estimability', *Statistics (Mathematische Operationsforschung und Statistik)* **5**(3), 223–233.

Calvo, G. A. (1983), 'Staggered prices in a utility-maximizing framework', *Journal of Monetary Economics* **12**, 383–398.

Canova, F. (2009), 'Comment', *Journal of Business & Economic Statistics* **27**(3), 311–315.

Cantrell, R. S., Burrows, P. M. and Vuong, Q. H. (1991), 'Interpretation and use of generalized Chow tests', *International Economic Review* **31**(3), 725–741.

Carlstrom, C. T., Fuerst, T. S. and Paustian, M. (2009), 'Monetary policy shocks, Choleski identification, and DNK models', *Journal of Monetary Economics* **56**, 1014–1021.

Casella, G. and Berger, R. L. (2002), *Statistical Inference*, Duxbury & Thomson Learning, Pacific Grove, California.

Chen, X., Chernozhukov, V., Lee, S. and Newey, W. (2011), Local identification of nonparametric and semiparametric models, Technical Report 1795, Cowles Foundation for Research in Economics, Yale University, New Haven, Connecticut.

Chernozhukov, V., Imbens, G. W. and Newey, W. K. (2007), 'Instrumental variable estimation of nonseparable models', *Journal of Econometrics* **139**, 4–14.

Chipman, J. S. (1964), 'On least squares with insufficient observations', *Journal of the American Statistical Association* **59**(308), 1078–1111.

Cho, J. S. and White, H. (2012), Directionally differentiable econometric models, Technical report, School of Economics, Yonsei University, Seoul, Korea.

Christ, Carl, F. (1985), 'Early progress in estimating quantitative economic relationshiips in america', *The American Economic Review* **75**(6), 39–52. Centennial Essays and 1985 Survey of Members (December ,1985).

Christensen, R. (1997), *Log-Linear Models and Logistic Regression*, second edn, Springer, New York.

Christensen, R. (2011), *Plane Answers to Complex Questions. The Theory of Linear Models*, fourth edn, Springer, New York.

Church, P. T. (1962), 'Differentiable open maps', *Bulletin of the American Mathematical Society* **68**, 468–469.

Church, P. T. (1963), 'Differentiable open maps on manifolds', *Transactions of the American Mathematical Society* **109**(1), 87–100.

Clarke, F. H. (1976), 'On the inverse function theorem', *Pacific Journal of Mathematics* **64**(1), 97–102.

Collett, D. (2003*a*), *Modelling Binary Data*, CRC Press, Boca Raton, Florida.

Collett, D. (2003*b*), *Modelling Survival Data in Medical Research*, second edn, Chapman and Hall/CRC, Boca Raton, Florida.

Coudin, E. and Dufour, J.-M. (2009), 'Finite-sample distribution-free inference in linear median regressions under heteroskedasticity and nonlinear dependence of unknown form', *Econometrics Journal, 10th anniversary special edition* **12**(S1), S19–S49.

Cox, D. R. and Snell, E. J. (1989), *Analysis of Binary Data*, second edn, Chapman and Hall/CRC, Boca Raton, Florida.

Dasgupta, A., Self, S. G. and Gupta, S. D. (2007), 'Non-identifiable parametric probability models and reparametrization', *Journal of Statistical Planning and Inference* **137**, 3380–3393.

Davidson, R. and MacKinnon, J. G. (2004), *Econometric Theory and Methods*, Oxford University Press, Oxford.

Dees, S., Pesaran, M. H., Smith, L. V. and Smith, R. P. (2009), 'Identification of New Keynesian Phillips curves from a global perspective', *Journal of Money, Credit and Banking* **41**(7), 1481–1502.

Diggle, P. J., Heagerty, P. J., Liang, K.-Y. and Zeger, S. L. (2002), *Analysis of Longitudinal Data*, second edn, Oxford University Press, Oxford.

Dobson, A. J. and Barnett, A. G. (2008), *An Introduction to Generalized Linear Models*, third edn, Chapman & Hall/CRC, Boca Raton, Florida.

Dufour, J.-M. (1982), 'Generalized Chow tests for structural change: A coordinate-free approach', *International Economic Review* **23**, 565–575.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**, 767–808.

Dufour, J.-M. and Hsiao, C. (2008), Identification, *in* L. E. Blume and S. N. Durlauf, eds, 'The New Palgrave Dictionary of Economics', 2nd edn, Palgrave Macmillan, Basingstoke, Hampshire, England.

Dufour, J.-M., Khalaf, L. and Kichian, M. (2006), 'Inflation dynamics and the New Keynesian Phillips curve: An identification robust econometric analysis', *Journal of Economic Dynamics & Control* **30**, 1707–1727.

Dufour, J.-M. and Liang, X. (2012), Parameter separability, canonical representations, estimability and global partial identification of linear parameters, Technical report, Department of Economics, McGill University.

Dufour, J.-M. and Liang, X. (2013*a*), Necessary and sufficient conditions for identification of parametric functions in iv regressions, Technical report, Department of Economics, McGill University.

Dufour, J.-M. and Liang, X. (2013*b*), Necessary and sufficient conditions for nonlinear parametric function identification, Technical report, Department of Economics, McGill University.

Eubank, R. L. and Webster, J. T. (1985), 'The singular-value decomposition as a tool for solving estimability problems', *The American Statistician* **39**, 64–66.

Fisher, F. M. (1966), *The Identification Problem in Econometrics*, McGraw-Hill, New York.

Florens, J.-P. (2003), Inverse problems and structural econometrics: The example of instrumental varialbes, *in* M. Dewatripont, L. P. Hansen and S. J. Turnovsky, eds, 'Advances in Economics and Econometrics, Theory and Applications, Eighth world congress', Vol. II, Cambridge University Press, pp. 284–311.

Florens, J.-P., Johanes, J. and Van Bellegem, S. (2012), 'Instrumental regression in partially linear models', *Econometrics Journal* **15**, 304–324.

Florens, J. P., Mouchart, M. and Rolin, J. M. (1985), 'On two definitions of identification', *Statistics* **16**(2), 213–218.

Florens, J.-P., Mouchart, M. and Rolin, J.-M. (1990), *Elements of Bayesian Statistics*, Marcel Dekker, New York.

Fox, K. (1968), Demand and supply: Econometric studies, *in* 'International Encyclopedia of the Social Sciences', Vol. 4, Macmillan, New York, pp. 104–111.

Friedman, B. M. (1977), 'Financial flow varialbes and the short-run determination of long-term interest rates', *Journal of Political Economy* **85**(4), 661–689.

Gale, D. and Nikaidô, H. (1965), 'The jacobian matrix and global univalence of mappings', *Mathematische Annalen* **159**, 81–93.

Galí, J. and Gertler, M. (1999), 'Inflation dynamics: A structural econometric analysis', *Journal of Monetary Economics* **44**(44), 195–222.

Garcia, C. B. and Zangwill, W. I. (1979), 'On univalence and p-matrices', *Linear Algebra and its Applications* **24**, 239–250.

Golub, G. H. and Van Loan, C. F. (2013), *Matrix Computation*, the fourth edition edn, The John Hopkins University Press, Baltimore and London.

Gouriéroux, C. (2000), *Econometrics of Qualitative Dependent Variables*, Themes in Morden Econometrics, Cambridge University Press, Cambridge, U.K. Translated by Paul B. Klassen.

Gouriéroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Volumes One and Two*, Cambridge University Press, Cambridge, U.K. Translated by Quang Vuong.

Handa, J. (2009), *Monetary Economics*, second edn, Routledge, London and New York.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Harville, D. A. (2008), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Horowitz, J. L. (2009), *Semiparametric and Nonparametric Models Methods in Econometrics*, Springer, New York.

Hosmer, D. W. and Lemeshow, S. (2000), *Applied Logistic Regression*, John Wiley & Sons, New York.

Ichimura, H. (1993), 'Semiparametric least squares and weighted sls estimation of single-index models', *Journal of Econometrics* **58**, 71–120.

Iskrev, N. (2010), 'Local identification in DSGE models', *Journal of Monetary Economics* **57**, 189–202.

Jiang, J. (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, Springer, New York.

Kleibergen, F. and Mavroeidis, S. (2009), 'Weak instrument robust tests in gmm and the new keynesian phillips curve', *Journal of Business & Economic Statistics* **27**, 293–311.

Klein, R. W. and Spady, R. H. (1993), 'An efficient semiparametric estimator for binary response models', *Econometrica* **61**(2), 387–421.

Koenker, R. (2005), *Quantile Regression*, Cambridge University Press, Cambridge, U.K.

Koenker, R. and Bassett, Jr., G. (1978), 'Regression quantiles', *Econometrica* **46**(1), 33–50.

Kojima, M. and Saigal, R. (1979), 'A study of PC^1 homeomorphisms on subdivided polyhedrons', *SIAM Journal on Mathematical Analysis* **10**(6), 1299–1312.

Komunjer, I. (2012), 'Global identificaion in nonlinear models with moment restrictions', *Econometric Theory* **28**, 719–729.

Komunjer, I. and Ng, S. (2011), 'Dynamic identification of dynamic stochastic general equilibrium models', *Econometrica* **79**, 1995–2032.

Koopmans, T. C. and Reiersøl, O. (1950), 'The identification of structural characteristics', *The Annals of Mathematical Statistics* **21**(2), 165–181.

Kounias, S. and Chalikias, M. (2008), 'Estimability of parameters in a linear model and related characterizations', *Statistics & Probability Letters* **78**, 2437–2439.

Krantz, S. G. and Parks, H. R. (2002), *The Implicit Function Theorem: History, theory and Applications*, Birkäuser, Boston.

Kruskal, W. (1961), The coordinate-free approach to Gauss-Markov estimation and its application to missing and extra observations, *in* J. Neyman, ed., 'Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability', Vol. 1, The University of California Press, Berkeley, California, pp. 435–451.

Kruskal, W. (1968), 'When are Gauss Markov and least squares estimates identical? a coordinate free approach', *Annals of Mathematical Statistics* **39**, 70–75.

Kydland, F. E. and Prescott, E. C. (1977), 'Rules rather than discretion: The inconsistency of optimal plans', *Journal of Political Economy* **85**, 473–491.

Lehmann, E. L. (2005), *Testing Statistical Hypotheses*, third edn, John Wiley & Sons, New York.

Lenoir, M. (1913), *Études sur la formation et le mouvement des prix*, Giard et Brière, Paris.

Lewbel, A. (2000), 'Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables', *Journal of Econometrics* **97**, 145–177.

Li, Q. and Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton, N.J.

Lieb, E. H. and Loss, M. (2000), *Analysis*, Vol. 14, American Mathematical Society, Providence, Rhode Island.

Lindé, J. (2005), 'Estimating New-Keynesian Phillips Curves: A full information maximum likelihood approach', *Journal of Monetary Economics* **52**, 1135–1149.

Lucas, Jr., R. E. (1976), Econometric policy evaluation: A critique, *in* K. Brunner and A. H. Meltzer, eds, 'The Phillips Curve and Labor Markets', number 1 *in* 'Carnegie-Rochester Conference Series on Public Policy', North-Holland, New York, pp. 19–46.

Ma, A. (2002), 'GMM estimation of the new Phillips curve', *Economics Letters* **76**, 411–417.

Magnus, J. R. and Neudecker, H. (1998), *Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition*, John Wiley & Sons, New York.

Mandel, J. (1982), 'Use of the singular value decomposition in regression analysis', *The American Statistician* **36**(1), 15–24.

Manski, C. F. (1995), *Identification Problems in the Social Sciences*, Harvard University Press, Cambridge, Mass.

Manski, C. F. (2003), *Partial Identification of Probability Distributions*, Springer, Berlin and New York.

Marsaglia, G. and Styan, G. P. H. (1974), 'Rank conditions for generalized inverses of partitioned matrices', *Sankhyā: The Indian Journal of Statistics, Series A* **36**(4), 437–442.

Martins, L. F. and Gabriel, V. J. (2009), 'New Keynesian Phillips curves and potential identification failures: A generalized empirical likelihood analysis', *Journal of Macroeconomics* **31**, 561–571.

Mavroeidis, S. (2005), 'Identification issues in forward-looking models estimated by GMM with an application to the Phillips curve', *Journal of Money, Credit and Banking* **37**(3), 421–448.

Mavroeidis, S., Plagborg-Møller, M. and Stock, J. H. (2014), 'Empirical evidence on inflation expectations in the New Keynesian Phillips Curve', *Journal of Economic Literature* **52**, 124–188.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, second edn, Chapman and Hall/CRC, London and New York.

Meyer, C. D. (2000), *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, PA.

Milliken, G. A. (1971), 'New criteria for estimability for linear models', *The Annals of Mathematical Statistics* **42**(5), 1588–1594.

Munkres, J. R. (2000), *Topology*, second edn, Prentice Hall, Upper Saddle River, NJ.

Nason, J. M. and Smith, G. W. (2008), 'Identifying the New Keynesian Phillips curve', *Journal of Applied Econometrics* **23**, 525–551.

Nelder, J. A. (1985), 'An alternative interpretation of the singular-value decomposition in regression', *The American Statistician* **39**(1), 63–64.

Nelder, J. A. and Wedderburn, R. W. M. (1972), 'Generalized linear models', *Journal of the Royal Statistical Society, Series A* **135**, 370–384.

Newey, W. and Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**, 1565–1578.

Nijenhuis, A. (1974), 'Strong derivatives and inverse mappings', *The American Mathematical Monthly* **81**(9), 969–980.

Nymoen, R., Swensen, A. R. and Tveter, E. (2012), 'Interpreting the evidence for New Keynesian models of inflation dynamics', *Journal of Macroeconomics* **34**, 253–263.

Oulhaj, A. and Mouchart, M. (2003), 'Partial sufficiency with connection to the identification problem', *International Journal of Statistics* **LXI**(2), 267–283.

Parthasarathy, T. (1983), *On Global Univalence Theorems*, Springer-Verlag, Berlin.

Paulino, C. D. M. and de Bragança Pereira, C. A. (1994), 'On identifiability of parametric statistical models', *Journal of the Italian Statistical Society* **1**, 125–151.

Phillips, P. C. B. (1989), 'Partially identified econometric models', *Econometric Theory* **5**(2), 181–240.

Plastock, R. (1974), 'Homeomorphisms between banach spaces', *Transactions of the American Mathematical Society* **200**, 169–183.

Pourciau, B. H. (1982), 'Hadamard's theorem for locally Lipschitzian maps', *Journal of Mathematical Analysis and Applications* **85**, 279–285.

Ramanathan, R. (2002), *Introductory Econometrics with Applications*, Roche, Michael P., Mason, Ohio.

Rao, C. R. (1945), 'Generalisation of Markoff's theorem and tests of linear hypotheses', *Sankhyā: The Indian Journal of Statistics* **7**(1), 9–16.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, second edn, John Wiley & Sons, New York.

Rao, C. R. and Mitra, S. K. (1971), *Generalized Inverse of Matrices and its Applications*, John Wiley & Sons, New York.

Reiersøl, O. (1963), 'Identifiability, estimability, pheno-restricting specifications, and zero Lagrange multipliers', *Skandinavisk Aktuarietidskrift* **46**, 131–142.

Richmond, J. (1974), 'Identifiability in linear models', *Econometrica* **42**, 731–736.

Robinson, P. M. (1988), 'Root-n-consistent semiparametric regression', *Econometrica* **56**(4), 931–954.

Romer, D. (1993), 'Openness and inflation: Theory and evidence', *Quarterly Journal of Economics* **CVII**, 869–903.

Rothenberg, T. J. (1971), 'Identification in parametric models', *Econometrica* **39**, 577–591.

Rudd, J. and Whelan, K. (2005), 'New tests of the New Keynesian Phillips Curve', *Journal of Monetary Economics* **52**, 1167–1181.

Rudd, J. and Whelan, K. (2006), 'Modeling inflation dynamics: A critical review of recent research', *Journal of Money, Credit and Banking* **39**(1), 155–170.

Rudin, W. (1976), *Principles of Mathematical Analysis, Third Edition*, McGraw-Hill, New York.

Sargent, T. J. and Wallace, N. (1975), 'Rational expectations, the optimal monetary instrument, and the optimal money supply rule', *Journal of Political Economy* **83**, 241–254.

Scheffé, H. (1959), *The Analysis of Variance*, John Wiley & Sons, New York.

Searle, S. R. (1971), *Linear Models*, John Wiley & Sons, New York.

Seely, J. (1970), 'Linear spaces and unbiased estimation', *The Annals of Mathematical Statistics* **41**(5), 1725–1734.

Seely, J. (1977), 'Estimability and linear hypotheses', *The American Statistician* **31**(3), 121–123.

Seely, J. and Birkes, D. (1980), 'Estimability in partitioned linear models', *The Annals of Statistics* **8**(2), 399–406.

Shapiro, A. (1986), 'Asymptotic theory of overparameterized structural models', *Journal of the American Statistical Association* **81**, 142–149.

Smets, F. and Wouters, R. (2003), 'An estimated dynamic stochastic general equalibrium model of the Euro area', *Journal of the European Economic Association* **1**(5), 1123–1175.

Stock, J. H. and Trebbi, F. (2003), 'Retrospectives who invented instrumental bariable regression?', *The Journal of Economic Perspectives* **17**(3), 177–194.

Tamer, E. (2010), 'Partial identification in econometrics', *The Annual Review of Economics* **2**, 167–195.

Tsoukis, C., Kapetanios, G. and Pearlman, J. (2011), 'Elusive persistence: Wage and price rigidities, the New Keynesian Phillips curve and inflation dynamics', *Journal of Economic Surveys* **25**(4), 737–768.

Tuncer, Y. (1985), 'Equivalence of parametric identifiability and estimability', *Communications in Statistics-Theory and Methods* **14**(2), 89–119.

Tyler, D. E. (1981), 'Asymptotic inference for eigenvectors', *The Annals of Statistics* **9**(4), 725–736.

Whyburn, G. T. (1942), *Analytic Topology*, Vol. XXVIII, American Mathematical Society, Providence, RI.

Wood, S. N. (2006), *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC, Boca Raton, Florida.

Wooldridge, J. M. (2006), *Introductory Econometrics: A Modern Approach*, South-Western, Mason, OH.

Working, E. J. (1927), 'What do statistical demand curves show?', *The Quarterly Journal of Economics* **41**(2), 212–235.

Wright, P. G. (1915), 'Moore's economic cycles', *Quarterly Journal of Economics* **29**(3), 631–641.

Zhang, C., Osborn, D. R. and Kim, D. H. (2007), 'The New Keynesian Phillips curve: From sticky inflation to sticky prices', *Journal of Money, Credit and Banking* **40**(4), 667–699.