Université de Montréal

Exogeneity, Weak Identification and Instrument Selection in Econometrics

par

Firmin Sabro Doko Tchatoka

Département de sciences économiques
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiae Doctor (Ph.D.)
en sciences économiques

Février 2010

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Exogeneity, Weak Identification and Instrument Selection in Econometrics

présentée par

Firmin Sabro Doko Tchatoka

a été évaluée par un jury composé des personnes suivantes:

Benoit Perron (Département de Sciences Économiques, Université de Montréal)
président-rapporteur
Jean-Marie Dufour (Department of Economics, McGill University)
directeur de recherche
Marine Carrasco (Département de Sciences Économiques, Université de Montréal)
membre du jury
Jan Kiviet (Department of ùeconomics, University of Amsterdam)
examinateur externe
François Perron (Département de Mathématiques et de Statistiques, Université de Montréal)
représentant du doyen de la FES

Thèse acceptée le: 12 Février 2010

Sommaire

La dernière décennie a connu un intérêt croissant pour les problèmes posés par les variables instrumentales faibles dans la littérature économétrique, c'est-à-dire les situations où les variables instrumentales sont faiblement corrélées avec la variable à instrumenter. En effet, il est bien connu que lorsque les instruments sont faibles, les distributions des statistiques de Student, de Wald, du ratio de vraisemblance et du multiplicateur de Lagrange ne sont plus standard et dépendent souvent de paramètres de nuisance. Plusieurs études empiriques portant notamment sur les modèles de rendements à l'éducation [Angrist et Krueger (1991, 1995), Angrist et al. (1999), Bound et al. (1995), Dufour et Taamouti (2007)] et d'évaluation des actifs financiers (C-CAPM) [Hansen et Singleton (1982,1983), Stock et Wright (2000)], où les variables instrumentales sont faiblement corrélées avec la variable à instrumenter, ont montré que l'utilisation de ces statistiques conduit souvent à des résultats peu fiables. Un remède à ce problème est l'utilisation de tests robustes à l'identification [Anderson et Rubin (1949), Moreira (2002), Kleibergen (2003), Dufour et Taamouti (2007)]. Cependant, il n'existe aucune littérature économétrique sur la qualité des procédures robustes à l'identification lorsque les instruments disponibles sont endogènes ou à la fois endogènes et faibles. Cela soulève la question de savoir ce qui arrive aux procédures d'inférence robustes à l'identification lorsque certaines variables instrumentales supposées exogènes ne le sont pas effectivement. Plus précisément, qu'arrive-t-il si une variable instrumentale invalide est ajoutée à un ensemble d'instruments valides? Ces procédures se comportent-elles différemment? Et si l'endogénéité des variables instrumentales pose des difficultés majeures à l'inférence statistique, peut-on proposer des procédures de tests qui sélectionnent les instruments lorsqu'ils sont à la fois forts et valides? Est-il possible de proposer les proédures de sélection d'instruments qui demeurent valides même en présence d'identification faible?

Cette thèse se focalise sur les modèles structurels (modèles à équations simultanées) et apporte des réponses à ces questions à travers quatre essais.

Le premier essai est publié dans *Journal of Statistical Planning and Inference 138 (2008)*

2649 – 2661. Dans cet essai, nous analysons les effets de l'endogénéité des instruments sur deux statistiques de test robustes à l'identification: la statistique d'Anderson et Rubin (AR, 1949) et la statistique de Kleibergen (K, 2003), avec ou sans instruments faibles. D'abord, lorsque le paramètre qui contrôle l'endogénéité des instruments est fixe (ne dépend pas de la taille de l'échantillon), nous montrons que toutes ces procédures sont en général convergentes contre la présence d'instruments invalides (c'est-à-dire détectent la présence d'instruments invalides) indépendamment de leur qualité (forts ou faibles). Nous décrivons aussi des cas où cette convergence peut ne pas tenir, mais la distribution asymptotique est modifiée d'une manière qui pourrait conduire à des distorsions de niveau même pour de grands échantillons. Ceci inclut, en particulier, les cas où l'estimateur des double moindres carrés demeure convergent, mais les tests sont asymptotiquement invalides. Ensuite, lorsque les instruments sont localement exogènes (c'est-à-dire le paramètre d'endogénéité converge vers zéro lorsque la taille de l'échantillon augmente), nous montrons que ces tests convergent vers des distributions chi-carré non centrées, que les instruments soient forts ou faibles. Nous caractérisons aussi les situations où le paramètre de non centralité est nul et la distribution asymptotique des statistiques demeure la même que dans le cas des instruments valides (malgré la présence des instruments invalides).

Le deuxième essai étudie l'impact des instruments faibles sur les tests de spécification du type Durbin-Wu-Hausman (DWH) ainsi que le test de Revankar et Hartley (1973). Nous proposons une analyse en petit et grand échantillon de la distribution de ces tests sous l'hypothèse nulle (niveau) et l'alternative (puissance), incluant les cas où l'identification est déficiente ou faible (instruments faibles). Notre analyse en petit échantillon founit plusieurs perspectives ainsi que des extensions des précédentes procédures. En effet, la caractérisation de la distribution de ces statistiques en petit échantillon permet la construction des tests de Monte Carlo exacts pour l'exogénéité même avec les erreurs non Gaussiens. Nous montrons que ces tests sont typiquement robustes aux intruments faibles (le niveau est contrôlé). De plus, nous fournissons une caractérisation de la puissance des tests, qui exhibe clairement les facteurs qui déterminent la puissance. Nous montrons que les tests n'ont pas de puissance lorsque tous les instruments sont faibles [similaire à Guggenberger(2008)].

Cependant, la puissance existe tant qu'au moins un seul instruments est fort. La conclusion de Guggenberger (2008) concerne le cas où tous les instruments sont faibles (un cas d'intérêt mineur en pratique). Notre théorie asymptotique sous les hypothèses affaiblies confirme la théorie en échantillon fini.

Par ailleurs, nous présentons une analyse de Monte Carlo indiquant que: (1) l'estimateur des moindres carrés ordinaires est plus efficace que celui des doubles moindres carrés lorsque les instruments sont faibles et l'endogenéité modérée [conclusion similaire à celle de Kiviet and Niemczyk (2007)]; (2) les estimateurs pré-test basés sur les tests d'exogenété ont une excellente performance par rapport aux doubles moindres carrés. Ceci suggère que la méthode des variables instrumentales ne devrait être appliquée que si l'on a la certitude d'avoir des instruments forts. Donc, les conclusions de Guggenberger (2008) sont mitigées et pourraient être trompeuses.

Nous illustrons nos résultats théoriques à travers des expériences de simulation et deux applications empiriques: la relation entre le taux d'ouverture et la croissance économique et le problème bien connu du rendement à l'éducation.

Le troisième essai étend le test d'exogénéité du type Wald proposé par Dufour (1987) aux cas où les erreurs de la régression ont une distribution non-normale. Nous proposons une nouvelle version du précédent test qui est valide même en présence d'erreurs non-Gaussiens. Contrairement aux procédures de test d'exogénéité usuelles (tests de Durbin-Wu-Hausman et de Rvankar-Hartley), le test de Wald permet de résoudre un problème courant dans les travaux empiriques qui consiste à tester l'exogénéité partielle d'un sous ensemble de variables. Nous proposons deux nouveaux estimateurs pré-test basés sur le test de Wald qui performent mieux (en terme d'erreur quadratique moyenne) que l'estimateur IV usuel lorsque les variables instrumentales sont faibles et l'endogénéité modérée. Nous montrons également que ce test peut servir de procédure de sélection de variables instrumentales. Nous illustrons les résultats théoriques par deux applications empiriques: le modèle bien connu d'équation du salaire [Angist et Krueger (1991, 1999)] et les rendements d'échelle [Nerlove (1963)]. Nos résultats suggèrent que l'éducation de la mère expliquerait le décrochage de son fils, que l'output est une variable endogène dans l'estimation du coût de la firme et que le prix du fuel en est un instrument valide pour l'output.

Le quatrième essai résout deux problèmes très importants dans la littérature économétrique. D'abord, bien que le test de Wald initial ou étendu permette de construire les régions de confiance et de tester les restrictions linéaires sur les covariances, il suppose que les paramètres du modèle sont identifiés. Lorsque l'identification est faible (instruments faiblement corrélés avec la variable à instrumenter), ce test n'est en général plus valide. Cet essai développe une procédure d'inférence robuste à l'identification (instruments faibles) qui permet de construire des régions de confiance pour la matrices de covariances entre les erreurs de la régression et les variables explicatives (possiblement endogènes). Nous fournissons les expressions analytiques des régions de confiance et caractérisons les conditions nécessaires et suffisantes sous lesquelles ils sont bornés. La procédure proposée demeure valide même pour de petits échantillons et elle est aussi asymptotiquement robuste à l'hétéroscédasticité et l'autocorrélation des erreurs.

Ensuite, les résultats sont utilisés pour développer les tests d'exogénéité partielle robustes à l'identification. Les simulations Monte Carlo indiquent que ces tests contrôlent le niveau et ont de la puissance même si les instruments sont faibles. Ceci nous permet de proposer une procédure valide de sélection de variables instrumentales même s'il y a un problème d'identification. La procédure de sélection des instruments est basée sur deux nouveaux estimateurs pré-test qui combinent l'estimateur IV usuel et les estimateurs IV partiels. Nos simulations montrent que: (1) tout comme l'estimateur des moindres carrés ordinaires, les estimateurs IV partiels sont plus efficaces que l'estimateur IV usuel lorsque les instruments sont faibles et l'endogénéité modérée; (2) les estimateurs pré-test ont globalement une excellente performance comparés à l'estimateur IV usuel. Nous illustrons nos résultats théoriques par deux applications empiriques: la relation entre le taux d'ouverture et la croissance économique et le modèle de rendements à l'éducation. Dans la première application, les études antérieures ont conclu que les instruments n'étaient pas trop faibles [Dufour et Taamouti (2007)] alors qu'ils le sont fortement dans la seconde [Bound (1995), Doko et Dufour (2009)]. Conformément à nos résultats théoriques, nous trouvons les régions de confiance non bornées pour la covariance dans le cas où les instruments sont assez faibles.

Mots clés: Modèle structurels, instruments faibles, instruments endogènes, tests robustes à l'identification, tests de Monte Carlo exacts pour l'exogénéité, régions de confiance, sélection d'instruments, exogénéité partielle, estimateurs IV partiels.

Summary

The last decade shows growing interest for the so-called weak instruments problems in the econometric literature, i.e. situations where instruments are poorly correlated with endogenous explanatory variables. More generally, these can be viewed as situations where model parameters are not identified or nearly so (see Dufour and Hsiao, 2008). It is well known that when instruments are weak, the limiting distributions of standard test statistics - like Student, Wald, likelihood ratio and Lagrange multiplier criteria in structural models - have non-standard distributions and often depend heavily on nuisance parameters. Several empirical studies including the estimation of returns to education [Angrist and Krueger (1991, 1995), Angrist et al. (1999), Bound et al. (1995), Dufour and Taamouti (2007)] and asset pricing model (C-CAPM) [Hansen and Singleton (1982, 1983), Stock and Wright (2000)], have showed that the above procedures are unreliable in presence of weak identification. As a result, identification-robust tests [Anderson and Rubin (1949), Moreira (2003), Kleibergen (2002), Dufour and Taamouti (2007)] are often used to make reliable inference. However, little is known about the quality of these procedures when the instruments are invalid or both weak and invalid. This raises the following question: what happens to inference procedures when some instruments are endogenous or both weak and endogenous? In particular, what happens if an invalid instrument is added to a set of valid instruments? How robust are these inference procedures to instrument endogeneity? Do alternative inference procedures behave differently? If instrument endogeneity makes statistical inference unreliable, can we propose the procedures for selecting "good instruments" (i.e. strong and valid instruments)? Can we propose instrument selection procedure which will be valid even in presence of weak identification?

This thesis focuses on structural models and answers these questions through four chapiters. The first chapter is published in *Journal of Statistical Planning and Inference 138 (2008) 2649 – 2661*. In this chapter, we analyze the effects of instrument endogeneity on two identification-robust procedures: Anderson and Rubin (1949, AR) and Kleibergen (2002, K) test statistics, with or without weak instruments. First, when the level of instrument endogeneity is fixed (does not depend on the sample size), we show that all these procedures are in general consistent against the presence of invalid instruments (hence asymptotically invalid for the hypothesis of interest), whether the instruments are "strong" or "weak". We also describe situations where this consistency may not hold, but the asymptotic distribution is modified in a way that would lead to size distortions in large samples. These include, in particular, cases where 2SLS estimator remains consistent, but the tests are asymptotically invalid. Second, when the instruments are locally exogenous (the level of instrument endogeneity approaches zero as the sample size increases), we find asymptotic non-central chi-square distributions with or without weak instruments, and describe situations where the non-centrality parameter is zero and the asymptotic distribution remains the same as in the case of valid instruments (despite the presence of invalid instruments).

The second chapter analyzes the effects of weak identification on Durbin-Wu-Hausman (DWH) specification tests an Revankar-Harttley exogeneity test. We propose a finite-and large-sample analysis of the distribution of DWH tests under the null hypothesis (level) and the alternative hypothesis (power), including when identification is deficient or weak (weak instruments). Our finite-sample analysis provides several new insights and extensions of earlier procedures. The characterization of the finite-sample distribution of the test-statistics allows the construction of exact identification-robust exogeneity tests even with non-Gaussian errors (Monte Carlos tests) and shows that such tests are typically robust to weak instruments (level is controlled).

Furthermore, we provide a characterization of the power of the tests, which clearly exhibits factors which determine power. We show that DWH-tests have no power when all instruments are weak [similar to Guggenberger(2008)]. However, power does exist as soon as we have one strong instruments. The conclusions of Guggenberger (2008) focus on the case where all instruments are weak (a case of little practical interest). Our asymptotic distributional theory under weaker assumptions confirms the finite-sample theory.

Moreover, we present simulation evidence indicating: (1) over a wide range cases, including weak IV and moderate endogeneity, OLS performs better than 2SLS [finding similar to Kiviet and Niemczyk (2007)]; (2) pretest-estimators based on exogeneity tests have an excellent overall performance compared with usual IV estimator.

We illustrate our theoretical results through simulation experiment and two empirical applications: the relation between trade and economic growth and the widely studied problem of returns to education.

In the third chapter, we extend the generalized Wald partial exogeneity test [Dufour (1987)] to non-gaussian errors. Testing whether a subset of explanatory variables is exogenous is an important challenge in econometrics. This problem occurs in many applied works. For example, in the well know wage model, one should like to assess if mother's education is exogenous without imposing additional assumptions on ability and schooling. In the growth model, the exogeneity of the constructed instrument on the basis of geographical characteristics for the trade share is often questioned and needs to be tested without constraining trade share and the other variables. Standard exogeneity tests of the type proposed by Durbin-Wu-Hausman and Revankar-Hartley cannot solve such problems. A potential cure for dealing with partial exogeneity is the use of the generalized linear Wald (GW) method (Dufour, 1987). The GW-procedure however assumes the normality of model errors and it is not clear how robust is this test to non-gaussian errors.

We develop in this chapter, a modified version of earlier procedure which is valid even when model errors are not normally distributed. We present simulation evidence indicating that when identification is strong, the standard GW-test is size distorted in presence of non-gaussian errors. Furthermore, our analysis of the performance of different pretest-estimators based on GW-tests allow us to propose two new pretest-estimators of the structural parameter. The Monte Carlo simulations indicate that these pretest-estimators have a better performance over a wide range cases compared with 2SLS. Therefore, this can be viewed as a procedure for selecting variable where a GW-test is used in the first stage to decide which variables should be instruments and which ones are valid instruments.

We illustrate our theoretical results through two empirical applications: the well known wage equation and the returns to scale in electricity supply. The results show that the GW-tests cannot reject the exogeneity of mother's education, *i.e.* mother's education may constitute a valid IV for schooling. However, the output in cost equation is endogenous and the price of fuel is a valid IV for estimating the returns to scale.

The fourth chapter develops identification-robust inference for the covariances between errors and regressors of an IV regression. The results are then applied to develop partial exogeneity tests and partial IV pretest-estimators which are more efficient than usual IV estimator.

When more than one stochastic explanatory variables are involved in the model, it is often necessary to determine which ones are independent of the disturbances. This problem arises in many empirical applications. For example, in the New Keynesian Phillips Curve, one should like to assess whether the interest rate is exogenous without imposing additional assumptions on inflation rate and the other variables. Standard Wu-Durbin-Hausman (DWH) tests which are commonly used in applied work are inappropriate to deal with such a problem. The generalized Wald (GW) procedure (Dufour, 1987) which typically allows the construction of confidence sets as well as testing linear restrictions on covariances assumes that the available instruments are strong. When the instruments are weak, the GW-test is in general size distorted. As a result, its application in models where instruments are possibly weak–returns to education, trade and economic growth, life cycle labor supply, New Keynesian Phillips Curve, pregnancy and the demand for cigarettes–may be misleading.

To answer this problem, we develop a finite-and large-sample valid procedure for building confidence sets for covariances allowing for the presence of weak instruments. We provide analytic

forms of the confidence sets and characterize necessary and sufficient conditions under which they are bounded.

Moreover, we propose two new pretest-estimators of structural parameters based on our above procedure. Both estimators combine 2SLS and partial IV-estimators. The Monte Carlo experiment shows that: (1) partial IV-estimators outperform 2SLS when the instruments are weak; (2) pretest-estimators have an excellent overall performance–bias and MSE– compared with 2SLS. Therefore, this can be viewed as a variable selection method where the projection-based techniques is used to decide which variables should be instrumented and which ones are valid instruments.

We illustrate our results through two empirical applications: the relation between trade and economic growth and the widely studied problem of returns to education. The results show unbounded confidence sets, suggesting that the IV are relatively poor in these models, as questioned in the literature [Bound (1995)].

Key words: Structural models, weak instruments, endogenous instruments, identification-robust tests, exact Monte Carlo exogeneity tests, confidence sets, instrument selection, partial exogeneity, partial IV estimators.

Contents

List of Tables

Remerciements

Je dois l'accomplissement de ce travail à plusieurs personnes que je voudrais remercier sans avoir la prétention de réussir à être exhaustif.

Tout d'abord mon directeur de recherche Jean-Marie Dufour pour sa disponibilité et tout ce qu'il m'a patiemment et généreusement enseigné. Sans son précieux soutien, cette thèse ne serait pas arrivée à ce stade.

Je voudrais également remercier Benoit Perron, Lynda Khalaf et Emanuela Cardia avec qui j'ai particulièrement collaboré et qui m'ont soutenu tout au long de cette thèse.

Merci aux Professeurs Marine Carrasco, Rui Castro, Sílvia Gonçalves et William McCausland pour leurs conseils et suggestions qui m'ont beaucoup apporté. Mes remerciements vont aussi aux Professeurs Atsushi Inoue, Russuel Davidson, Victoria Zende-Walsh pour leurs commentaires sur le deuxième essai et aux Professeurs Tze Leung Lai et Ingram Olkin pour leurs commentaires très utiles sur le premier essai.

Je voudrais aussi remercier les différents organismes qui m'ont apporté un support financier et logistique tout le long de ma formation: la Banque Laurentienne, la Chaire de Recherche du Canada en Économétrie, la Chaire de Recherche William Dow en Économie Politique (McGill University), le Centre Interuniversitaire de Recherche en Analyse des Organisations (CIRANO), le Centre Interuniversitaire de Recherche en Économie Quantitative (CIREQ), the Mathematics of Information Technology and Complex Systems (MITACS) et le Fonds de Recherch sur la Société et la Culture (FQRSC).

Un Merci tout particulier :

À mon épouse Carole et mes enfants Alexia, Laurie et Juvénal pour le soutien quotidien et inconditionnel dont ils ont fait preuve pour moi.

À ma mère Yoro, mon feu père Doko Tcharo, mes frère et soeurs en particulier Nicolas, Boniface, Sahi, Moussa, Joseph et André pour leurs conseils et aussi leur patience en ces périodes où nous n'avons pu nous voir aussi souvent qu'on aurait souhaité.

À mes oncles et tantes Raphaèl, Odandou, Albert, Nicolas Tchoroué, feu Tchesso Tchati, Kassa et Sékou pour leur régulier soutien.

À Mes beaux parents Iréné et Léontine pour leur soutien quotidien et inconditionnel durant cette thèse.

À mes amis Prosper Dovonon et Olivia, Assa Sena, Constant Lonkeng et Judith, Balguissa Ouédraogo, Rachidi Kotchoni, Messan Agbaglah, Modeste Somé, Octave Keutiben, Sali Sanou, Georges Tsafack, Isabelle Akafou, Mame Astou Diouf, Christine, Johan Latulipe et Julien Edja pour les encouragements mutuels et les discussions enrichissantes pour la réalisation de cette thèse.

À Abdoulaye Diaw, Ernest Tafolong et Pascal Martinolli pour leur soutien technique inconditionnel, et Suzanne Larouche-Sidoti pour les bonbons et les belles histoires pour changer de la routine.

À tout le personnel du département de sciences économiques de l'université de Montréal en particulier Josée Lafontaine et lyne Racine et celui du CIREQ pour leur dynamisme et leur efficacité.

Et enfin à mes collègues Professeurs de du Département d'Économie de l'Université de Sherbrooke pour leur accueil convivial et leurs encouragements.

Introduction générale

Cette thèse est composée de quatre essais et s'inscrit dans le cadre des modèles structurels (modèles à équations simultanées). Elle contribue aux récents développements en économétrie lié aux problème posé par les variables instrumentales faibles, c'est-à-dire les situations où les variables instrumentales sont faiblement corrélées avec la variable à instrumenter.

Le premier essai est publié dans *Journal of Statistical Planning and Inference 138 (2008) 2649 – 2661*. Il aborde des questions relatives aux effets de l'endogénéity des variables instrumentales sur les statistiques de test robustes à l'identification. Le deuxième essai de cette thèse étudie le comportement des tests d'exogénéité du type Durbin-Wu-Hausman (DWH) and Revankar-Hartley (RH) en présence d'identification faible. Le troisième et le quatrième essais abordent des questions relatives aux tests l'exogénéité partielle, à la construction des régions de confiance pour la covariance entre les erreurs et les régresseurs et aux estimateurs pré-test.

Dans le premier essai, nous analysons les effets de l'endogénéité des instruments sur les statistiques de test robustes à l'identification.

Les procédures robustes à l'identification – incluant la méthode d'Anderson et Rubin – suppose que les variables instrumentales sont exogène. Ce qui soulève la question de savoir ce qui arrive à ces proédures lorsque certaines variables instrumentales sont endogènes. En particulier, qu'est-ce qui arrive aux procédures de test robustes à l'identification si un instrument endogène est ajouté à un ensemble d'instruments exogènes? Est-ce que ces procédures alternatives se comportent différemment? Si oui, quelle est leur performance relative en présence d'instruments endogènes?

Nous voyons le probème de l'endogénéité des instruments comme important parce qu'en pratique, il est difficile d'évaluer si une variable instrumentale est exogène ou non. La validité des instruments ou le test d'orthogonalité habituel est construit avec l'hypothèse qu'un ensemble d'instruments (au moins égal au nombre de coefficients à estimer dans le modèle) est disponible, alors que la validité de ce dernier ensemble est intestable. Dans la littérature économétrique, on connaît très peu des procédures de test lorsque les instruments sont à la fois endogènes et faibles.

Cet essai étudie les effets de l'endogénéité des instruments sur deux statistiques de test robustes à l'identification: la statistique d'Anderson et Rubin (AR, 1949) et la statistique de Kleibergen (K, 2003), avec ou sans instruments faibles.

D'abord, lorsque le paramètre qui contrôle l'endogénéité des instruments est fixe (ne dépend pas de la taille de l'échantillon), nous montrons que toutes ces procédures sont en général consistantes contre la présence d'instruments invalides (c'est-à-dire qu'elles détectent la présence d'instruments invalides) indépendamment de leur qualité (forts ou faibles). Nous décrivons aussi des cas où cette consistance peut ne pas tenir, mais la distribution asymptotique est modifiée d'une manière qui pourrait conduire aux distorsions de niveau en grands échantillons. Ceci inclut en particulier, les cas où l'estimateur des double moindres carrés demeure consistant, mais les tests sont asymptotiquement invalides. Ensuite, lorsque les instruments sont localement exogènes (c'est-à-dire le paramètre d'endogénéité converge vers zéro lorsque la taille de l'échantillon augmente), nous montrons que ces tests convergent vers des distributions de chi carré non centré, que les instruments soient forts ou faibles. Nous caractérisons aussi les situations où le paramètre de non centralité est nul et la distribution asymptotique des statistiques demeure la même que dans le cas des instruments valides (malgré la présence des instruments invalides).

Le deuxième essai étudie l'impact des instruments faibles sur les tests de spécification du type Durbin-Wu-Hausman (DWH) ainsi que le test de Revankar et Hartley (1973). Dans le modèle de régression linéaire, les tests d'exogénéité de Durbin-Wu-Hausman et Revankar-Hartley sont souvent utilisés dans le but de corriger la corrélation entre les variables explicatives et les erreurs, habituellement en recourant à la méthode d'estimation par variables instrumentales. Un autre problème courant dans les travaux empiriques et où ces tests sont utilisés consiste à pré-tester l'exogénéité d'un instrument, voir par exemple Bradford (2003). Cependant, ces tests reposent sur l'hypothèse que les paramètres du modèle sont identifiés, c'est-à-dire que les instruments disponibles sont fortement corrélés avec la variable à instrumenter. Ce qui pose la question de savoir ce qui arrive à ces procédures lorsque les instruments sont faibles.

Dans cet essai, nous proposons une analyse en petit-et grand-échantillon de la distribution de ces tests sous l'hypothèse nulle (niveau) et l'alternative (puissance), incluant les cas où l'identification est déficient ou faible (instruments faibles). Notre analyse en petit-échantillon founit plusieurs nouvelles perspectives et extensions des précédentes procédures. En effet, la caractérisation de la distribution de ces statistiques en petit-échantillon permet la construction des tests de Monte Carlo exacts pour l'exogénéité même avec les erreurs non Gaussiens. Nous montrons que ces tests sont typiquement robustes aux intruments faibles (le niveau est contrôlé). De plus, nous fournissons une caractérisation de la puissance des tests, qui exhibe clairement les facteurs qui déterminent la puissance. Nous montrons que les tests n'ont pas de puissance lorsque tous les instruments sont faibles [similaire à Guggenbergen (2008)]. Cependant, la puissance existe dès que au moins des instruments est fort. La conclusion de Guggenberger (2008) se focalise sur le cas où tous les instruments sont faibles (un cas d'intérêt mineur en pratique). Notre théorie asymptotique sous les hypothèses affaiblies confirme la théorie en échantillon fini.

Par ailleurs, nous présentons une analyse de Monte Carlo indiquant que: (1) l'estimateur des moindres carrés ordinaires est plus efficace que celui des doubles moindres carrés lorsque les instruments sont faibles et l'endogenéité modérée [conclusion similaire celle de Kiviet and Niemczyk (2007)]; (2) les estimateurs pré-test basés sur les tests d'exogénéité ont une excellente performance comparés aux doubles moindres carrés. Ce qui suggère que la méthode des variables instrumentales ne devrait être appliquée que si l'on a la certitude d'avoir des instruments fort. Donc, les conclusions de Guggenberger (2008) sont mitigées et pourraient être trompeuses.

Nous illustrons nos résultats théoriques par des expériences de simulation et deux applications empiriques: la relation entre le taux d'ouverture et la croissance économique et le problème bien connu de rendements à l'éducation.

Dans le troisième essai, nous proposons une extension du test d'exogénéité partielle de Wald proposé par Dufour (1987) aux cas où les erreurs de la régression ont une distribution non-Gaussienne. Dans les travaux empiriques, un problème recurrent consiste à tester l'exogénéité d'un sous-ensemble de régresseurs. Par exemple, dans le modèle de rendement à l'éducation, on aimerait tester l'exogénéité du nombre d'années de fréquentation de l'individu sans contraindre son habileté (mesurée par le QI) et les autres variables. Dans le modèle de croissance économique [Franckel et Romer (1999)], la validité de l'instrument utilisé pour le taux d'ouverture est souvent mis en cause et l'on a besoin de savoir si cet instrument est effectivement exogène.

Les tests d'exogénéité du type Durbin-Wu-Hausman et Rvankar-Hartley ne fournissent pas une réponse à ces questions. Le test de Wald [Dufour (1987)] répond à ces préoccupations. Cependant, ce test est construit sous l'hypothèse de normalité des erreurs. Qu'arrive-t-il alors à ce test lorsque les erreurs ne sont plus Gaussiens?

Dans cet essai, nous développons une nouvelle version du précédent test qui est valide même en présence d'erreurs non-Gaussiens. Nous proposons également deux nouveaux estimateurs pré-test basés sur ce test qui performent mieux que l'estimateur de variables instrumentales usuel lorsque les instruments sont faibles et l'endogénéité modérée. Nous montrons que le test peut servir de procédure de sélection de variables instrumentales. Nous illustrons les résultats théoriques à travers deux applications empiriques: le modèle bien connu d'équation du salaire [Angist et Krueger (1991, 1999)] et les rendements d'échelle [Nerlove (1963)]. Nos résultats suggèrent que l'éducation de la mère expliquerait le décrochage de son fils, que l'output est une variable endogène dans l'estimation du coût de la firme et que le prix du fuel en est un instrument valide pour l'output.

Le quatrième essai se démarque du troisième par deux contributions majeures. D'abord, cet essai développe une procédure d'inférence valide sur les covariances entre les erreurs et les régresseurs possiblement endogènes tant en grands échantillons qu'en petits échantillons. Ensuite, contrairement au test du type Wald, cette procédure est robuste à l'identification, C'est-à-dire, qu'elle demeure valide même si les instruments sont faibles. Bien que le test de Wald initial ou étendu permet de construire les régions de confiance et de tester les restrictions linéaires sur les covariances, il est construit avec l'hypothèse questionable que les paramètres du modèle sont identifiés. Lorsque l'identification est faible (instruments faiblement corrélée avec la variable à instrumenter), ce test n'est en général plus valide. Cet essai développe une procédure d'inférence robuste à l'identification (instruments faibles) pour construire des régions de confiance pour la matrices de covariances entre les erreurs de la régression et les variables explicatives (possiblement endogènes). Nous fournissons les expressions analytiques des régions de confiance et caractérisons les conditions nécessaires et suffisantes sous lesquelles ils sont bornés. La procédure proposée demeure valide même pour les petits échantillons et est aussi asymptotiquement robuste à l'hétéroscédasticité et l'autocorrélation des erreurs. Ces résultats sont alors utilisés pour développer les tests d'exogénéité partielle robustes à l'identification. Les simulations Monte Carlo indiquent que les tests contrôlent le niveau et ont de la puissance même si les instruments sont faibles. Ce qui nous permet de proposer une procédure de sélection de variables instrumentales valide même s'il y a un problème d'identification. La procédure de sélection des instruments est basée sur deux nouveaux estimateurs pré-test qui combinent l'estimateur IV usuel et des estimateurs IV partiels. Nos simulations montrent que: (1) tout comme l'estimateur des moindres carrés ordinaires, les estimateurs IV partiels sont plus efficaces que l'estimateur IV usuel lorsque les instruments sont faibles et l'endogénéité modérée; (2) les estimateurs pré-test ont globalement une performance excellente comparés à l'estimateur IV usuel. Nous illustrons nos résultats théoriques à travers deux apllications empiriques: la relation entre le taux d'ouverture et la croissance économique et le modèle de rendements à l'éducation. Dans la première application, les études antérieures ont conclu que les instruments n'étaient pas trop faibles [Dufour et Taamouti (2007)] alors qu'ils le sont fortement dans la second [Bound (1995), Doko et Dufour (2009)]. Conformément à notre théorie, nous trouvons les régions de confiance non bornées pour la covariance dans le cas où les instruments sont assez faibles.

Contribution du Coauteur

Je suis le premier auteur de tous les quatre articles de cette thèse. Cependant, mon coauteur Jean-Marie Dufour et moi-même contribuons à part égale.

Doko Tchatoka, Firmin Sabro

Chapter 1

Instrument endogeneity and identification-robust tests: some analytical results, *article published in Journal of Statistical Planning and Inference 138 (2008) 2649 – 2661*

1. Introduction

The last decade shows growing interest for so-called *weak instruments* problems in the econometric literature, *i.e.* situations where “instruments” are poorly correlated with endogenous explanatory variables; see the reviews of Dufour (2003) and Stock, Wright and Yogo (2002). More generally, these can be viewed as situations where model parameters are not identified or close not to being *identifiable*, as meant in the econometric literature [see Dufour and Hsiao (2008)]. When instruments are weak, the limiting distributions of standard test statistics – like Student, Wald, likelihood ratio and Lagrange multiplier criteria in structural models – often depend heavily on nuisance parameters; see *e.g.* Phillips (1989), Bekker (1994), Dufour (1997), Staiger and Stock (1997) and Wang and Zivot (1998). In particular, standard Wald-type procedures based on the use of asymptotic standard errors are very unreliable in the presence of weak identification. As a result, several authors have worked on proposing more reliable statistical procedures that would be applicable in such contexts.

Interestingly, in the early days of simultaneous-equations econometrics, Anderson and Rubin (1949, AR) proposed a procedure which is completely robust to weak instruments as well as to other difficulties such as missing instruments [see Dufour (2003), Dufour and Taamouti (2005, 2006)]. But the AR procedure may suffer from power losses when too many instruments are used. So alternative methods largely try to palliate this difficulty, for example: pseudo-pivotal LM-type and LR-type statistics [Wang and Zivot (1998), Kleibergen (2002), Moreira (2003)], sample-splitting methods [Dufour and Jasiak (2001)], approximately optimal instruments [Dufour and Taamouti (2003)], systematic search methods for identifying relevant instruments and excluding unimportant instruments [Hall, Rudebusch and Wilcox (1996), Hall and Peixe (2003), Dufour and Taamouti (2003), Donald and Newey (2001)].

However, all these procedures – including the AR method – rely on the availability on valid (exogenous) instruments. This raises the question: what happens to these procedures when some of the instruments are endogenous? In particular, what happens if an invalid instrument is added to a set of valid instruments? How robust are these inference procedures to instrument endogeneity? Do alternative inference procedures behave differently? If yes, what is their relative performance in the presence of instrument endogeneity?

We view the problem of instrument endogeneity as important because it is hard in practice to assess whether an instrumental variable is valid, *i.e.* whether it is uncorrelated with the disturbance term. Instrument validity or orthogonality tests are built on the availability of a number of undisputed valid instruments, at least as great as the number of coefficients to be estimated, whereas the validity of those initial instruments is not testable.

In the econometric literature, little is known about test procedures when some instruments are both invalid and weak. Hausman and Hahn (2002) deal with both instrument endogeneity and weakness, but they focus on estimation. Ashley (2006) proposed a sensitivity analysis of IV estimators when instruments are imperfect, his results however are only applicable if the covariance between the structural error term and some instruments is known, which is not necessary the case as it is showed in this paper. Analyzing the effect of instrument invalidity on the limiting and empirical distribution of IV estimators, Kiviet and Niemczyk (2006) conclude that for the accuracy of asymptotic approximations, instrument weakness is much more detrimental than instrument invalidity and that the realizations of IV estimators based on strong but possibly invalid instruments seem usually much closer to the true parameter values than those obtained from valid but weak instruments. However, this finding of Kiviet and Niemczyk leaves open crucial questions: is it really possible to make reliable inference with endogenous instruments? Is instrument endogeneity really more detrimental than its weakness on inference procedures like a general family of Anderson-Rubin-type procedures? Swanson and Chao (2005) proposed a weak-instrument unified framework, but they do not take into account possible invalidity of some instruments. Finally, Small (2007) has recently studied the properties of tests for identifying restrictions [Sargan (1958), Anderson and

Kadane (1977)], which can be sensitive to the use of "endogenous instruments", and he proposed a sensitivity analysis to assess the importance of the issue. These results, however, do not allow for weak identification.

In this paper, we focus on structural models and analyze the effects of instrument endogeneity on the Anderson and Rubin (1949) and Kleibergen (2002) tests, in the presence of possibly weak instruments. After formulating a general asymptotic framework which allows one to study these issues in a convenient way, we consider two main setups: (1) the one where the level of "instrument" endogeneity is fixed (*i.e.*, it does not depend on the sample size), and (2) the one where the instruments are *locally exogenous*, *i.e.* the parameter which controls instrument endogeneity approaches zero (at rate $T^{-1/2}$) as the sample size increases. In the first setup, we show that both test procedures studied are in general consistent against the presence of invalid instruments (hence asymptotically invalid for the hypothesis of interest), whether the instruments are "strong" or "weak". We also observe there are cases where consistency may not hold, but the asymptotic distribution is modified in a way that would lead to size distortions in large samples. In the second setup, asymptotic non-central chi-square distributions are derived, and we give conditions under which the non-centrality parameter is zero and the asymptotic distribution remains the same as in the case of valid instruments (despite the presence of invalid instruments). Overall, our results underscore the importance of checking for the presence of possibly invalid instruments when applying "identification-robust" tests.

The paper is organized as follows. Section 2 formulates the model considered. Section 3 describes briefly the statistics. Section 4 studies the asymptotic distribution of the statistics (under the null hypothesis) when some instruments are invalid. We conclude in section 9. Proofs are presented in the Appendix.

2. Framework

We consider the following standard simultaneous equation framework, which has been the basis of much work on inference in model with possibly weak instruments [see the reviews of Dufour (2003) and Stock et al. (2002)]:

$$y = Y\beta + Z\gamma + u\,, \tag{2.1}$$

$$Y = X\mathit{\Pi} + Z\mathit{\Gamma} + V\,, \tag{2.2}$$

where y is a $T \times 1$ vector of observations on the dependent variable, $Y = [Y_1, \ldots, Y_T]'$ is a $T \times G$ matrix of observations on explanatory (possibly) endogenous variables $(G \geq 1)$, Z is a $T \times r$ matrix of observations on the included exogenous variables, $X = [X_1, \ldots, X_T]'$ is a $T \times k$ $(k \geq G)$ full-column-rank matrix of observations on (supposedly) "exogenous variables" (instruments) excluded from the structural equation (2.1), $u = [u_1, \ldots, u_T]'$ and $V = [V_1, \ldots, V_T]' = [v_1, \ldots, v_G]$ are respectively $T \times 1$ vector and $T \times G$ disturbance matrices, β and γ are $G \times 1$ and $r \times 1$ vectors of unknown coefficients, $\mathit{\Pi}$ and $\mathit{\Gamma}$ are $k \times G$ and $r \times G$ matrices of unknown coefficients. The usual necessary and sufficient condition for identification of this model is $\mathrm{rank}(\mathit{\Pi}) = G$.

Since we focus on the parameter β in our analysis, we can simplify the presentation of the results without notable loss of generality by setting $\gamma = 0$ and $\mathit{\Gamma} = 0$, so that Z drops from the model. With this simplification, model (2.1)-(2.2) reduces to

$$y = Y\beta + u\,, \tag{2.3}$$

$$Y = X\mathit{\Pi} + V\,. \tag{2.4}$$

We also assume that

$$u_t = V_t' a + \varepsilon_t\,,\ t = 1, \ldots, T\,, \tag{2.5}$$

$$X_t = X_{0t} + W_t,\ t = 1, \ldots, T\,, \tag{2.6}$$

$$u_t = W_t' b + e_t,\ t = 1, \ldots, T, \tag{2.7}$$

where $X_0 = [X_{01}, \ldots, X_{0T}]'$ is a $T \times k$ matrix of exogenous variables, ε_t is uncorrelated with V_t, and e_t are uncorrelated with W_t . V_t and W_t have mean zero and covariance matrices Σ_V and Σ_W, ε_t and e_t have mean zero and variances σ_ε^2 and σ_e^2 respectively, while a and b are $G \times 1$ and $k \times 1$ vectors of unknown coefficients. (2.5)-(2.7) can be rewritten in matrix form as:

$$u = Va + \varepsilon\,, \tag{2.8}$$

$$X = X_0 + W\,, \tag{2.9}$$

$$u = Wb + e\,, \tag{2.10}$$

where X_0 is uncorrelated with W, V, ε and e, while $W = [W_1, \ldots, W_T]'$ is uncorrelated with e but may be correlated with u (when $b \neq 0$). So a controls the endogeneity of the variable Y, whereas b represents the possible endogeneity of the instruments X. If $b = 0$, the instruments X are valid; otherwise, they are invalid (endogenous). More precisely, if $b \neq 0$, *i.e.*, there exists at least one i such that $b_i \neq 0$, $i = 1, \ldots, k$, and the corresponding variable X_i does not constitute a valid instrument.

We also make the following generic assumptions on the asymptotic behaviour of model variables [where $A > 0$ for a matrix A means that A is positive definite (p.d.), and $\to$ refers to limits as $T \to \infty$]:

$$\frac{1}{T}\begin{bmatrix} V & \varepsilon \end{bmatrix}'\begin{bmatrix} V & \varepsilon \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \Sigma_V & 0' \\ 0 & \sigma_\varepsilon^2 \end{bmatrix} > 0\,, \tag{2.11}$$

$$\frac{1}{T}\begin{bmatrix} X_0 & W \end{bmatrix}'\begin{bmatrix} X_0 & W \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \Sigma_0 & 0' \\ 0 & \Sigma_W \end{bmatrix}, \quad \Sigma_0 > 0\,, \tag{2.12}$$

$$\frac{1}{T}X_0'\begin{bmatrix} V & \varepsilon & e \end{bmatrix} \xrightarrow{p} 0\,, \tag{2.13}$$

$$\frac{1}{T}X'X \xrightarrow{p} \Sigma_X\,, \tag{2.14}$$

$$\frac{1}{T}\begin{bmatrix} W & e \end{bmatrix}'\begin{bmatrix} W & e \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \Sigma_W & 0' \\ 0 & \sigma_e^2 \end{bmatrix}, \tag{2.15}$$

$$\frac{1}{T}W'V \xrightarrow{p} \Sigma_{WV}\,, \tag{2.16}$$

$$\frac{1}{\sqrt{T}}\begin{bmatrix} X'e \\ (X'W-\Sigma_W)b \end{bmatrix} \xrightarrow{L} \begin{bmatrix} S_e \\ S_b \end{bmatrix} \sim N\left[0\,,\,\Sigma_S\right]\,, \tag{2.17}$$

$$S_e \sim N\left[0,\,\sigma_e^2\Sigma_X\right]\,,\quad S_b \sim N\left[0,\,\sigma_e^2\Sigma_b\right]\,, \tag{2.18}$$

where Σ_V is $G \times G$ fixed matrix, Σ_0 and Σ_W are $k \times k$ fixed matrices, S_e and S_b are $k \times 1$ random vectors. Note that Σ_W may be singular, and S_b may not be independent of S_e.

From the above assumptions, it is easy to see that:

$$\frac{1}{T}X_0'u \xrightarrow{p} 0\,,\quad \frac{1}{T}X_0'e \xrightarrow{p} 0\,, \tag{2.19}$$

$$\frac{1}{T}X'u \xrightarrow{p} \varphi = \Sigma_W b\,,\quad \frac{X'V}{T} \xrightarrow{p} \Sigma_{WV}\,, \tag{2.20}$$

$$\frac{1}{T}\begin{bmatrix} u & V \end{bmatrix}'\begin{bmatrix} u & V \end{bmatrix} \xrightarrow{p} \Sigma = \begin{bmatrix} \sigma_u^2 & \delta' \\ \delta & \Sigma_V \end{bmatrix} > 0\,, \tag{2.21}$$

$$\frac{1}{T}\begin{bmatrix} u & W \end{bmatrix}'\begin{bmatrix} u & W \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \sigma_u^2 & \varphi' \\ \varphi & \Sigma_W \end{bmatrix}\,, \tag{2.22}$$

$$\frac{1}{T}\begin{bmatrix} W & V \end{bmatrix}'\begin{bmatrix} \varepsilon & e \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \delta_{W\varepsilon} & 0 \\ 0 & \delta_{Ve} \end{bmatrix}\,, \tag{2.23}$$

where

$$\delta = \Sigma_V a\,,\; \sigma_u^2 = a'\Sigma_V a + \sigma_\varepsilon^2 = \sigma_e^2 + b'\Sigma_W b\,, \tag{2.24}$$

$$\Sigma_V a = \Sigma_{WV}' b + \delta_{Ve}\,,\; \Sigma_W b = \Sigma_{WV} a + \delta_{W\varepsilon}\,, \tag{2.25}$$

$$\Sigma_X = \Sigma_0 + \Sigma_W > 0\,,\quad \Sigma_{XY} = \Sigma_X \Pi + \Sigma_{WV}\,, \tag{2.26}$$

$$\Sigma_Y = \Pi'\Sigma_X\Pi + \Sigma_V + \Sigma_{WV}'\Pi + \Pi'\Sigma_{WV}\,. \tag{2.27}$$

Finally, we denote by $\mathcal{N}(\Sigma_W)$ the null set of the linear map on $\mathbb{R}^k$ characterized by the matrix Σ_W :

$$\mathcal{N}(\Sigma_W) = \{x \in \mathbb{R}^k : \Sigma_W x = 0\}\,. \tag{2.28}$$

If Σ_W is a full-column-rank matrix, then $\mathcal{N}(\Sigma_W) = \{0\}$; otherwise, there is at least one $x_0 \neq 0$ such that $\Sigma_W x_0 = 0$.

The setup described above is quite wide and does allow one to study several questions associated with the possible presence of "invalid" instruments. In particular, an important practical problem consists in studying the effect on inference of adding an "invalid" instrument to a list of valid (possibly identifying) instruments. Note that this problem is distinct from studying the effect of imposing "incorrect" overidentifying restrictions [as done by Small (2007)]. To better see the issues studied here, it will be useful to consider a simple example.

Example 2.1 Consider a model with one endogenous explanatory variable $(G = 1)$ and two candidate instruments $(k = 2)$. Then Y and V are $T \times 1$ vectors, $X = [X_1,\, X_2]$ and $W = [W_1,\, W_2]$ are $T \times 2$ matrices, $\mathit{\Pi} = [\pi_1,\, \pi_2]'$ and $b = [b_1,\, b_2]'$ are vectors of dimension 2, and

$$Y = X\mathit{\Pi} + V = X_1\pi_1 + X_2\pi_2 + V\,, \tag{2.29}$$

$$u = Wb + e = W_1 b_1 + W_2 b_2 + e\,. \tag{2.30}$$

Let us further assume that X_1 is a valid instrument (with $W_1 = 0$), $\mathsf{E}[u \,|\, X_1] = 0$, $X_2 = W_2$, $\pi_2 = 0$ and $b_1 = 0$, where e is independent of X_1 and X_2 (with finite mean zero), so that

$$Y = X\mathit{\Pi} + V = X_1\pi_1 + V\,, \tag{2.31}$$

$$u = Wb + e = W_2 b_2 + e\,. \tag{2.32}$$

Here W_2 is not a "valid" instrument when $b_2 \neq 0$. But the structural equation (2.3) may in principle be estimated using only X_1 as an instrument, because $\mathsf{E}[u \,|\, X_1] = 0$; if X_1 is not a weak instrument $(\pi_1 \neq 0)$ and satisfies usual regularity conditions, a consistent estimate of β can be obtained. Among other things, we study below the effect (on some identification-robust tests) of taking X_2 as an instrument when $b_2 \neq 0$, *i.e.* when X_2 is correlated with u. Note that the condition $E[u \,|\, X_1] = 0$ *does not entail* $\mathsf{E}[e \,|\, X_1,\, X_2] = 0$, which is a maintained hypothesis used by Small (2007). So the problem considered here is distinct from the problem of testing overidentifying restrictions [studied, for example, by Sargan (1958), Anderson and Kadane (1977) and Small (2007)].

3. Test statistics

We consider in this paper the problem of testing

$$H_0 : \beta = \beta_0 \tag{3.1}$$

where some of the "instruments" used are in fact endogenous $(b \neq 0)$. We analyze the behavior of the Anderson-Rubin and Kleibergen statistics. The Anderson and Rubin (1949) test for H_0 in equation (2.3) involves considering the transformed equation

$$y - Y\beta_0 = X\varDelta + \varepsilon \tag{3.2}$$

where $\varDelta = \varPi(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. H_0 can then be assessed by testing $H_0' : \varDelta = 0$. The AR-statistic for H_0' is given by

$$AR(\beta_0) = \frac{1}{k}\frac{(y - Y\beta_0)'P_X(y - Y\beta_0)}{(y - Y\beta_0)'M_X(y - Y\beta_0)/(T-k)} \tag{3.3}$$

where $M_B = I - P_B$ and $P_B = B(B'B)^{-1}B'$ is the projection matrix on the space spanned by the columns of B. If $b = 0$, the asymptotic distribution of AR(β_0) is a $\chi^2(k)/k$ under H_0. If furthermore $u \sim N[0\,,\sigma^2 I_T]$ and X is independent of u, then $AR(\beta_0) \sim F(k, T-k)$ under H_0 irrespective of whether the instruments are strong or weak. However, when some instruments are invalid, the distribution of the AR statistic may be affected.

Kleibergen (2002) proposed a modification of the AR statistic to take into account the fact that this statistic may have low power when there are too many instruments in the model. The modified statistic for testing H_0 can be written

$$K(\beta_0) = \frac{(y - Y\beta_0)'P_{\tilde{Y}(\beta_0)}(y - Y\beta_0)}{(y - Y\beta_0)'M_X(y - Y\beta_0)/(T-k)} \tag{3.4}$$

where

$$\tilde{Y}(\beta_0) = X\tilde{\Pi}(\beta_0)\,,\; \tilde{\Pi}(\beta_0) = (X'X)^{-1}X'\left[Y - (y - Y\beta_0)\frac{S_{uV}(\beta_0)}{S_{uu}(\beta_0)}\right]\,, \tag{3.5}$$

$$S_{uu}(\beta_0) = \frac{1}{T-k}(y - Y\beta_0)'M_X(y - Y\beta_0)\,,\; S_{uV}(\beta_0) = \frac{1}{T-k}(y - Y\beta_0)'M_X Y\,. \tag{3.6}$$

Unlike the AR statistic which projects $y - Y\beta_0$ on the k columns of X, the K statistic projects $y - Y\beta_0$ on the G columns of $X\tilde{\Pi}(\beta_0)$. If the instruments X are exogenous, $\tilde{\Pi}(\beta_0)$ is both a consistent estimator of Π and asymptotically independent of $X'(y - Y\beta_0)$ under H_0, and $K(\beta_0)$ converges to a $\chi^2(G)$. However, if some instruments are invalid ($b \neq 0$), $\tilde{\Pi}(\beta_0)$ may not be asymptotically independent of $X'(y - Y\beta_0)$ and the asymptotic distribution of the K statistic may not be a $\chi^2(G)$.[1]

If the model contains only one instrument and one endogenous variable $(G = k = 1)$, the AR and K statistics are equivalent and pivotal even in finite samples whenever $b = 0$. When $k > 1$, even if $b = 0$, the K statistic is not pivotal in finite samples but is asymptotically pivotal, whereas the AR statistic is pivotal even in finite samples (when X is independent of u). Following Staiger and Stock (1997), we refer to the *locally weak-instrument* asymptotic setup by considering a limiting sequence of Π where Π is local-to-zero. We also consider a limiting sequence of b where b is local-to-zero. We refer to this later limiting sequence as *locally exogenous* instruments asymptotic.

4. Asymptotic theory with invalid and weak instruments

In this section, we study the large-sample properties of the statistics described above when some of the instruments used are invalid. Two setups are considered. The first is the possibly invalid instrument setup, *i.e.*, the endogeneity parameter b is a fixed vector. The second is the locally exogenous instrument setup, *i.e.*, b is local-to-zero.

[1]We do not study this paper conditional tests such as those proposed by Moreira (2003), because the distributional theory for such tests is considerably more complex and would go beyond the scope of a short paper like the present one.

4.1. Possibly invalid instruments

We consider first the case where the endogeneity parameter b is a constant vector and we analyze the asymptotic distributions of the statistics. Our results cover both strong and weak-instrument asymptotic. Theorem **4.1** below summarizes the asymptotic behavior of the AR statistic when some instruments may be endogenous. For a random variable S whose distribution depends on the sample size T, the notation $S \xrightarrow{L} +\infty$ means that $P[S > x] \to 1$ as $T \to \infty$, for any x.

Theorem 4.1 ASYMPTOTIC DISTRIBUTION OF THE AR STATISTIC. *Suppose that the assumptions* (2.3)-(2.18) *hold, with* $b = b_0$ *and* $\beta = \beta_0$*, where* b_0 *and* β_0 *are given vectors.*
If $b_0 \notin \mathcal{N}(\Sigma_W)$*, then*

$$AR(\beta_0) \xrightarrow{L} +\infty\,. \tag{4.1}$$

If $b_0 \in \mathcal{N}(\Sigma_W)$*, then*

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k\sigma_u^2}(S_e + S_b)'\Sigma_X^{-1}(S_e + S_b) \tag{4.2}$$

where S_e *and* S_b *are defined in* (2.17)-(2.18).
If $b_0 = 0$*, then*

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k}\chi^2(k)\,. \tag{4.3}$$

In the above theorem, no restriction is imposed on the rank of Π. In particular, the result holds even if Π is not a full-column rank matrix. When $b_0 \notin \mathcal{N}(\Sigma_W)$, the AR statistic diverges under the null hypothesis H_0. When $b_0 \in \mathcal{N}(\Sigma_W)$, the limiting distribution of the AR statistic does not diverge, but the AR test is not valid unless $S_b = 0$. Of course, when $b_0 = 0$ – which is the classical exogenous instrument setup – $S_b = 0$ and the AR test is asymptotically valid.

Theorem **4.2** below summarizes the asymptotic behavior of the K statistic when some instruments are possibly invalid.

Theorem 4.2 ASYMPTOTIC DISTRIBUTION OF THE K STATISTIC. *Suppose that the assumptions* (2.3)-(2.18) *hold, with* $b = b_0$ *and* $\beta = \beta_0$, *where* b_0 *and* β_0 *are given vectors.*
(A) *If* $b_0 \notin \mathcal{N}(\Sigma_W)$ *then*

$$K(\beta_0) \xrightarrow{L} +\infty \tag{4.4}$$

when at least one of the following two conditions holds: (i) $\Pi = \Pi_0 \neq 0$ *with* $\text{rank}(\tilde{\Sigma}_{XY}) = G$, *or* (ii) $\Pi = \Pi_0/\sqrt{T}$ *with* $\text{rank}(\Sigma^{*}_{XY}) = G$, *where*

$$\begin{aligned} \tilde{\Sigma}_{XY} &= \Sigma_{XY} - \Sigma_W b_0 (q_{uV}/\bar{\sigma}_u^2)\,, \quad \Sigma^{*}_{XY} = \Sigma_{WV} - \Sigma_W b_0 (q_{uV}/\bar{\sigma}_u^2)\,, \\ q_{uV} &= \delta' - b_0'\Sigma_W \Sigma_X^{-1}\Sigma_{WV}\,, \quad \bar{\sigma}_u^2 = \sigma_u^2 - b_0'\Sigma_W\Sigma_X^{-1}\Sigma_W b_0\,. \end{aligned}$$

(B) *If* $b_0 \in \mathcal{N}(\Sigma_W)$, *then*

$$K(\beta_0) \xrightarrow{L} \frac{1}{\sigma_u^2}(S_e + S_b)'\Sigma_X^{-1}\Sigma_{XY}(\Sigma_{XY}'\Sigma_X^{-1}\Sigma_{XY})^{-1}\Sigma_{XY}'\Sigma_X^{-1}(S_e + S_b) \tag{4.5}$$

when $\Pi = \Pi_0 \neq 0$ *and* $\text{rank}(\Sigma_{XY}) = G$, *and*

$$K(\beta_0) \xrightarrow{L} \frac{1}{\sigma_u^2}(S_e + S_b)'\Sigma_X^{-1}\Sigma_{WV}(\Sigma_{WV}'\Sigma_X^{-1}\Sigma_{WV})^{-1}\Sigma_{WV}'\Sigma_X^{-1}(S_e + S_b) \tag{4.6}$$

when $\Pi = \Pi_0/\sqrt{T}$ *and* $\text{rank}(\Sigma_{WV}) = G$. (C) *If* $b_0 = 0$, *then*

$$K(\beta_0) \xrightarrow{L} \chi^2(G) \tag{4.7}$$

when at least one of the following two conditions holds: (i) $\Pi = \Pi_0 \neq 0$ *with* $\text{rank}(\Sigma_{XY}) = G$, *or (ii)* $\Pi = \Pi_0/\sqrt{T}$ *with* $\text{rank}(\Sigma_{WV}) = G$.

Unlike Theorem **4.1** for the AR statistic, Theorem **4.2** requires an additional rank assumption. When $b_0 \notin \mathcal{N}(\Sigma_W)$, the null limiting distribution of the K statistic diverges. This means that the K test often rejects H_0 asymptotically when $b_0 \notin \mathcal{N}(\Sigma_W)$. Furthermore, when $b_0 \in \mathcal{N}(\Sigma_W)$, the K test is not asymptotically valid unless $S_b = 0$. As expected, if $b_0 = 0$ (*i.e.*, $S_b = 0$), the K statistic converges to a $\chi^2(G)$. It is worthwhile to note that the case where the rank assumption fails [*e.g.*, the partial identification of β] is not covered in this paper.

Finally, it is interesting to observe that the limiting value of the two-stage least-squares (2SLS) estimator of β,

$$\tilde{\beta} = \big(\hat{Y}'\hat{Y}\big)^{-1}\hat{Y}'y = \big[Y'X(X'X)^{-1}X'Y\big]^{-1}Y'X(X'X)^{-1}X'y\,, \tag{4.8}$$

is given by

$$\operatorname*{plim}_{T\to\infty}\tilde{\beta} = \beta + \big[\Sigma_{XY}'\Sigma_X^{-1}\Sigma_{XY}\big]^{-1}\Sigma_{XY}'\Sigma_X^{-1}\Sigma_W b \tag{4.9}$$

provided $\mathrm{rank}(\Sigma_{XY}) = G$, so that $\tilde{\beta}$ is consistent when $b_0 \in \mathcal{N}(\Sigma_W)$ and Σ_{XY} has full column rank (even if some instruments are invalid). If $b_0 \in \mathcal{N}(\Sigma_W)$ but $b_0 \neq 0$, the asymptotic level of the Anderson-Rubin and Kleibergen tests can be affected.

4.2. Locally exogenous instruments

We consider now the case where the endogeneity parameter b is local-to-zero. As in the previous subsection, we analyze the limiting distributions of the statistics. The results also cover two setups: locally exogenous instruments [$\Pi = \Pi_0 \neq 0$, $b = b_0/\sqrt{T}$], and weak locally exogenous instruments [$\Pi = \Pi_0/\sqrt{T}$, $b = b_0/\sqrt{T}$]. Theorem **4.3** and Theorem **4.4** below derive the distributions of the statistics for both setups.

Theorem 4.3 ASYMPTOTIC DISTRIBUTIONS WITH LOCALLY EXOGENOUS INSTRUMENTS. *Suppose that the assumptions* (2.3)-(2.18) *hold, with* $b = b_0/\sqrt{T}$, $\Pi = \Pi_0 \neq 0$ *and* $\beta = \beta_0$, *where* b_0 *and* β_0 *are given vectors, and* Π_0 *is a given matrix. If* $b_0 \notin \mathcal{N}(\Sigma_W)$, *then*

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k}\chi^2(k\,,\mu_1)\,, \tag{4.10}$$

$$K(\beta_0) \xrightarrow{L} \chi^2(G\,,m'm) \quad \textit{if } \mathrm{rank}(\Sigma_{XY}) = G\,, \tag{4.11}$$

where

$$\mu_1 = \frac{1}{\sigma_e^2} b_0' \Sigma_W \Sigma_X^{-1} \Sigma_W b_0 \,,\; m = \frac{1}{\sigma_e} (\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY})^{-1/2} \Sigma_{XY}' \Sigma_X^{-1} \Sigma_W b_0 \,, \tag{4.12}$$

and Σ_X*,* Σ_{XY}*, and* Σ_W *are given in* (2.11)-(2.27)*. If* $b_0 \in \mathcal{N}(\Sigma_W)$*, then*

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k}\chi^2(k), \tag{4.13}$$

$$K(\beta_0) \xrightarrow{L} \chi^2(G) \quad \textit{if } \mathrm{rank}(\Sigma_{XY}) = G \,. \tag{4.14}$$

Theorem 4.4 ASYMPTOTIC DISTRIBUTIONS WITH WEAK LOCALLY EXOGENOUS INSTRUMENTS. *Suppose that the assumptions* (2.3)-(2.18) *hold, with* $b = b_0/\sqrt{T}$*,* $\Pi = \Pi_0/\sqrt{T}$ *and* $\beta = \beta_0$*, where* b_0 *and* β_0 *are given vectors, and* Π_0 *is a given matrix* ($\Pi_0 = 0$ *is allowed*)*. If* $b_0 \notin \mathcal{N}(\Sigma_W)$*, then*

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k}\chi^2(k \,, \mu_1) \,, \tag{4.15}$$

$$K(\beta_0) \xrightarrow{L} \chi^2(G \,, \tilde{m}'\tilde{m}) \quad \textit{if } \mathrm{rank}(\Sigma_{WV}) = G \,, \tag{4.16}$$

where

$$\tilde{m} = \frac{1}{\sigma_e} (\Sigma_{WV}' \Sigma_X^{-1} \Sigma_{WV})^{-1/2} \Sigma_{WV}' \Sigma_X^{-1} \Sigma_W b_0 \,, \tag{4.17}$$

and Σ_X*,* Σ_{WV}*,* Σ_W *and* μ_1 *are defined in Theorem* **4.3***. If* $b_0 \in \mathcal{N}(\Sigma_W)$*, then*

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k}\chi^2(k) \,, \tag{4.18}$$

$$K(\beta_0) \xrightarrow{L} \chi^2(G) \quad \textit{if } \mathrm{rank}(\Sigma_{WV}) = G \,. \tag{4.19}$$

We make the following remarks concerning Theorem **4.3** and Theorem **4.4**. First, the endogeneity parameter b is local-to-zero, and for $b_0 \in \mathcal{N}(\Sigma_W)$ the AR and K tests are asymptotically valid. However, unlike the AR test, note that the validity of the K test is established under an additional rank assumption (the case where this additional rank assumption fails is not covered in this paper). So, when $b_0 \in \mathcal{N}(\Sigma_W)$, the inference with locally exogenous instruments using the AR and K tests is feasible (at least in large samples). Second, if $b_0 \notin \mathcal{N}(\Sigma_W)$, the results in both theorems are different from those of Theorems **4.1** and **4.2** because the limiting distributions of both statistics do not diverge. Third, even though the AR and K statistics have non-central chi-square limiting distributions when $b_0 \notin \mathcal{N}(\Sigma_W)$, they are not pivotal since the non-centrality parameters depend on nuisance parameters. In addition, the limiting distributions of both statistics cannot be bounded by any pivotal distribution.

It will be useful to see how the above theorems apply in a simple example.

Example 4.1 Consider again model (2.29)-(2.30), which involves one endogenous explanatory variable and two instruments. If the matrix Σ_W is invertible, then $\mathcal{N}(\Sigma_W) = \{0\}$, and Theorem **4.1** entails that $AR(\beta_0) \stackrel{L}{\to} +\infty$ under the null hypothesis $\beta = \beta_0$. Similarly, if $\tilde{\Sigma}_{XY} \neq 0$, then $\text{rank}(\tilde{\Sigma}_{XY}) = G = 1$ and Theorem **4.2** entails that $K(\beta_0) \stackrel{L}{\to} +\infty$ when $\beta = \beta_0$. If X_1 is a valid instrument (with $W_1 = 0$) and $X_2 = W_2$ with $W_2'W_2/T \stackrel{p}{\to} \sigma^2_{W_2} > 0$, we have

$$\Sigma_W = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2_{W_2} \end{bmatrix} \tag{4.20}$$

which is a matrix of rank one, and $\mathcal{N}(\Sigma_W) = \{(x_1, x_2)' : x_2 = 0\}$. If $b_2 = 0$, then $b_0 \in \mathcal{N}(\Sigma_W)$ and Theorem **4.1** entails that the asymptotic distribution given by (4.2) holds for $AR(\beta_0)$, while for $K(\beta_0)$ part B of Theorem **4.2** is applicable. Of course, when $b_0 = 0$, $AR(\beta_0)$ follows the usual $\chi^2(2)/2$ asymptotic distribution, while $K(\beta_0)$ follows a $\chi^2(1)$ distribution. For locally exogenous instruments, theorems **4.3** and **4.4** can be applied in a similar way.

5. Conclusion

In this paper, we have established conditions under which the AR and K tests are asymptotically valid even if some instruments used are endogenous. We have also showed that when these conditions fail, the limiting distributions of both statistics may diverge. Furthermore, when these conditions fail, under locally exogenous instruments setup, the limiting distributions of the statistics depend on nuisance parameters and cannot be bounded by any pivotal distribution. In consequence, the weak-instrument procedure proposed by Wang and Zivot (1998), the unified weak instruments framework of Swanson and Chao (2005) and the inference with imperfect instruments suggested by Ashley (2006) are not applicable. Overall, our results underscore the importance of checking for the presence of possibly invalid instruments when applying "identification-robust" tests. They also suggest that sensitivity analyses where different sets of instruments are considered [Ashley (2006), Small (2007)] can be quite useful for the interpretation of empirical results based on instrumental variables.

A. Appendix: Proofs

PROOF OF THEOREM **4.1** Note first that

$$\frac{(y-Y\beta_0)'M_X(y-Y\beta_0)}{T-k} = \frac{u'u}{T-k} - \frac{T}{T-k}\left(\frac{u'X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{T}\right) \tag{A.1}$$

where, by the assumptions (2.3)-(2.18),

$$\frac{u'u}{T-k} \xrightarrow{p} \sigma_u^2 > 0\,,\ \frac{X'X}{T} \xrightarrow{p} \Sigma_X > 0\,,\ \frac{X'u}{T} = \frac{X_0'u}{T} + \frac{W'W}{T}b_0 + \frac{W'e}{T} \xrightarrow{p} \Sigma_W b_0\,, \tag{A.2}$$

$$\left(\frac{u'X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{T}\right) \xrightarrow{p} b_0'\Sigma_W\Sigma_X^{-1}\Sigma_W b_0\,, \tag{A.3}$$

$$\frac{(y-Y\beta_0)'M_X(y-Y\beta_0)}{T-k} \xrightarrow{p} \bar{\sigma}_u^2 = \sigma_u^2 - b_0'\Sigma_W\Sigma_X^{-1}\Sigma_W b_0 \geq 0\,. \tag{A.4}$$

(A) Suppose now that $b_0 \notin \mathcal{N}(\Sigma_W)$. Then $b_0'\Sigma_W\Sigma_X^{-1}\Sigma_W b_0 > 0$ and the numerator of the AR statistic diverges:

$$(y-Y\beta_0)'P_X(y-Y\beta_0) = T\left(\frac{u'X}{T}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{T}\right) \xrightarrow{L} +\infty\,, \tag{A.5}$$

hence

$$AR(\beta_0) \xrightarrow{L} +\infty\,. \tag{A.6}$$

(B) If $b_0 \in \mathcal{N}(\Sigma_W)$, we have $\Sigma_W b_0 = 0$ and $\bar{\sigma}_u^2 = \sigma_u^2$. Further,

$$X'u = X'(e + Wb_0) = X'e + X'Wb_0\,, \tag{A.7}$$

$$\frac{1}{\sqrt{T}}X'u = \frac{1}{\sqrt{T}}[X'u - \Sigma_W b_0] = \frac{1}{\sqrt{T}}X'e + \frac{1}{\sqrt{T}}(X'W - \Sigma_W)b_0 \xrightarrow{L} S = S_e + S_b\,. \tag{A.8}$$

Then,

$$(y-Y\beta_0)'P_X(y-Y\beta_0) = \left(\frac{u'X}{\sqrt{T}}\right)\left(\frac{X'X}{T}\right)^{-1}\left(\frac{X'u}{\sqrt{T}}\right) \xrightarrow{L} S'\Sigma_X^{-1}S\,, \tag{A.9}$$

$$\frac{(y-Y\beta_0)'M_X(y-Y\beta_0)}{T-k} \xrightarrow{p} \sigma_u^2\,, \tag{A.10}$$

hence

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k\sigma_u^2} S' \Sigma_X^{-1} S\,. \tag{A.11}$$

(C) Finally, if $b_0 = 0$, we have $b_0 \in \mathcal{N}(\Sigma_W)$, with the extra restrictions $u = e$, $\sigma_u^2 = \sigma_e^2$,

$$S = \frac{1}{\sqrt{T}} X'u = \frac{1}{\sqrt{T}} X'e \xrightarrow{L} N\left[0,\, \sigma_e^2 \Sigma_X\right]\,,$$

hence

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k\sigma_e^2} S_e' \Sigma_X^{-1} S_e \sim \frac{1}{k}\chi^2(k)\,. \tag{A.12}$$

□

PROOF OF THEOREM **4.2** We note first, as in (A.1)-(A.4), that

$$S_{uu}(\beta_0) = \frac{(y - Y\beta_0)' M_X (y - Y\beta_0)}{T - k} \xrightarrow{p} \bar{\sigma}_u^2\,, \ \frac{X'X}{T} \xrightarrow{p} \Sigma_X > 0\,, \ \frac{X'u}{T} \xrightarrow{p} \Sigma_W b_0\,. \tag{A.13}$$

(A) Suppose that $b_0 \notin \mathcal{N}(\Sigma_W)$. (i) Let $\Pi = \Pi_0 \neq 0$. Then, we have

$$S_{uV}(\beta_0) = \frac{1}{T-k}(y - Y\beta_0)' M_X Y \xrightarrow{p} q_{uV} = \delta' - b_0' \Sigma_W \Sigma_X^{-1} \Sigma_{WV}\,, \tag{A.14}$$

$$\tilde{\Pi}(\beta_0) = \left(\frac{X'X}{T}\right)^{-1} \frac{X'Y}{T} - \left(\frac{X'X}{T}\right)^{-1} \frac{X'u}{T} \frac{S_{uV}(\beta_0)}{S_{uu}(\beta_0)} \xrightarrow{p} \Sigma_X^{-1} \tilde{\Sigma}_{XY}, \tag{A.15}$$

where $\tilde{\Sigma}_{XY} = \Sigma_{XY} - \Sigma_W b_0 (q_{uV}/\bar{\sigma}_u^2)$, and

$$\frac{\tilde{Y}(\beta_0)'u}{T} = \tilde{\Pi}(\beta_0)' \frac{X'u}{T} \xrightarrow{p} \tilde{\Sigma}_{XY}' \Sigma_X^{-1} \Sigma_W b_0\,, \tag{A.16}$$

$$\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{p} \tilde{\Sigma}_{XY}' \Sigma_X^{-1} \tilde{\Sigma}_{XY}\,. \tag{A.17}$$

If $\operatorname{rank}(\tilde{\Sigma}_{XY}) = G$, then $\tilde{\Sigma}_{XY}' \Sigma_X^{-1} \tilde{\Sigma}_{XY} > 0$ and $\Sigma_X^{-1} \tilde{\Sigma}_{XY} \Sigma_W b_0 \neq 0$ for $b_0 \notin \mathcal{N}(\Sigma_W)$, hence

$$\frac{u'\tilde{Y}(\beta_0)}{T} \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T}\right]^{-1} \frac{\tilde{Y}(\beta_0)'u}{T} \xrightarrow{p} b_0' \Sigma_W \Sigma_X^{-1} \tilde{\Sigma}_{XY} (\tilde{\Sigma}_{XY}' \Sigma_X^{-1} \tilde{\Sigma}_{XY})^{-1} \tilde{\Sigma}_{XY}' \Sigma_X^{-1} \Sigma_W b_0 > 0\,.$$

Consequently, the numerator of the K statistic diverges:

$$(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0) = T \frac{u'\tilde{Y}(\beta_0)}{T} \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \right]^{-1} \frac{\tilde{Y}(\beta_0)'u}{T} \xrightarrow{p} +\infty \tag{A.18}$$

and

$$K(\beta_0) \xrightarrow{L} +\infty\,. \tag{A.19}$$

(ii) Let $\Pi = \Pi_0/\sqrt{T}$. Then

$$(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0) = T \frac{u'\tilde{Y}(\beta_0)}{T} \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \right]^{-1} \frac{\tilde{Y}(\beta_0)'u}{T}\,, \tag{A.20}$$

where

$$\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{L} \Sigma_{XY}^{*\prime} \Sigma_X^{-1} \Sigma_{XY}^{*}\,,\ \frac{\tilde{Y}(\beta_0)'u}{T} \xrightarrow{p} \Sigma_X^{-1} \Sigma_{XY}^{*} \Sigma_W b_0,$$

with $\Sigma_{XY}^{*} = \Sigma_{WV} - \Sigma_W b_0 (q_{uV}/\bar{\sigma}_u^2)$. If $\mathrm{rank}(\Sigma_{XY}^{*}) = G$, then the numerator of the K statistic diverges, and $K(\beta_0) \xrightarrow{L} +\infty$.

(B) If $b_0 \in \mathcal{N}(\Sigma_W)$, we have $\Sigma_W b_0 = 0$, $\bar{\sigma}_u^2 = \sigma_u^2$ and $\frac{1}{\sqrt{T}} X'u \xrightarrow{L} S = S_e + S_b$ as in (A.7)-(A.8).

(i) If $\Pi = \Pi_0 \neq 0$, we have when $b_0 \in \mathcal{N}(\Sigma_W)$, the denominator of the K statistic satisfies

$$\frac{1}{T}(y - Y\beta_0)' M_X (y - Y\beta_0) \xrightarrow{p} \sigma_u^2\,, \tag{A.21}$$

while the denominator can be written

$$(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0) = \frac{u'X}{\sqrt{T}} \tilde{\Pi}(\beta_0) \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \right]^{-1} \tilde{\Pi}(\beta_0)' \frac{X'u}{\sqrt{T}} \tag{A.22}$$

where

$$\tilde{\Pi}(\beta_0) \xrightarrow{p} \Sigma_X^{-1} \Sigma_{XY}\,, \quad \frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{p} \Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY}\,, \quad \frac{\tilde{Y}(\beta_0)'u}{\sqrt{T}} \xrightarrow{p} \Sigma_X^{-1} \Sigma_{XY} S\,. \tag{A.23}$$

If $\mathrm{rank}(\Sigma_{XY}) = G$, we have $\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY} > 0$, hence

$$K(\beta_0) \xrightarrow{L} \frac{1}{\sigma_u^2} S' \Sigma_X^{-1} \Sigma_{XY} (\Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY})^{-1} \Sigma_{XY}' \Sigma_X^{-1} S\,. \tag{A.24}$$

(ii) If $\Pi = \Pi_0/\sqrt{T}$, the numerator of the K statistic is

$$(y - Y\beta_0)' P_{\tilde{Y}(\beta_0)} (y - Y\beta_0) = \frac{u'X}{\sqrt{T}} \tilde{\Pi}(\beta_0) \left[\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T}\right]^{-1} \tilde{\Pi}(\beta_0)' \frac{X'u}{\sqrt{T}}\,, \tag{A.25}$$

hence

$$\tilde{\Pi}(\beta_0) \xrightarrow{p} \Sigma_X^{-1}\Sigma_{WV}\,, \quad \frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{p} \Sigma_{WV}'\Sigma_X^{-1}\Sigma_{WV}\,, \quad \frac{\tilde{Y}(\beta_0)'u}{\sqrt{T}} \xrightarrow{p} \Sigma_X^{-1}\Sigma_{WV}S\,. \tag{A.26}$$

If $\mathrm{rank}(\Sigma_{WV}) = G$, then

$$K(\beta_0) \xrightarrow{L} \frac{1}{\sigma_u^2} S'\Sigma_X^{-1}\Sigma_{WV}(\Sigma_{WV}'\Sigma_X^{-1}\Sigma_{WV})^{-1}\Sigma_{WV}'\Sigma_X^{-1}S\,. \tag{A.27}$$

(C) Finally, if $b_0 = 0$, we have $b_0 \in \mathcal{N}(\Sigma_W)$, with the extra restrictions $u = e$, $\sigma_u^2 = \sigma_e^2$,

$$S = \frac{1}{\sqrt{T}} X'u \xrightarrow{L} N\left[0,\, \sigma_e^2\Sigma_X\right]\,,$$

hence, if $\Pi = \Pi_0 \neq 0$,

$$K(\beta_0) \xrightarrow{L} \frac{1}{\sigma_e^2} S_e'\Sigma_X^{-1}\Sigma_{XY}(\Sigma_{XY}'\Sigma_X^{-1}\Sigma_{XY})^{-1}\Sigma_{XY}'\Sigma_X^{-1}S_e \sim \chi^2(G)\,, \tag{A.28}$$

and if $\Pi = \Pi_0/\sqrt{T}$ (where $\Pi_0 = 0$ is allowed),

$$K(\beta_0) \xrightarrow{L} \frac{1}{\sigma_e^2} S_e'\Sigma_X^{-1}\Sigma_{WV}(\Sigma_{WV}'\Sigma_X^{-1}\Sigma_{WV})^{-1}\Sigma_{WV}'\Sigma_X^{-1}S_e \sim \chi^2(G)\,. \tag{A.29}$$

□

PROOF OF THEOREM **4.3** Since b is now local-to-zero, we have

$$\frac{X'u}{\sqrt{T}} \xrightarrow{L} S_e + \Sigma_W b_0\,, \ \frac{X'X}{T} \xrightarrow{p} \Sigma_X\,, \ \frac{X'u}{T} \xrightarrow{p} 0\,, \ \frac{(y - Y\beta_0)'M_X(y - Y\beta_0)}{T-k} \xrightarrow{p} \sigma_u^2 > 0\,. \tag{A.30}$$

Further, we have

$$\frac{u'u}{T-k} = \frac{(e + W\frac{b_0}{\sqrt{T}})'(e + W\frac{b_0}{\sqrt{T}})}{T-k}$$

$$= \quad \frac{e'e}{T-k} + \frac{b_0'W'e}{\sqrt{T}(T-k)} + \frac{e'Wb_0}{\sqrt{T}(T-k)} + \frac{b_0'e'Wb_0}{T(T-k)} \xrightarrow{p} \sigma_e^2 = \sigma_u^2 \,. \qquad \text{(A.31)}$$

(A) Let $b_0 \notin \mathcal{N}(\Sigma_W)$. Then,

$$AR(\beta_0) \xrightarrow{L} \frac{1}{k\sigma_e^2}(S_e + \Sigma_W b_0)'\Sigma_X^{-1}(S_e + \Sigma_W b_0) \sim \frac{1}{k}\chi^2(k\,, \mu_1) \qquad \text{(A.32)}$$

where $\mu_1 = \frac{1}{\sigma_e^2} b_0' \Sigma_W \Sigma_X^{-1} \Sigma_W b_0 \neq 0$. Similarly, we have $\frac{\tilde{Y}(\beta_0)'\tilde{Y}(\beta_0)}{T} \xrightarrow{p} \Sigma_{XY}' \Sigma_X^{-1} \Sigma_{XY}$ and $\frac{\tilde{Y}(\beta_0)'u}{\sqrt{T}} \xrightarrow{L} \Sigma_X^{-1} \Sigma_{XY}(S_e + \Sigma_W b_0)$. So, if $\mathrm{rank}(\Sigma_{XY}) = G$, we have

$$K(\beta_0) \xrightarrow{L} \frac{1}{\sigma_e^2}(S_e + \Sigma_W b_0)'\Sigma_X^{-1}\Sigma_{XY}(\Sigma_{XY}'\Sigma_X^{-1}\Sigma_{XY})^{-1}\Sigma_{XY}'\Sigma_X^{-1}(S_e + \Sigma_W b_0) \sim \chi^2(G\,, m'm) \qquad \text{(A.33)}$$

where $m = \frac{1}{\sigma_e}(\Sigma_{XY}'\Sigma_X^{-1}\Sigma_{XY})^{-1/2}\Sigma_{XY}'\Sigma_X^{-1}\Sigma_W b_0 \neq 0$.

(B) If $b_0 \in \mathcal{N}(\Sigma_W)$, we have $\Sigma_W b_0 = 0$. Then $\mu_1 = 0$ and $m = 0$, hence $AR(\beta_0) \xrightarrow{L} \frac{1}{k}\chi^2(k)$ and $K(\beta_0) \xrightarrow{L} \chi^2(G)$. □

PROOF OF THEOREM **4.4** The proof of Theorem **4.3** for the AR statistic covers Theorem **4.4**. The proof for the K statistic is similar to the one in Theorem **4.3**. □

Chapter 2

Exogeneity tests, non Gaussian distributions and weak identification : finite-sample and asymptotic distributional theory

1. Introduction

A basic problem in econometrics is estimating an equation of the form

$$y = X\beta + u \tag{1.1}$$

where the explanatory variables X and the errors u might be correlated. In order to make corrections for correlation between explanatory variables and disturbances, a common practice consists in applying an exogeneity test, usually by resorting to instrumental variable (IV) methods. Exogeneity tests of the type proposed by Durbin (1954), Wu (1973), Hausman (1978), Revankar and Hartley (1973) are often used for this purpose. However, such tests rely on the assumption that model parameters are identified by the available instruments. So, an interesting question is how do standard exogeneity tests behave when the instruments are weak?

In a recent paper, Hahn, Ham and Moon (2008) consider the problem of testing the exogeneity of a subset of excluded IV using Durbin-Wu-Hausman-type tests. By referring to Theorem 4 in the Appendix C.1, the authors conclude in Section 5 that standard Hausman pre-tests [$\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$] are not valid in presence of weak instruments and propose a modified version which does not exhibit this problem. With a close look of this, it is likely that the conclusions of Hahn et al. (2008) underline the non validity of DWH-type procedures for partial exogeneity hypotheses (*i.e.* DWH-tests are unusable for testing the exogeneity of a subset of variables) [Doko and Dufour (2009c), Doko and Dufour (2009b)]. It is not clear from Hahn et al. (2008), how behave DWH-type tests in presence of weak IV when testing the exogeneity of endogenous explanatory variables. More precisely, are DWH-type tests robust to weak IV when testing the exogeneity of (possibly) included endogenous regressors?

Moreover, Guggenberger (2008) shows that the two-stage t-tests, where DWH-type tests [including the modified version in Hahn et al. (2008)] are used in the first stage as a pre-test, are unreliable from the view point of size control when IV are weak. This suggests that only identification-robust procedures [Anderson and Rubin (1949, AR-test), Kleibergen (2002, K-test), Moreira (2003, CLR-test), projection-based techniques, see Dufour (1997, 2003), Dufour (2005, 2006), split-sample methods, see Dufour and Jasiak (2001)] should be used and the practice of pre-testing for explanatory variable exogeneity be abandoned.

In this paper, we argue that this type of conclusions may go too far. First, Guggenberger (2008) paper focuses on testing and does not explore the problem in the viewpoint of estimation. An interesting question is: are usual IV estimators more efficient than pretest-estimators based on DWH-tests when IV are weak? Second, the conclusions of Guggenberger (2008) are based on the weak instruments asymptotic setup as in Staiger and Stock (1997). However, this framework assumes that the reduced form parameters which control instrument weakness approach zero (at rate $T^{-1/2}$) as the sample size increases. Clearly, this framework assumes that all structural parameters are not asymptotically identified (all IV are weak). The question now is what happens to DWH-type tests (in finite-and large-sample) when model parameters are partially identified? In particular, what happens to these tests if at least one instrument is strong?

In this paper, we propose a finite-and large-sample analysis of the distribution of DWH tests under the null hypothesis (level) and the alternative hypothesis (power), including when identification is deficient or weak (weak instruments). Our finite-sample analysis provides several new insights and extensions of earlier procedures. The characterization of the finite-sample distribution of DWH-tests statistics allows the construction of exact identification-robust exogeneity tests even with non-Gaussian errors [Monte Carlos exogeneity (MCE) tests]. This characterization also shows that DWH-tests are typically robust to weak instruments (level is controlled). Thus, the conclusions of Hahn et al. (2008) is inaccurate. Furthermore, we provide a characterization of the power of the tests, which clearly exhibits factors which determine power. We show that DWH tests have no power when all instruments are weak [similar to Guggenberger (2008)]. But power may exist as soon as we have one strong instruments (partial identification). The conclusions of Guggenberger

(2008) focus on the case where all instruments are weak, a case of little practical interest. Our asymptotic distributional theory under weaker assumptions confirms the finite-sample theory.

We present simulation evidence indicating that: (1) over a wide range cases, including weak IV and moderate endogeneity, OLS performs better than 2SLS [finding similar to Kiviet and Niemczyk (2007)]; (2) pretest-estimators based on exogeneity tests have an excellent overall performance. Hence, the conclusions of Guggenberger (2008) may be misleading.

We illustrate our theoretical results through two empirical applications: the relation between trade and economic growth [see, Dufour and Taamouti (2006), Irwin and Tervio (2002), Frankel and Romer (1999), Harrison (1996), Mankiw and al. (1992)] and the widely studied problem of returns to education [Dufour and Taamouti (2006), Angrist and Krueger (1991), Angrist and Krueger (1995), Angrist and al. (1999), Mankiw and al. (1992)].

The paper is organized as follows. Section 2 formulates the model studied. Section 3 describes the statistics. Section 4 studies the finite-sample properties of the tests with (possibly) weak instruments. Section 5 presents the Monte Carlo exogeneity (MCE) tests while Section 6 explores the asymptotic behaviour of the test statistics. Section 7 presents a simulation experiment and Section 8 illustrates our theoretical results through two important applications. We conclude in Section 9 and proofs are presented in the Appendix.

2. Model

We consider the following standard simultaneous equations framework:

$$y = Y\beta + Z_1\gamma + u\,, \tag{2.1}$$

$$Y = Z_1\Pi_1 + Z_2\Pi_2 + V\,, \tag{2.2}$$

where $y \in R^T$ is a vector of observations on a dependent variable, $Y \in R^{T\times G}$ is a matrix of observations on (possibly) endogenous explanatory variables $(G \geq 1)$, $Z_1 \in R^{T\times k_1}$ is a matrix of observations on exogenous variables included in the structural equation of interest (2.1), $Z_2 \in R^{T\times k_2}$ is a matrix of observations on the exogenous variables excluded from the structural equation, $u = (u_1, \ldots, u_T)' \in R^T$ and $V = [V_1, \ldots, V_T]' \in R^{T\times G}$ are disturbance matrices with mean zero, $\beta \in R^G$ and $\gamma \in R^{k_1}$ are vectors of unknown coefficients, $\Pi_1 \in R^{k_1\times G}$ and $\Pi_2 \in R^{k_2\times G}$ are matrices of unknown coefficients. We suppose that the "instrument matrix"

$$Z = [Z_1 : Z_2] \in \mathbb{R}^{T\times k} \text{ has full-column rank} \tag{2.3}$$

where $k = k_1 + k_2$ and

$$T - k_1 - k_2 > G\,, \quad k_2 \geq G\,. \tag{2.4}$$

The usual necessary and sufficient condition for identification of this model is $rank(\Pi_2) = G\,.$

The reduced form for $[y, Y]$ can be written as

$$y = Z_1\pi_1 + Z_2\pi_2 + v\,, \tag{2.5}$$

$$Y = Z_1\Pi_1 + Z_2\Pi_2 + V\,, \tag{2.6}$$

where $\pi_1 = \gamma + \Pi_1\beta,\ \pi_2 = \Pi_2\beta,$ and $v = u + V\beta = [v_1, \ldots, v_T]'$. Let

$$M = M_Z = I - Z(Z'Z)^{-1}Z', \quad M_1 = M_{Z_1} = I - Z_1(Z_1'Z_1)^{-1}Z_1'. \tag{2.7}$$

Then, we have

$$M_1 - M = M_1Z_2(Z_2'M_1Z_2)^{-1}Z_2'M_1. \tag{2.8}$$

We now study the problem of testing the exogeneity of Y in model (2.1) - (2.2).

3. Exogeneity test statistics

We consider Durbin-Wu-Hausman (DWH) test statistics, namely three versions of Hausman-type statistics $[\mathrm{H}_i, i = 1, 2, 3]$, the four statistics proposed by Wu (1973) $[\mathcal{T}_l, l = 1, 2, 3, 4]$ and the test statistic proposed by Revankar and Hartley (1973, RH). First, we propose a unified presentation of DWH test statistics. And second, we provide the regression interpretation of all above statistics (including RH test statistic).

3.1. Unified presentation

This subsection proposes a unified presentation of DWH test statistics. The proof of this representation is attached in Appendix A-1.1. The four statistics proposed by Wu (1973) are given by

$$\mathcal{T}_l = \kappa_l(\tilde{\beta} - \hat{\beta})'\tilde{\Sigma}_l^{-1}(\tilde{\beta} - \hat{\beta}), \quad l = 1, 2, 3, 4; \tag{3.1}$$

the three versions of Hausman-type statistics are defined as

$$\mathcal{H}_i = T(\tilde{\beta} - \hat{\beta})'\hat{\Sigma}_i^{-1}(\tilde{\beta} - \hat{\beta}), \quad i = 1, 2, 3, \tag{3.2}$$

and the Revankar and Hartley (1973, RH) statistic is given by

$$\mathcal{RH} = \kappa_R y'\hat{\Sigma}_R y, \tag{3.3}$$

where $\hat{\beta} = (Y'M_1Y)^{-1}Y'M_1y$ is the ordinary least squares (OLS) estimator of β, $\tilde{\beta} = [Y'(M_1 - M)Y]^{-1}Y'(M_1 - M)y$ is the two-stage least squares (2SLS) estimator of β,

$$\tilde{\Sigma}_1 = \tilde{\sigma}_1^2\hat{\Delta}, \quad \tilde{\Sigma}_2 = \tilde{\sigma}_2^2\hat{\Delta}, \quad \tilde{\Sigma}_3 = \tilde{\sigma}^2\hat{\Delta}, \quad \tilde{\Sigma}_4 = \hat{\sigma}^2\hat{\Delta}, \tag{3.4}$$

$$\hat{\Sigma}_1 = \tilde{\sigma}^2\hat{\Omega}_{IV}^{-1} - \hat{\sigma}^2\hat{\Omega}_{LS}^{-1}, \quad \hat{\Sigma}_2 = \tilde{\sigma}^2\hat{\Delta}, \quad \hat{\Sigma}_3 = \hat{\sigma}^2\hat{\Delta}, \tag{3.5}$$

$$\hat{\Sigma}_R = \frac{1}{\hat{\sigma}_R^2}D_1Z_2(Z_2'D_1Z_2)^{-1}Z_2'D_1, \tag{3.6}$$

$$\hat{\Omega}_{IV} = \frac{1}{T}Y'(M_1 - M)Y, \quad \hat{\Omega}_{LS} = \frac{1}{T}Y'M_1Y, \tag{3.7}$$

$$\hat{\Delta} = \hat{\Omega}_{IV}^{-1} - \hat{\Omega}_{LS}^{-1}, \; D_1 = \frac{1}{T}M_1M_{M_1Y}, \tag{3.8}$$

$$\tilde{\sigma}^2 = (y - Y\tilde{\beta})'M_1(y - Y\tilde{\beta})/T, \quad \hat{\sigma}^2 = (y - Y\hat{\beta})'M_1(y - Y\hat{\beta})/T, \tag{3.9}$$

$$\tilde{\sigma}_1^2 = (y - Y\tilde{\beta})'(M_1 - M)(y - Y\tilde{\beta})/T = \tilde{\sigma}^2 - \tilde{\sigma}_e^2\,, \tag{3.10}$$

$$\tilde{\sigma}_2^2 = \hat{\sigma}^2 - (\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = \hat{\sigma}^2 - \tilde{\sigma}^2(\tilde{\beta} - \hat{\beta})'\hat{\Sigma}_2^{-1}(\tilde{\beta} - \hat{\beta})\,, \tag{3.11}$$

$$\tilde{\sigma}_e^2 = (y - Y\tilde{\beta})'M(y - Y\tilde{\beta})/T,\ \hat{\sigma}_R^2 = yM_{\bar{X}}y'/T\,, \tag{3.12}$$

$$M_{M_1Y} = I - M_1Y(Y'M_1Y)^{-1}Y'M_1\,, \tag{3.13}$$

and $\kappa_1 = (k_2 - G)/G\,,\ \kappa_2 = (T - k_1 - 2G)/G\,,\ \kappa_3 = \kappa_4 = T - k_1 - G\,, \kappa_R = (T - k_1 - k_2 - G)/k_2$. The corresponding tests reject H_0 when the test statistic is "large".

In the above definitions, $\hat{\sigma}^2$ is the OLS-based estimator of σ_u^2, $\tilde{\sigma}^2$ is the usual 2SLS-based estimator of σ_u^2 (both without correction for degrees of freedom), while $\tilde{\sigma}_1^2$, $\tilde{\sigma}_2^2$ and $\hat{\sigma}_R^2$ may be interpreted as alternative IV-based scaling factors. $\hat{\sigma}^2$ is consistent when Y is exogenous while $\tilde{\sigma}^2$ is consistent with strong instruments irrespective of whether Y is exogenous or not. Apart from $\mathcal{RH}$, the other statistics are based on comparing the OLS and 2SLS estimators of β. They differ through the use of different "covariance matrices". $\mathcal{H}_1$ uses two different estimators of σ_u^2, while the other statistics resort to a single scaling factor (or estimator of σ_u^2). The expressions of the Durbin-Wu test statistics in (3.1) are different from those in Wu (1973, Econometrica). The link between Wu (1973) notations and ours is established in Appendix 1.1. We use the above notations to better see the link between Hausman-type tests and the Wu tests. In particular, it is easy to see that $\tilde{\Sigma}_3 = \hat{\Sigma}_2$ and $\tilde{\Sigma}_4 = \hat{\Sigma}_3$, so $\mathcal{T}_3 = (\kappa_3/T)\mathcal{H}_2$ and $\mathcal{T}_4 = (\kappa_4/T)\mathcal{H}_3$. Since $\kappa_3/T = \kappa_4/T \to 1$ as $T \to +\infty$, $\mathcal{T}_3$ is asymptotically equivalent with $\mathcal{H}_2$, and $\mathcal{T}_4$ is asymptotically equivalent with $\mathcal{H}_3$.

Finite-sample distributions are available for $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{RH}$ (under a Gaussian distributional assumption), while $\mathcal{T}_3$ and $\mathcal{T}_4$ can be interpreted as asymptotically justified modifications of $\mathcal{T}_1$ and $\mathcal{T}_2$. More precisely, if $u \sim N[0,\, \sigma^2 I_T]$ and Z is independent of u, then

$$\mathcal{T}_1 \sim F(G,\, k_2 - G)\,, \quad \mathcal{T}_2 \sim F(G,\, T - k_1 - 2G)\,, \quad \mathcal{RH} \sim F(k_2,\, T - k_1 - k_2 - G) \tag{3.14}$$

under H_0. If furthermore, $rank(\Pi_2) = G$ and the sample size is large, we have (with standard regularity conditions): under H_0,

$$\mathcal{H}_i \xrightarrow{L} \chi^2(G),\ i = 1,\, 2,\, 3\,, \tag{3.15}$$

$$\mathcal{T}_l \xrightarrow{L} \chi^2(G),\ l = 2,\, 3,\, 4. \tag{3.16}$$

The question is now what happens to $\mathcal{H}_i$, $\mathcal{T}_l$ and $\mathcal{RH}$ when $rank(\Pi_2) < G$?

3.2. Regression interpretation

It is interesting to observe that $\mathcal{T}_2$ is the usual F-statistic for testing $a = 0$ in the extended "regression":

$$y = Y\beta + Z_1\gamma + \hat{V}a + e_* = X\theta + e_*\,, \tag{3.17}$$

where $\hat{V} = MY$, $e_* = P_Z V a + \varepsilon$, $P_Z = Z(Z'Z)^{-1}Z'$, $Z = [Z_1 : Z_2]$, $X = [X_1 : \hat{V}]$, $X_1 = [Y : Z_1]$, $\theta = [\beta',\, \gamma',\, a']'$ and ε is independent of V with mean zero and variance σ_ε^2 ; see Dufour (1987, eqs. (3.1)-(3.2) and (4.2)). Since $Y = \hat{Y} + \hat{V}$, (3.17) can also be written as

$$y = \hat{Y}\beta + Z_1\gamma + \hat{V}b + e_* = \hat{X}\theta_* + e_*\,, \tag{3.18}$$

where $\hat{X} = [\hat{X}_1 : \hat{V}]$, $\hat{X}_1 = [\hat{Y} : Z_1]$, $\theta_* = [\beta',\, \gamma',\, b']'$, and $b = \beta + a$. We see that $H_0 : a = 0$ can be assessed by testing $H_b : \beta = b$ in (3.18). Let define

$\hat{\theta}$: the OLS estimate of θ in (3.17), $\hat{\theta}_0$: the restricted OLS estimate of θ under H_0 in (3.17);

$\hat{\theta}_*$: the OLS estimate of θ_* in (3.18), $\hat{\theta}_{*0}$: the restricted OLS estimate of θ_* under H_b in (3.18);

$\hat{\theta}_*^0$: the restricted OLS estimate of θ_* under $b = 0$ in (3.18) or $\beta = -a$ in (3.17);

$S(\theta) = (y - X\theta)'(y - X\theta),\ S_*(\theta_*) = (y - \hat{X}\theta_*)'(y - \hat{X}\theta_*)\,.$

Then, we have

$$S(\hat{\theta}) \;=\; S_*(\hat{\theta}_*),\; S(\hat{\theta}_0) = S_*(\hat{\theta}_{*0})\,, \tag{3.19}$$

$$S(\hat{\theta}) \;=\; T\tilde{\sigma}_2^2,\; S(\hat{\theta}_0) = T\hat{\sigma}^2,\; S_*(\hat{\theta}_*^0) = T\tilde{\sigma}^2\,, \tag{3.20}$$

where $\tilde{\sigma}^2, \hat{\sigma}^2, \tilde{\sigma}_2^2$ are defined in (3.9) - (3.13). So, except for $\mathcal{H}_1, \mathcal{H}_i,\; i = 2,\, 3$ and $\mathcal{T}_l,\; l = 1,\, 2,\, 3,\, 4$ can be expressed as [see Appendix A-1.2 for further details]:

$$\mathcal{T}_1 \;=\; \frac{[S(\hat{\theta}_0) - S(\hat{\theta})]/G}{[S_*(\hat{\theta}_*^0) - S_e(\hat{\theta})]/(k_2 - G)}, \quad \mathcal{T}_2 = \frac{[S(\hat{\theta}_0) - S(\hat{\theta})]/G}{S(\hat{\theta})/(T - k_1 - 2G)}\,, \tag{3.21}$$

$$\mathcal{T}_3 \;=\; \frac{S(\hat{\theta}_0) - S(\hat{\theta})}{S_*(\hat{\theta}_*^0)/(T - k_1 - G)}, \quad \mathcal{T}_4 = \frac{S(\hat{\theta}_0) - S(\hat{\theta})}{S(\hat{\theta}_0)/(T - k_1 - G)}\,, \tag{3.22}$$

$$\mathcal{H}_2 \;=\; \frac{S(\hat{\theta}_0) - S(\hat{\theta})}{S_*(\hat{\theta}_*^0)/T}, \quad \mathcal{H}_3 = \frac{S(\hat{\theta}_0) - S(\hat{\theta})}{S(\hat{\theta}_0)/T}\,, \tag{3.23}$$

where $S_*(\hat{\theta}_*^0) - S_e(\hat{\theta}) = Q_1 = T\tilde{\sigma}_1^2,\; S_e(\hat{\theta}) = T\tilde{\sigma}_e^2$ and $\tilde{\sigma}_1^2,\; \tilde{\sigma}_e^2$, are defined in (3.9) - (3.13). Equations (3.23) - (3.22) are the regression formulation of DWH statistics.

We now derive the same expression for $\mathcal{RH}$. Assume that

$$u \;=\; Va + \varepsilon\,, \tag{3.24}$$

where a is a $G \times 1$ vector of unknown coefficients, and ε is independent of V with mean zero and variance σ_ε^2. From (2.2), we have $V = Y - Z_1\Pi_1 - Z_2\Pi_2$ and (3.24) becomes

$$u = Ya - Z_1\Pi_1 a - Z_2\Pi_2 a + \varepsilon\,. \tag{3.25}$$

If we replace (3.25) in (2.1), we get

$$y = Yb + Z_1\bar{\gamma} + Z_2\bar{a} + \varepsilon = \bar{X}\bar{\theta} + \varepsilon\,, \tag{3.26}$$

where $b = a + \beta,\; \bar{\gamma} = \gamma - \Pi_1 a,\; \bar{a} = -\Pi_2 a,\; \bar{\theta} = [b',\, \bar{\gamma}',\, \bar{a}']'$ and $\bar{X} = [X_1\,,\, Z_2] = [Y\,,\, Z]$.

Revankar and Hartley (1973) propose to test the exogeneity of Y (i.e., $H_0 : a = 0$) by testing the hypothesis

$$H_0^* \quad : \quad \bar{a} = -\Pi_2 a = 0\,. \tag{3.27}$$

If $rank(\Pi_2) = G$, $a = 0$ if and only if $\bar{a} = 0$, hence H_0 is equivalent to H_0^*. However, if $rank(\Pi_2) < G$, i.e. if identification is deficient (weak instruments), $\bar{a} = 0$ does not entail that $a = 0$. So, the RH test may not test the exogeneity of Y in the model if the instruments are weak. As we can see from (3.27), H_0^* is the joint hypothesis that all coefficients of $Z_{2j},\ j = 1,\ \ldots,\ k_2$ in (3.26) are zero. Clearly, H_0 is equivalence to H_0^* only when identification is strong: $rank(\Pi_2) = G$ [see Revankar and Hartley (1973)]. Moreover, equation (3.26) illustrates clearly that the endogeneity of the regressors Y may be viewed as a problem of omitted variables [see Dufour (1987)].

Now, define

$\hat{\bar{\theta}}$: the OLS estimate of $\bar{\theta}$ in (3.26), $\hat{\bar{\theta}}_0$: the restricted OLS estimate of $\bar{\theta}$ under H_0^* in (3.26); $\bar{S}(\bar{\theta}) = (y - X\bar{\theta})'(y - X\bar{\theta})$.

Then, the $\mathcal{RH}$ statistic is expressed as

$$\mathcal{RH} = \frac{\kappa_R(\bar{S}(\hat{\bar{\theta}}_0) - \bar{S}(\hat{\bar{\theta}}))}{\bar{S}(\hat{\bar{\theta}}_0)}\,. \tag{3.28}$$

Equation (3.28) is the regression interpretation of $\mathcal{RH}$ statistic.

In addition to Staiger-Stock(1997) weak instruments framework– as in Hahn et al. (2008) and Guggenberger (2008)– this paper proposes a finite-sample analysis of the distribution of DWH and RH tests under the null hypothesis (level) and the alternative hypothesis (power), including when identification is deficient or weak (weak instruments).

4. Finite-sample theory

To obtain finite-sample results on the distributions of the test statistics, we shall consider two setups on the disturbances distributions: the strict exogeneity setup where the structural error is independent of the endogenous regressors and the set of instruments and the second setup where the reduced form errors belong to Cholesky error family.

4.1. Strict exogeneity

The strict exogeneity hypothesis is expressed as

$$H_0 \quad : \quad u \text{ is independent of } [Y, Z] \tag{4.1}$$

vs

$$H_1 \quad : \quad u = Va + \varepsilon\,, \tag{4.2}$$

where a is a $G \times 1$ vector of unknown coefficients, ε is independent of V with mean zero and variance σ_ε^2. The hypothesis H_0 can also be expressed as

$$H_0 \quad : \quad a = 0\,. \tag{4.3}$$

It is important to note that (4.1) - (4.2) does not require any assumption on the functional form of Y. So, we could assume that Y obeys a general model of the form

$$Y \quad = \quad g(Z_1,\, Z_2,\, V,\, \mathit{\Pi})\,, \tag{4.4}$$

where $g(.)$ is a possibly unspecified non-linear function, $\mathit{\Pi}$ is an unknown parameter matrix and V follows an arbitrary distribution. This setup is quite wide and does allow one to study several situations where V does not follow a Gaussian distribution.

We shall now analyze the distributions of DWH and RH statistics under the null hypothesis (level) and the alternative hypothesis (power).

4.1.1. Pivotality under strict exogeneity

As mentioned above, we study here the distribution of exogeneity tests under H_0 without imposing any restriction on instrument strength. Theorem below establishes the pivotality of all statistics, including situations where identification is deficient or weak (weak instruments).

Theorem 4.1 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions* (2.1), (2.3) - (2.4) *and the null hypothesis* (4.1), *the conditional distributions given* $[Y\ ,\ Z_1\ ,\ Z_2]$ *of the statistics defined in* (3.1) - (3.13) *depend only on the distribution of* u/σ_u *irrespective of whether the instruments are strong or weak.*

Theorem **4.1** shows that under strict exogeneity, DWH- and RH-type tests are typically robust to weak instruments (level is controlled) whether the instruments are strong or weak. This pivotality result allows the construction of exact identification-robust exogeneity tests even with non-Gaussian errors [Monte Carlo tests, see Section 5]. The Monte Carlo exogeneity (MCE) tests proposed do not require any restriction on the distribution of V and the functional form of Y. More generally, assumption (2.2) may not hold and one could assume that Y obeys a general non-linear model as defined in (4.4) and that $V_1, \ldots, V_T$ are heteroskedastic.

Section 4.1.2 below focuses on the power of the tests.

4.1.2. Power and large endogeneity

We characterize the distributions of DWH and RH tests under the alternative hypothesis (4.2) with or without weak instruments. Two main results are presented. First, we show that the conditional distributions given Y, $[Z_1, Z_2]$ only depend on the endogeneity parameter a and derive cases where the tests have power even if identification is deficient or weak. Second, we introduce the concept of "large exogeneity" and analyze its effects on the tests. Theorem **4.2** below characterizes the power of the tests.

Theorem 4.2 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Let the assumptions* (2.1) - (2.4) *hold. If furthermore* H_1 *in* (4.2) *is satisfied, then we can write*

$$\mathcal{H}_1 = T(Va+\varepsilon)'(\bar{A}_1 - A_1)'\Sigma_1^{-1}(\bar{A}_1 - A_1)(Va+\varepsilon)\,, \tag{4.5}$$

$$\mathcal{H}_2 = T\frac{(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)}\,, \tag{4.6}$$

$$\mathcal{H}_3 = T\frac{(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'D_1(Va+\varepsilon)}\,, \tag{4.7}$$

$$\mathcal{T}_1 = \frac{\kappa_1(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'(\bar{D}_1 - D_1)(Va+\varepsilon)}\,, \tag{4.8}$$

$$\mathcal{T}_2 = \frac{\kappa_2(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'(D_1 - C_0)(Va+\varepsilon)}\,, \tag{4.9}$$

$$\mathcal{T}_3 = \frac{\kappa_3(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)}\,, \tag{4.10}$$

$$\mathcal{T}_4 = \frac{\kappa_4(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'D_1(Va+\varepsilon)}\,, \tag{4.11}$$

$$\mathcal{RH} = \frac{\kappa_R(Va+\varepsilon)'P_{D_1Z_2}(Va+\varepsilon)}{(Va+\varepsilon)'(D_1 - P_{D_1Z_2})(Va+\varepsilon)}\,, \tag{4.12}$$

where

$$\Sigma_1 = (Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)\hat{\Omega}_{IV}^{-1} - (Va+\varepsilon)'D_1(Va+\varepsilon)\hat{\Omega}_{LS}^{-1}\,,$$

$$C_0 = (\bar{A}_1 - A_1)'\hat{\Delta}^{-1}(\bar{A}_1 - A_1),\ \bar{A}_1 = [Y'(M_1 - M)Y]^{-1}Y'(M_1 - M),$$

$$A_1 = (Y'M_1Y)^{-1}Y'M_1,\ \bar{D}_1 = \frac{1}{T}M_1M_{(M_1-M)Y},\ D_1 = \frac{1}{T}M_1M_{M_1Y},$$

$$P_B = B(B'B)^{-1}B' \text{ and } M_B = I - P_B \text{ for any matrix } B\,,$$

$\hat{\Omega}_{IV},\ \hat{\Omega}_{LS}$ *and* $\hat{\Delta}$ *are defined in (*3.7*)* - *(*3.8*),* $\kappa_R = (T-k-G)/k_2$, *and* $\kappa_l,\ l = 1, 2, 3, 4$, *are defined in* (3.1) - (3.13).

We note first that Theorem **4.2** follows from algebraic arguments only, so Y, $[Z_1, Z_2]$ can be random in any arbitrary way. Second, we remark that given $[Y, Z_1, Z_2]$, the distributions of the statistics only depend on a as nuisance parameter. This characterization of the power of the tests clearly exhibits $(\bar{A}_1 - A_1)Va$, C_0Va, D_1Va, $\bar{D}_1Va$, $P_{D_1Z_2}Va$ as factors which determine power. So, we can observe that the tests have no power if all instruments are weak [similar to Guggenberger (2008)]. This result is proved in Corollary **4.3** below.

Corollary 4.3 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions of Theorem 4.***2***, if* $\mathit{\Pi}_2 = 0$, *then we have* $(\bar{A}_1 - A_1)V = C_0V = D_1V = \bar{D}_1V = P_{D_1Z_2}V = 0$ *so that*

$$\mathcal{H}_1 = T\varepsilon'(\bar{A}_1 - A_1)'\mathit{\Sigma}_{1*}^{-1}(\bar{A}_1 - A_1)\varepsilon\,, \tag{4.13}$$

$$\mathcal{H}_2 = T\frac{\varepsilon' C_0\varepsilon}{\varepsilon'\bar{D}_1\varepsilon}, \quad \mathcal{H}_3 = T\frac{\varepsilon' C_0\varepsilon}{\varepsilon' D_1\varepsilon}\,, \tag{4.14}$$

$$\mathcal{T}_1 = \frac{\kappa_1\varepsilon' C_0\varepsilon}{\varepsilon'(\bar{D}_1 - D_1)\varepsilon}, \quad \mathcal{T}_2 = \frac{\kappa_2\varepsilon' C_0\varepsilon}{\varepsilon'(D_1 - C_0)\varepsilon}\,, \tag{4.15}$$

$$\mathcal{T}_3 = \frac{\kappa_3\varepsilon' C_0\varepsilon}{\varepsilon'\bar{D}_1\varepsilon}, \quad \mathcal{T}_4 = \frac{\kappa_4\varepsilon' C_0\varepsilon}{\varepsilon' D_1\varepsilon}\,, \tag{4.16}$$

$$\mathcal{RH} = \frac{\kappa_R\varepsilon' P_{D_1Z_2}\varepsilon}{\varepsilon'(D_1 - P_{D_1Z_2})\varepsilon}\,, \tag{4.17}$$

where

$$\mathit{\Sigma}_{1*} = \varepsilon'\bar{D}_1\varepsilon\hat{\mathit{\Omega}}_{IV}^{-1} - \varepsilon' D_1\varepsilon\hat{\mathit{\Omega}}_{LS}^{-1}\,.$$

When $\mathit{\Pi}_2 = 0$ (irrelevant instruments), the conditional distributions given $[Y,\, Z_1,\, Z_2]$, of the statistics are the same under the null hypothesis and the alternative hypothesis. This entails that the unconditional distributions are also the same under the null and the alternative hypotheses. Consequently, the power of the tests can not exceed the nominal level [similar to Guggenberger (2008)].

We now introduce the concept of large endogeneity and study its effects on the test statistics. Without loss of generality, less assume that $cov(u,\, V) = \delta$ and $E(V_t'V_t) = \Sigma_V,$ where Σ_V is a fixed positive matrix. From (4.2), we have

$$\Sigma_V a = \delta \;\Leftrightarrow\; a = \Sigma_V^{-1}\delta. \tag{4.18}$$

Let $\|.\|$ be the euclidian norm in R^G. Since Σ_V is fixed (does not depend neither a nor δ), from (4.18), we have

$$\|a\| \;=\; \|\Sigma_V^{-1}\delta\| \tag{4.19}$$

$$\text{and } \|\delta\| \;\to\; +\infty \Leftrightarrow \|a\| \to +\infty. \tag{4.20}$$

When condition (4.20) is satisfied, we say that the endogeneity is large.

The question is how behave DWH-type test statistics in finite-sample when this condition is satisfied.

Let first introduce the following notations and definitions:

$$\mathcal{V}_\lambda[\mathcal{A},\mathcal{B}] \;=\; \left\{\lambda \in \mathbb{R}^+ : |\mathcal{A} - \lambda\mathcal{B}| = 0\right\} \tag{4.21}$$

where A is any positive semidefinite random matrix and B is any positive definite random matrix of dimensions G. Let

$$\lambda_2 \;=\; \max_{\lambda\in\mathbb{R}^+} \mathcal{V}_\lambda[V'C_0V\,,\, V'D_1V]\,, \tag{4.22}$$

$$\lambda_p \;=\; \max_{\lambda\in\mathbb{R}^+} \mathcal{V}_\lambda[V'C_0V\,,\, V'(\bar{D}_1 - D_1)V]\,, \tag{4.23}$$

$$\lambda_p^* \;=\; \max_{\lambda\in\mathbb{R}^+} \mathcal{V}_\lambda[V'P_{D_1Z_2}V\,,\, V'D_1V]\,, \tag{4.24}$$

$$\tilde{\lambda}_1 \;=\; \min_{\lambda\in\mathbb{R}^+} \mathcal{V}_\lambda[V'C_0V\,,\, V'\bar{D}_1V]\,, \tag{4.25}$$

$$\lambda_1^* \;=\; \min_{\lambda\in\mathbb{R}^+} \mathcal{V}_\lambda[V'P_{D_1Z_2}V\,,\, V'D_1V]\,, \tag{4.26}$$

where $C_0,\ D_1,\ \bar{D}_1$, and $P_{D_1 Z_2}$ are defined in Theorem **4.2**. Note that the extremum defined in (4.22) - (4.26) are positive since the matrices in the arguments of $V_\lambda(.,.)$ are all symmetric, positive definite whenever $Va \neq 0$ with probability 1.

Theorem **4.4** below gives the distribution of DWH-tests under condition (4.20).

Theorem 4.4 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Let the assumptions of Theorem* **4.2** *hold and assume that* $Va \neq 0$ *with probability 1.*

(A) *If* $\Pi_2 \neq 0$, *then*

$$0 \quad < \quad \lim_{\|a\|\to\infty} \mathcal{H}_1 \leq \lim_{\|a\|\to\infty} \mathcal{H}_2 \leq \lim_{\|a\|\to\infty} \mathcal{H}_3 \leq T\lambda_2 < +\infty\,, \tag{4.27}$$

$$0 \quad < \quad \tilde{\lambda}_1 \leq \frac{1}{\kappa_3} \lim_{\|a\|\to\infty} \mathcal{T}_3 \leq \frac{1}{\kappa_1} \lim_{\|a\|\to\infty} \mathcal{T}_1 \leq \lambda_p < +\infty\,, \tag{4.28}$$

$$0 \quad < \quad \tilde{\lambda}_1 \leq \frac{1}{\kappa_3} \lim_{\|a\|\to\infty} \mathcal{T}_3 \leq \frac{1}{\kappa_4} \lim_{\|a\|\to\infty} \mathcal{T}_4 \leq \frac{1}{\kappa_2} \lim_{\|a\|\to\infty} \mathcal{T}_2 \leq \frac{\lambda_2}{1+\lambda_2} < +\infty\,, \tag{4.29}$$

$$0 \quad < \quad \frac{\lambda_1^*}{1+\lambda_1^*} \leq \frac{1}{\kappa_R} \lim_{\|a\|\to\infty} \mathcal{RH} \leq \frac{\lambda_p^*}{1+\lambda_p^*} < +\infty\,. \tag{4.30}$$

where λ_2, λ_p, λ_p^*, $\tilde{\lambda}_1$ *and* λ_1^* *are defined in* (4.22) - (4.26).

(B) *If* $\Pi_2 = 0$, *we have*

$$\lim_{\|a\|\to\infty} \mathcal{H}_1 \quad = \quad T\varepsilon'(\bar{A}_1 - A_1)'\Sigma_{1*}^{-1}(\bar{A}_1 - A_1)\varepsilon,\ \lim_{\|a\|\to\infty} \mathcal{H}_2 = T\frac{\varepsilon' C_0 \varepsilon}{\varepsilon' \bar{D}_1 \varepsilon},$$
$$\lim_{\|a\|\to\infty} \mathcal{H}_3 = T\frac{\varepsilon' C_0 \varepsilon}{\varepsilon' D_1 \varepsilon}\,, \tag{4.31}$$

$$\lim_{\|a\|\to\infty} \mathcal{T}_1 \quad = \quad \frac{\kappa_1 \varepsilon' C_0 \varepsilon}{\varepsilon'(\bar{D}_1 - D_1)\varepsilon},\ \lim_{\|a\|\to\infty} \mathcal{T}_2 = \frac{\kappa_2 \varepsilon' C_0 \varepsilon}{\varepsilon'(D_1 - C_0)\varepsilon},\ \lim_{\|a\|\to\infty} \mathcal{T}_3 = \frac{\kappa_3 \varepsilon' C_0 \varepsilon}{\varepsilon' \bar{D}_1 \varepsilon}\,, \tag{4.32}$$

$$\lim_{\|a\|\to\infty} \mathcal{T}_4 \quad = \quad \frac{\kappa_4 \varepsilon' C_0 \varepsilon}{\varepsilon' D_1 \varepsilon}, \quad \lim_{\|a\|\to\infty} \mathcal{RH} = \frac{\kappa_R \varepsilon' P_{D_1 Z_2} \varepsilon}{\varepsilon'(D_1 - P_{D_1 Z_2})\varepsilon}\,, \tag{4.33}$$

*where Σ_{1*}, C_0, D_1, $\bar{D}_1$ and $P_{D_1 Z_2}$ are defined in Theorem 4.**2**.*

We note that the distributions of all tests are finite for fixed T whether the instruments are strong or weak and the endogeneity large. Thus, the power of the tests does not converge with a large endogeneity. Furthermore, when $\Pi_2 = 0$, the distribution of the tests does not involve the endogeneity parameter a. Consequently, the tests have no power. More interestingly, equation (4.27) indicates that $\mathcal{H}_3$ dominates (in term of power) $\mathcal{H}_2$ and $\mathcal{H}_2$ dominates $\mathcal{H}_1$. By the same way, we can see from equations (4.28)-(4.29) that $\mathcal{T}_1$ dominates $\mathcal{T}_3$; $\mathcal{T}_2$ dominates $\mathcal{T}_4$ and $\mathcal{T}_4$ dominates $\mathcal{T}_3$. However, any theoretical power comparison is provided for $\mathcal{T}_1$ and $\mathcal{T}_4$ or $\mathcal{T}_1$ and $\mathcal{T}_2$. Nevertheless, the simulation experiment in Section 7 suggests that $\mathcal{T}_1$ has less power than $\mathcal{T}_2$ and $\mathcal{T}_4$.

We now focus on the Cholesky error families setup.

4.2. Cholesky error families

In this section, we impose more restrictions on the distributions of the errors $[u\,,\,V]$ defined in model (2.1)-(2.2).

Let

$$U = [u,\, V] = [U_1,\, \dots\,,\, U_T]'\,, \tag{4.34}$$

$$W = [v,\, V] = [u + V\beta,\, V] = [W_1,\, W_2,\, \dots\,,\, W_T]'\,. \tag{4.35}$$

We shall assume that the vectors $U_t = [u_t,\, V_t']'$, $t = 1,\, \dots\,,\, T$, have the same nonsingular covariance matrix:

$$E[U_t U_t^{'}] = \Sigma = \begin{bmatrix} \sigma_u^2 & \delta' \\ \delta & \Sigma_V \end{bmatrix} > 0\,, \quad t = 1,\, \dots\,,\, T, \tag{4.36}$$

where Σ_V has dimension G. Then the covariance matrix of the reduced-form disturbances $W_t = [v_t, V_t']'$ errors in (2.5) - (2.6) also have the same covariance matrix, which takes the form:

$$\Omega = \begin{bmatrix} \sigma_u^2 + \beta'\Sigma_V\beta + 2\beta'\delta & \beta'\Sigma_V + \delta' \\ \Sigma_V\beta + \delta & \Sigma_V \end{bmatrix} \tag{4.37}$$

where Ω is positive definite. In this framework, the exogeneity hypothesis can be expressed as

$$H_0 : \delta = 0\,. \tag{4.38}$$

Under (4.2), we have from (4.36)

$$\delta = \Sigma_V a, \quad \sigma_u^2 = \sigma_\varepsilon^2 + a'\Sigma_V a = \sigma_\varepsilon^2 + \delta'\Sigma_V^{-1}\delta\,. \tag{4.39}$$

So, the null hypothesis in (4.38) can be expressed as

$$H_a : a = 0\,. \tag{4.40}$$

Assume that

$$W_t = J\bar{W}_t\,,\ t = 1, \ldots, T\,, \tag{4.41}$$

where the vector $W_{(T)} = vec(\bar{W}_1, \ldots, \bar{W}_T)$ has a known distribution $F_{\bar{W}}$ and $J \in R^{(G+1)\times(G+1)}$ is an unknown upper triangular nonsingular matrix [for a similar assumption in the context of multivariate linear regressions, see Dufour and Khalaf (2002) and Dufour, Khalaf and Beaulieu (2008)].[1] When the errors W_t obeys (4.41), we say that W_t belongs to Cholesky error family.

If the covariance matrix of $\bar{W}_t$ is an identity matrix I_{G+1}, the covariance matrix of W_t is

$$\Omega = E[W_t W_t'] = JJ'. \tag{4.42}$$

In particular, these conditions are satisfied when

$$\bar{W}_t \overset{i.i.d.}{\sim} N[0,\, I_{G+1}]\,,\ t = 1, \ldots, T\,. \tag{4.43}$$

Since the J matrix is upper triangular, its inverse J^{-1} is also upper triangular. Let

$$P = (J^{-1})'. \tag{4.44}$$

[1]In Section 6 below, we consider alternative assumptions to derive an asymptotic distributional theory for the test statistics.

Clearly, P is a $(G+1)\times(G+1)$ lower triangular matrix and it allows one to orthogonalize JJ' :

$$P'JJ'P = I_{G+1}\,, \quad (JJ')^{-1} = PP^{'}\,. \tag{4.45}$$

In (4.45), P' can be interpreted as the Cholesky factor of Ω^{-1}, so P is the unique lower triangular matrix that satisfies equation (4.45); see Harville (1997, Section 14.5, Theorem 14.5.11). We will find useful to consider the following partition of P :

$$P = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix} \tag{4.46}$$

where $P_{11} \neq 0$ is a scalar and P_{22} is a nonsingular $G \times G$ matrix. In particular, if (4.42) holds, we see [using (4.37)] that an appropriate P matrix is obtained by taking:

$$P_{11} \;=\; (\sigma_u^2 - \delta' \Sigma_V^{-1}\delta)^{-1/2} = \sigma_\varepsilon\,, \quad P_{22}'\Sigma_V P_{22} = I_G\,, \tag{4.47}$$

$$P_{21} \;=\; -(\beta + \Sigma_V^{-1}\delta)(\sigma_u^2 - \delta'\Sigma_V^{-1}\delta)^{-1/2} = -(\beta + a)\sigma_\varepsilon^{-1}. \tag{4.48}$$

Further this choice is unique. From (4.48), P_{22} only depends on Σ_V and $P_{11}\beta + P_{21} = -(\Sigma_V^{-1}\delta)\sigma_\varepsilon^{-1} = -a\sigma_\varepsilon^{-1}$. In particular, if $\delta = 0,$ we have $P_{11} = 1/\sigma_u, P_{21} = -\beta/\sigma_u$ and $P_{11}\beta + P_{21} = 0.$

If we postmultiply $[y,\, Y]$ by P, we obtain from (2.5) - (2.6):

$$\left[\bar{y},\, \bar{Y}\right] = [y,\, Y]\, P = [yP_{11} + YP_{21},\, YP_{22}] = [Z_1,\, Z_2] \begin{bmatrix} \gamma + \Pi_1\beta & \Pi_1 \\ \Pi_2\beta & \Pi_2 \end{bmatrix} P + \bar{W} \tag{4.49}$$

where

$$\bar{W} = UP = [\bar{v},\, \bar{V}] = [\bar{W}_1,\, \ldots\,,\, \bar{W}_T]^{'},\; \bar{W}_t = [\bar{v}_t,\, \bar{V}_t^{'}]^{'}, \tag{4.50}$$

$$\bar{v} = vP_{11} + VP_{21} = [\bar{v}_1,\, \ldots\,,\, \bar{v}_T]'\,, \quad \bar{V} = VP_{22} = \left[\bar{V}_1,\, \ldots\,,\, \bar{V}_T\right]'\,, \tag{4.51}$$

Then, we can rewrite (4.49) as

$$\bar{y} \;=\; Z_1(\gamma P_{11} + \Pi_1\zeta) + Z_2\Pi_2\zeta + \bar{v}\,, \tag{4.52}$$

$$\bar{Y} \;=\; Z_1 \Pi_1 P_{22} + Z_2 \Pi_2 P_{22} + \bar{V} \,, \tag{4.53}$$

where

$$\zeta = \beta P_{11} + P_{21} = -(\Sigma_V^{-1}\delta)/(\sigma_u^2 - \delta' \Sigma_V^{-1}\delta)^{1/2} = -a\sigma_\varepsilon^{-1}. \tag{4.54}$$

Since $MZ = 0$,

$$M\bar{y} \;=\; M\bar{v} \,,\; M\bar{Y} = M\bar{V} \,, \tag{4.55}$$

$$M_1\bar{y} \;=\; M_1(\mu_1 + \bar{v}) \,,\; M_1\bar{Y} = M_1(\mu_2 + \bar{V}) \,. \tag{4.56}$$

where

$$\mu_1 = M_1 Z_2 \Pi_2 \zeta = -\sigma_\varepsilon^{-1} M_1 Z_2 \Pi_2 a \,, \quad \mu_2 = M_1 Z_2 \Pi_2 P_{22} \,. \tag{4.57}$$

Clearly, $\zeta = 0 \Leftrightarrow \delta = a = 0$ and $\mu_1 = 0$. This condition holds under H_0 $(\delta = a = 0)$. Furthermore, if $\Pi_2 = 0$ (complete non identification of model parameters), we have $\mu_1 = 0$ and $\mu_2 = 0$, irrespective of the value of δ. In this case,

$$M\bar{y} = M\bar{v} \,,\; M\bar{Y} = M\bar{V} \,,\; M_1\bar{y} = M_1\bar{v} \,,\; M_1\bar{Y} = M_1\bar{V} \,. \tag{4.58}$$

We can now prove the following Cholesky invariance property of all test statistics.

Lemma 4.5 CHOLESKY INVARIANCE OF EXOGENEITY TESTS. *Let*

$$R = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix} \tag{4.59}$$

be a lower triangular matrix such that $R_{11} \neq 0$ *is a scalar and* R_{22} *is a nonsingular* $G \times G$ *matrix. If we replace* y *and* Y *by* $y_* = yR_{11} + YR_{21}$ *and* $Y_* = YR_{22}$ *in* (3.1) - (3.13)*, then the statistics* H_i $(i = 1, 2, 3)$, T_l $(l = 1, 2, 3, 4)$ *and* RH *do not change.*

The above invariance holds irrespective of the choice of lower triangular matrix R. In particular, one can choose $R = P$ as defined in (4.44).

Let introduce the following notations:

$$\Omega_{IV} \equiv \Omega_{IV}(\mu_2, \bar{V}) = (\mu_2 + \bar{V})'(M_1 - M)(\mu_2 + \bar{V})\,, \tag{4.60}$$

$$\Omega_{LS} \equiv \Omega_{LS}(\mu_2, \bar{V}) = (\mu_2 + \bar{V})'M_1(\mu_2 + \bar{V})\,, \tag{4.61}$$

where $\Delta \equiv \Delta(\mu_2, \bar{V})$ is defined in (3.8),

$$\omega_{IV}^2 \equiv \omega_{IV}(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v})^2 = (\mu_1 + \bar{v})'D_*'D_*(\mu_1 + \bar{v})\,, \tag{4.62}$$

$$\omega_{LS}^2 \equiv \omega_{LS}(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v})^2 = (\mu_1 + \bar{v})'C_*(\mu_1 + \bar{v})\,, \tag{4.63}$$

where

$$C_* = M_1 - M_1(\mu_2 + \bar{V})\Omega_{LS}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'M_1$$

$$D_* = M_1 - M_1(\mu_2 + \bar{V})\Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'(M_1 - M),$$

$$\omega_1^2 \equiv \omega_1(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v})^2 = (\mu_1 + \bar{v})'E(\mu_1 + \bar{v})\,, \tag{4.64}$$

$$\omega_2^2 \equiv \omega_2(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v})^2 = (\mu_1 + \bar{v})'[C_* - C'\Delta^{-1}C](\mu_1 + \bar{v})\,, \tag{4.65}$$

$$\omega_R^2 \equiv \omega_R(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v})^2 = (\mu_1 + \bar{v})'[D_1 - P_{D_1 Z_2}](\mu_1 + \bar{v})\,, \tag{4.66}$$

$$C = \Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'(M_1 - M) - \Omega_{LS}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'M_1, \tag{4.67}$$

$$E = (M_1 - M)\big[I - (\mu_2 + \bar{V})\Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'\big](M_1 - M), \tag{4.68}$$

and finally

$$\omega_3^2 \equiv \omega_3(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v})^2 = \omega_{IV}^2,\ \omega_4^2 \equiv \omega_4(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v})^2 = \omega_{LS}^2\,. \tag{4.69}$$

We can now prove the following general theorem on the distributions of the test statistics.

Theorem 4.6 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions* (2.1) - (2.4) *and assumption* (4.41)*, the statistics defined in* (3.1) - (3.13) *have the following representations:*

$$\mathcal{H}_i = T[\mu_1 + \bar{v}]' \Gamma_i(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}] \,, \quad i = 1,\, 2,\, 3,$$

$$\mathcal{T}_l \;=\; \kappa_l[\mu_1 + \bar{v}]' \bar{\Gamma}_l(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}] \,, \quad l = 1,\, 2,\, 3,\, 4,$$

$$\mathcal{RH} \;=\; \kappa_R[\mu_1 + \bar{v}]' \Gamma_R(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}] \,,$$

where $[\bar{v}, \bar{V}]$, μ_1, μ_2 *are defined in* (4.50) *and* (4.57),

$$\begin{aligned}
\Gamma_1(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v}) &= C'[\omega_{IV}^2 \Omega_{IV}^{-1} - \omega_{LS}^2 \Omega_{LS}^{-1}]^{-1} C \,, \\
\Gamma_2(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v}) &= \frac{1}{\omega_{IV}^2} C' \Delta^{-1} C \,, \\
\Gamma_3(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v}) &= \frac{1}{\omega_{LS}^2} C' \Delta^{-1} C \,, \\
\bar{\Gamma}_l(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v}) &= \frac{1}{\omega_l^2} C' \Delta^{-1} C \;, \quad l = 1,\, 2,\, 3,\, 4 \,, \\
\Gamma_R(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v}) &= \frac{1}{\omega_R^2} (\mu_1 + \bar{v})' P_{D_1 Z_2} (\mu_1 + \bar{v}) \,,
\end{aligned}$$

$\kappa_R = (T - k_1 - k_2 - G)/k_2,\ \kappa_l,\ l = 1, 2, 3, 4$ *are defined in* (3.1) - (3.13) *and* M_1, M *in* (2.7).

The above theorem entails that the distributions of the statistics do not depend on neither β nor γ. Observe that Theorem **4.6** follows from algebraic arguments only, so $[Y,\, Z]$ and $[\bar{v}, \bar{V}]$ can be random in an arbitrary way. If the distributions of Z and $[\bar{v}, \bar{V}]$ do not depend on other model parameters, the theorem entails that the distributions of the statistics depend on model parameters only through μ_1 and μ_2. Since μ_2 does not involve δ, μ_1 is the only factor which determine power. If $\mu_1 \neq 0$, the tests have power. This may be the case when at least one instrument is strong (partial identification of model parameters). However, we can observe that when $M_1 Z_2 \Pi_2 a = 0$,

$\mu_1 = 0$ and exogeneity tests have no power. We now provide a formal characterization of the set of parameters in which exogeneity tests have no power.

Let $(M_1 Z_2 \varPi_2)^-$ be any generalized-inverse of $M_1 Z_2 \varPi_2$ and define

$$\mathcal{A} \;=\; (M_1 Z_2 \varPi_2)^- M_1 Z_2 \varPi_2 \,. \tag{4.70}$$

Then, a general solution of the homogeneous equation $M_1 Z_2 \varPi_2 a = 0$ is

$$a \;=\; (I_G - \mathcal{A}) a^* \,, \tag{4.71}$$

where a^* is any arbitrary $G \times 1$ vector [see Rao and Mitra (1971, Theorem 2.3.1)]. Let

$$\mathcal{N}_a \;=\; \{a \in \mathbb{R}^G : a = (I_G - \mathcal{A}) a^*,\, a^* \in \mathbb{R}^G\} \,. \tag{4.72}$$

Note that if the matrix of instruments Z_2 has a full column rank (say k_2), we have $\mathcal{N}_a = \{a \in \mathbb{R}^G : \varPi_2 a = 0\}$. So, provided identification is strong and Rank$(Z_2) = k_2$, $\mathcal{N}_a = \{0\}$. However, in general $\mathcal{N}_a \neq \{0\}$ unless Rank$(M_1 Z_2 \varPi_2) < G$. Corollary **4.7** below characterizes the power of the tests when $a \in N_a$.

Corollary 4.7 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions of Theorem 4.**6**, if $a \in N_a$, we have $\mu_1 = 0$ and the statistics defined in (3.1) - (3.13) have the following representations:*

$$\mathcal{H}_i = T \bar{v}' \varGamma_i(\mu_2, \bar{v}, \bar{V}) \bar{v} \,, \quad i = 1,\, 2,\, 3,$$

$$\mathcal{T}_l \;=\; \kappa_l \bar{v}' \bar{\varGamma}_l(\mu_2, \bar{v}, \bar{V}) \bar{v} \,, \quad l = 1,\, 2,\, 3,\, 4,$$

$$\mathcal{RH} \;=\; \kappa_R \bar{v}' \varGamma_R(\mu_2, \bar{v}, \bar{V}) \bar{v}$$

irrespective of whether the instruments are weak or strong, where

$$\begin{aligned}
\Gamma_i(\mu_2, \bar{v}, \bar{V}) &\equiv \Gamma_i(0, \mu_2, \bar{v}, \bar{V}),\ \bar{\Gamma}_l(\mu_2, \bar{v}, \bar{V}) \equiv \Gamma_l(0, \mu_2, \bar{v}, \bar{V}),\\
\Gamma_R(\mu_2, \bar{v}, \bar{V}) &= \Gamma_R(0, \mu_2, \bar{v}, \bar{V}),\\
\zeta &= -(\Sigma_V^{-1}\delta)/(\sigma_u^2 - \delta'\Sigma_V^{-1}\delta)^{1/2},
\end{aligned}$$

$\Gamma_i(\mu_1, \mu_2, \bar{v}, \bar{V})$, $\Gamma_l(\mu_1, \mu_2, \bar{v}, \bar{V})$ *and* $\Gamma_R(\mu_1, \mu_2, \bar{v}, \bar{V})$ *are defined in Theorem 4.***6**.

First, note that when $a \in N_a$, i.e. when $M_1 Z_2 \Pi_2 a = 0$, the conditional distributions given Z and $\bar{V}$ of exogeneity tests only depend on μ_2 irrespective of the quality of the instruments. In particular, this condition is satisfied when $\Pi_2 = 0$ (complete non identification of model parameters) or $\delta = a = 0$ (under the null hypothesis). Since μ_2 does not depend on δ nor a, all exogeneity test statistics have the same distribution under both the null hypothesis $(\delta = a = 0)$ and the alternative $(\delta \neq 0)$ when $a \in N_a$: the power of these tests cannot exceed the nominal levels. So, the practice of pretesting based on exogeneity tests is unreliable in this case.

Theorem **4.8** below characterizes the distributions of the statistics when the errors are Gaussian, *i.e.* under the assumption (4.43).

Theorem 4.8 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Let the assumptions of Theorem* **4.6** *hold. If furthermore the normality assumption* (4.43) *holds and* $Z = [Z_1, Z_2]$ *is fixed, then*

$$\begin{aligned}
\mathcal{H}_1 &= T[\mu_1 + \bar{v}]'\Gamma_1(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}],\\
\mathcal{H}_2 &= T[\mu_1 + \bar{v}]'\Gamma_2(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}] \sim T\frac{\phi_1(\bar{v},\, \nu_1)}{\phi_2(\bar{v},\, \nu_3)},\\
\mathcal{H}_3|\,\bar{V} &\sim \frac{T}{1 + \kappa_2^{-1}F(T - k_1 - 2G,\, G; \upsilon_2,\, \nu_1)} \leq \bar{\kappa}_1^* F(G,\, T - k_1 - 2G; \nu_1,\, \upsilon_2),
\end{aligned}$$

$$\mathcal{T}_1|\,\bar{V} \sim F(G,\, k_2 - G; \nu_1,\, \upsilon_1),$$

$$
\begin{aligned}
\mathcal{T}_2|\,\bar{V} &\sim F(G,\,T-k_1-2G;\nu_1,\,\upsilon_2),\\
\mathcal{T}_3 &= \kappa_2[\mu_1+\bar{v}]'\varGamma_2(\mu_1,\mu_2,\bar{v},\bar{V})[\mu_1+\bar{v}] \sim \kappa_2\frac{\phi_1(\bar{v},\,\nu_1)}{\phi_2(\bar{v},\,\nu_3)}\,,\\
\mathcal{T}_4|\,\bar{V} &\sim \frac{\kappa_4}{1+\kappa_2^{-1}F(T-k_1-2G,\,G;\upsilon_2,\,\nu_1)} \leq \bar{\kappa}_2^{*}F(G,\,T-k_1-2G;\nu_1,\,\upsilon_2)\,,
\end{aligned}
$$

$$
\mathcal{RH}|\,\bar{V} \sim F(k_2,\,T-k-G;\nu_R,\,\upsilon_R)\,,
$$

where

$$
\begin{aligned}
\phi_1(\bar{v},\,\nu_1)|\,\bar{V} &= [\mu_1+\bar{v}]'C'\varDelta^{-1}C[\mu_1+\bar{v}]|\,\bar{V} \sim \chi^2(G;\nu_1),\\
\phi_2(\bar{v},\,\nu_3)|\,\bar{V} &= \omega_{IV}^2|\,\bar{V} \sim \chi^2(T-k_1-G;\nu_3)\,,
\end{aligned}
$$

$$
\begin{aligned}
\nu_1 &= \mu_1'C'\varDelta^{-1}C\mu_1,\,\nu_3=\mu_1'(D_*'D^*)\mu_1,\,\upsilon_1=\mu_1'E\mu_1,\,\upsilon_2=\mu_1'(C_*-C'\varDelta^{-1}C)\mu_1\,,\\
\nu_R &= \mu_1'P_{D_1Z_2}\mu_1,\,\upsilon_R=\mu_1'(D_1-P_{D_1Z_2})\mu_1,\,\bar{\kappa}_1^{*}=\frac{TG}{T-k_1-2G},\,\bar{\kappa}_2^{*}=\frac{(T-k_1-G)G}{T-k_1-2G}\,,
\end{aligned}
$$

and $[\bar{v},\bar{V}]$, $\mu_1,\ \mu_2,\ \varGamma_i(\mu_1,\ \mu_2,\ \bar{v},\ \bar{V}),\ i=1,\,2,\ \ C,\ \ \varDelta,\ \ C_*,\ \ D_*,\ \ E,\ \ D_1,\ \ P_{D_1\bar{Z}_2},\ \ \kappa_2$ *and* κ_4 *are given in Theorem* **4.6**.

The above theorem entails that given $\bar{V}$, the statistics $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{RH}$ follow double noncentral F-distributions, while $\mathcal{T}_4$ and $\mathcal{H}_3$ are bounded by a double noncentral F-type distribution. However, the distributions of $\mathcal{T}_3$, $\mathcal{H}_2$ and $\mathcal{H}_1$ cannot be characterized by standard distributions. As in Theorem **4.6**, μ_1 is the factor which determine power. If $\mu_1 \neq 0$, the exogeneity tests have power. However, when $\mu_1 = 0$, all tests have no power as showed in Corollary **4.9** below.

Corollary 4.9 FINITE-SAMPLE DISTRIBUTIONS OF EXOGENEITY TESTS. *Under the assumptions of Theorem 4.***8**, *if* $a \in N_a$, *we have* $\nu_1=\nu_3=\upsilon_1=\upsilon_2=\nu_R=\upsilon_R=0$ *so that*

$$
\mathcal{H}_1 = T\bar{v}'\varGamma_1(\mu_2,\bar{v},\bar{V})\bar{v}\,,
$$

$$\begin{aligned}
\mathcal{H}_2 &= T\bar{v}'\Gamma_2(\mu_2, \bar{v}, \bar{V})\bar{v} \sim T\frac{\phi_1(\bar{v})}{\phi_2(\bar{v})}\,, \\
\mathcal{H}_3 &\sim \frac{T}{1+\kappa_2^{-1}F(T-k_1-2G,\, G)} \leq \bar{\kappa}_1^{*}F(G,\, T-k_1-2G)\,,
\end{aligned}$$

$$\begin{aligned}
\mathcal{T}_1 &\sim F(G,\, k_2-G)\,,\, \mathcal{T}_2 \sim F(G,\, T-k_1-2G), \\
\mathcal{T}_3 &= \kappa_2\bar{v}'\Gamma_2(\mu_2, \bar{v}, \bar{V})\bar{v} \sim \kappa_2\frac{\phi_1(\bar{v})}{\phi_2(\bar{v})}\,, \\
\mathcal{T}_4 &\sim \frac{\kappa_4}{1+\kappa_2^{-1}F(T-k_1-2G,\, G)} \leq \bar{\kappa}_2^{*}F(G,\, T-k_1-2G)\,,
\end{aligned}$$

$$\mathcal{RH} \sim F(k_2,\, T-k-G)\,,$$

where

$$\phi_1(\bar{v}) \equiv \phi_1(\bar{v},\, 0), \quad \phi_2(\bar{v}) \equiv \phi_2(\bar{v},\, 0)\,,$$

$\phi_1(\bar{v},\, \nu_1)$, $\phi_1(\bar{v},\, \nu_3)$, $\Gamma_i(\mu_1, \mu_2, \bar{v}, \bar{V})$, $i = 1, 2$ *are defined in Theorem 4.***8**.

Observe that when $a \in N_a$, the non-centrality parameters in the F-distributions vanish. In particular, under the null hypothesis H_0, we have $a = 0 \in N_a$ and all exogeneity tests are pivotal. Furthermore, all exogeneity test statistics have the same distribution under the null hypothesis ($\delta = a = 0$) and the alternative ($\delta \neq 0$): the power of the tests cannot exceed the nominal levels.

We now describe the exact procedure for testing exogeneity even with non Gaussian errors: Monte Carlo exogeneity (MCE) tests.

5. Exact Monte Carlo exogeneity tests

The finite-sample characterization of the distribution of exogeneity test statistics in the previous section show that the tests are typically robust to weak instruments (level is controlled). However, these distributions (under the null hypothesis) of the statistics are not standard if the errors are non Gaussian. Furthermore, even for Gaussian errors, $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{T}_3$ cannot be characterized by standard distributions. This section develops exact Monte Carlo tests which are identification-robust even if the errors are non Gaussian.

Consider again eq.(2.1) and assume that we test the strict exogeneity of Y, i.e. the hypothesis:

$$H_0 \quad : \quad u \text{ is independent of } [Y, Z]\,. \tag{5.1}$$

If the distribution under H_0 of u/σ_u is given, the conditional distributions of exogeneity test statistics given $[Y,\, Z]$ are pivotal and therefore can be simulated [see Theorem **4.1**]. Let

$$\mathcal{W} \quad \in \quad \{\mathcal{H}_i,\, \mathcal{H}_l,\, \mathcal{RH},\, i = 1,\, 2,\, 3\,; l = 1,\, 2,\, 3,\, 4\,\}\,. \tag{5.2}$$

We shall consider two cases: first, the support of W is continuous and second, the support may be a discrete set.

We first focus on the case where exogeneity tests have continuous distributions. Let $W_1,\, \ldots\,,\, W_N$ be a sample of N replications of identically distributed exchangeable random variables with the same distribution as W [for more details on exchangeability, see Dufour (2006)]. Define $W(N) = (W_1,\, \ldots\,,\, W_N)'$ and let W_0 be the value of W based on the observed data. Let

$$\hat{p}_N(x) = \frac{N\hat{G}_N(x) + 1}{N+1}, \tag{5.3}$$

$$\hat{G}_N(x) = \hat{G}_N[x; \mathcal{W}(N)], \tag{5.4}$$

where the survival function $\hat{G}_N$ is given by

$$\hat{G}_N[x; \mathcal{W}(N)] \quad = \quad \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[\mathcal{W}_i \geq x], \tag{5.5}$$

$$\begin{aligned} \mathbf{1}[C] \quad &= \quad 1 \quad \text{if condition C holds,} \\ &= \quad 0 \quad \text{otherwise.} \end{aligned} \tag{5.6}$$

Then, we can show that

$$P[\hat{p}_N(\mathcal{W}_0) \le \alpha] \;=\; \frac{I[\alpha(N+1)]}{N+1} \quad \text{for } \; 0 \le \alpha \le 1\,, \tag{5.7}$$

[see Dufour (2006, Proposition 2.2)], where $I[x]$ is the largest integer less than or equal to x. So, $\hat{p}_N(W_0) \le \alpha$ is the critical region of the MC-test with level $1-\alpha$ and $\hat{p}_N(W_0)$ is the MC-test p-value.

We shall now extend this procedure to the general case where the distribution of the statistic W may be discrete. Assume that $W(N) = (W_1, \ldots, W_N)'$ is a sequence of exchangeable random variables which may exhibit ties with positive probability. More precisely

$$P(\mathcal{W}_j = \mathcal{W}_{j'}) > 0 \quad \text{for } j \ne j',\, j, j' = 1, \ldots, N\,. \tag{5.8}$$

Let us associate with each variable $W_j,\ j = 1, \ldots, N,$ a random variable $U_j,\ j = 1, \ldots, N$ such that

$$\mathcal{U}_j, \ldots, \mathcal{U}_N \overset{i.i.d}{\sim} \mathcal{U}(0,1)\,, \tag{5.9}$$

$U(N) = (U_1, \ldots, U_N)'$ is independent of $W(N) = (W_1, \ldots, W_N)'$ where $U(0,1)$ is the uniform distribution on the interval $(0,\ 1)$. Then, we consider the pairs

$$\mathcal{Z}_j = (\mathcal{W}_j, \mathcal{U}_j\,),\ j = 1, \ldots, N\,, \tag{5.10}$$

which are ordered according to the lexicographic order:

$$(\mathcal{W}_j, \mathcal{U}_j\,) \le (\mathcal{W}_{j'}, \mathcal{U}_{j'}\,) \Longleftrightarrow \{\mathcal{W}_j < \mathcal{W}_{j'} \;\text{ or } (\mathcal{W}_j = \mathcal{W}_{j'} \;\text{ and } \mathcal{U}_j \le \mathcal{U}_{j'})\}\,. \tag{5.11}$$

Let us define the randomized p-value function as

$$\tilde{p}_N(x) = \frac{N\tilde{G}_N(x) + 1}{N+1}, \tag{5.12}$$

where the tail-area function $\tilde{G}_N$ is given by

$$\tilde{G}_N(x) = \tilde{G}_N[x; \mathcal{U}_0, \mathcal{W}(N), \mathcal{U}(N)], \tag{5.13}$$

and

$$\tilde{G}_N[x; \mathcal{U}_0, \mathcal{W}(N), \mathcal{U}(N)] \;=\; \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}[\mathcal{Z}_j \geq (x, \mathcal{U}_0)], \tag{5.14}$$

U_0 is a $U(0,1)$ random variable independent of $W(N)$ and $U(N)$. Then, we have

$$P[\tilde{p}_N(\mathcal{W}_0) \leq \alpha] \;=\; \frac{I[\alpha(N+1)]}{N+1} \quad \text{for } \; 0 \leq \alpha \leq 1\,, \tag{5.15}$$

[see Dufour (2006, Proposition 2.4)]. So, $\tilde{p}_N(W_0) \leq \alpha$ is the critical region of the MC-test with level $1-\alpha$ and $\tilde{p}_N(W_0)$ is the MC-test p-value.

The algorithm[2] for computing Monte Carlo exogeneity tests p-values in continue distributions setup is described as follows:

1. compute the test statistic W_0 based on the observed data;

2. generate $i.i.d.$ variables $u^{(j)} = [u_1^{(j)}, \ldots, u_T^{(j)}]',\ \ j = 1, \ldots, N\,,$ according to the selected distribution—for example, $u_t^{(j)} \sim N[0,1]$ for all $t = 1, \ldots, T$ and $j = 1, \ldots, N$. Since the distribution of W under H_0 does not involve neither β nor γ, compute the pseudo-samples as functions of the OLS estimators $\hat{\beta}$ and $\hat{\gamma}$ from the observed data, i.e.

$$y_t^{(j)} = Y_t^{(j)\prime}\hat{\beta} + Z_{1t}^{(j)\prime}\hat{\gamma} + u_t^{(j)},\ t = 1,\ \ldots\ ,\ T,\ j = 1,\ \ldots\ ,\ N\,, \tag{5.16}$$

   given the observed data Y and Z_1;

3. compute the corresponding test statistics $W^{(j)},\ j = 1, \ldots, N$;

[2]This algorithm can easily be generalized to discrete case

4. compute the MC p-value

$$\hat{p}_{MC} = \hat{p}_N[\mathcal{W}_0]; \tag{5.17}$$

5. reject the null hypothesis H_0 at level α_1 if $\hat{p}_{MC} \leq \alpha_1$.

The following section analyzes the distribution of exogeneity test statistics in large-sample with or without weak instruments.

6. Asymptotic theory

We now study large-sample properties of DWH-type test statistics. Let us consider again model (2.1) - (2.6) and assume that (4.2) holds. Let us also replace the strong independence assumption between ε and V in (4.2) by the following weaker one:

$$\varepsilon_t \text{ is uncorrelated with } V_t' \text{ and has mean zero and variance } \sigma_\varepsilon^2 \text{ for all } t = 1, \ldots, T\,. \tag{6.1}$$

Without lost of generality, let us define

$$\bar{Z}_2 = M_1 Z_2, \;\; \bar{Z} = [Z_1\,, \bar{Z}_2]\,, \tag{6.2}$$

where M_1 is the projection matrix defined in (2.7). Observe that in (6.2), we have $Z_1 \perp \bar{Z}_2$.

We make the following generic assumptions on the asymptotic behaviour of model variables [where $B > 0$ for a matrix B means that B is positive definite (p.d.), and $\to$ refers to limits as $T \to \infty$]:

$$\frac{1}{T}\left[\begin{array}{cc} V & \varepsilon \end{array}\right]'\left[\begin{array}{cc} V & \varepsilon \end{array}\right] \xrightarrow{p} \left[\begin{array}{cc} \Sigma_V & 0' \\ 0 & \sigma_\varepsilon^2 \end{array}\right] > 0\,, \tag{6.3}$$

$$\frac{1}{T}\bar{Z}'\left[\begin{array}{cc} V & \varepsilon \end{array}\right] \xrightarrow{p} 0\,, \tag{6.4}$$

$$\frac{1}{T}\bar{Z}'\bar{Z} \xrightarrow{p} \Sigma_{\bar{Z}} = \left[\begin{array}{cc} \Sigma_{Z_1} & 0 \\ 0 & \Sigma_{\bar{Z}_2} \end{array}\right] > 0\,, \tag{6.5}$$

$$\frac{1}{\sqrt{T}} V' \varepsilon \xrightarrow{L} S_{V\varepsilon} \,, \tag{6.6}$$

$$\frac{1}{\sqrt{T}} \bar{Z}'[u,\, V,\, \varepsilon] \xrightarrow{L} [S_u,\, S_V,\, S_\varepsilon] \,, \tag{6.7}$$

$$\mathrm{vec}[S_u,\, S_V,\, S_\varepsilon,\, S_{V\varepsilon}] \sim N\left[0\,,\, \Sigma_S\right] \,, \tag{6.8}$$

$$S_\varepsilon \text{ and } S_V \text{ are independent,} \tag{6.9}$$

$$S_u = \begin{bmatrix} S_{1u} \\ S_{2u} \end{bmatrix}, \quad S_V = \begin{bmatrix} S_{1V} \\ S_{2V} \end{bmatrix}, \quad S_\varepsilon = \begin{bmatrix} S_{1\varepsilon} \\ S_{2\varepsilon} \end{bmatrix}, \tag{6.10}$$

$$S_{1u} \sim N\left[0,\, \sigma_u^2 \Sigma_{Z_1}\right] \,, \quad S_{2u} \sim N\left[0,\, \sigma_u^2 \Sigma_{\bar{Z}_2}\right] , \tag{6.11}$$

$$S_{1\varepsilon} \sim N\left[0,\, \sigma_\varepsilon^2 \Sigma_{Z_1}\right] \,, \quad S_{2\varepsilon} \sim N\left[0,\, \sigma_\varepsilon^2 \Sigma_{\bar{Z}_2}\right] \,, \tag{6.12}$$

S_{iu} is a $k_i \times 1$ random vector, S_{iV} is a $k_i \times G$ random matrix matrix $(i = 1,\, 2)$, Σ_V is $G \times G$ positive definite matrix, and $\sigma_u^2 > 0$. Note that the covariance matrix Σ_S may be singular.

From the above assumptions, it is easy to see that

$$\frac{1}{T} \bar{Z}_2' u \xrightarrow{p} 0 \,, \quad \frac{1}{T} \begin{bmatrix} u & V \end{bmatrix}' \begin{bmatrix} u & V \end{bmatrix} \xrightarrow{p} \Sigma = \begin{bmatrix} \sigma_u^2 & \delta' \\ \delta & \Sigma_V \end{bmatrix} > 0 \,, \tag{6.13}$$

where

$$\delta = \Sigma_V a \,,\, \sigma_u^2 = a' \Sigma_V a + \sigma_\varepsilon^2 \,,\, S_u = S_V a + S_\varepsilon = S_V (\Sigma_V^{-1} \delta) + S_\varepsilon \,. \tag{6.14}$$

Under assumptions (6.3) - (6.12),

$$\operatorname*{plim}_{T \to \infty} \hat{\beta} = \beta + (\Pi_2' \Sigma_{\bar{Z}_2} \Pi_2 + \Sigma_V)^{-1} \delta \tag{6.15}$$

and $\hat{\beta}$ is consistent if and only if (iff) $\delta = 0$, irrespective of the rank of Π_2. Furthermore,

$$\tilde{\beta} = \beta + [Y'(M_1 - M)Y]^{-1} Y'(M_1 - M)u = \beta + \left[\frac{Y'(M_1 - M)Y}{T} \right]^{-1} \frac{Y'(M_1 - M)u}{T} \,, \tag{6.16}$$

so, provided the identification condition $rank(\Pi_2) = G$ holds,

$$\frac{Y'(M_1 - M)Y}{T} \xrightarrow{p} \Pi_2' \Sigma_{\bar{Z}_2} \Pi_2 > 0, \; \frac{Y'(M_1 - M)u}{T} \xrightarrow{p} 0 \,. \tag{6.17}$$

and

$$\operatorname*{plim}_{T\to\infty} \tilde{\beta} = \beta\,. \tag{6.18}$$

However, $\tilde{\beta}$ does not generally converge to β when $rank(\Pi_2) < G$.

We now study the asymptotic distributions of the statistics under two main setups: (1) $\Pi_2 = \Pi_0$ with rank$(\Pi_0) = G$ and (2) $\Pi_2 = \Pi_0/\sqrt{T}$ ($\Pi_0 = 0$ is allowed), where Π_0 is a $k_2 \times G$ constant matrix. The second case corresponds to locally weak instruments [Staiger and Stock (1997)]. For a random variable K whose distribution depends on the sample size T, the notation $K \xrightarrow{L} +\infty$ means that $P[K > x] \to 1$ as $T \to \infty$, for any x. Theorem **6.1** below summarizes the asymptotic behaviour of the statistics under H_0.

Theorem 6.1 ASYMPTOTIC DISTRIBUTIONS UNDER THE NULL HYPOTHESIS. *Suppose that the assumptions* (2.1) - (2.4) *and* (6.3) - (6.12) *hold, and let* $\delta = 0$. (A) *If* $\Pi_2 = \Pi_0$ *where* Π_0 *is a fixed* $k_2 \times G$ *matrix with rank* G, *then*

$$\mathcal{H}_i \xrightarrow{L} \chi^2(G),\; i = 1,\, 2,\, 3\;, \tag{6.19}$$

$$\mathcal{T}_1 \xrightarrow{L} F(G,\, k_2 - G), \quad \mathcal{T}_2 \xrightarrow{L} \frac{1}{G}\chi^2(G), \quad \mathcal{T}_l \xrightarrow{L} \chi^2(G),\; l = 3,\, 4\,, \tag{6.20}$$

$$\mathcal{RH} \xrightarrow{L} \frac{1}{k_2}\chi^2(k_2)\,. \tag{6.21}$$

(B) *If* $\Pi_2 = \Pi_0/\sqrt{T}$, *where* Π_0 *is a fixed* $k_2 \times G$ *matrix, then*

$$\mathcal{H}_i \xrightarrow{L} \frac{1}{\bar{\sigma}_u^2} S_{2u}' \Sigma_A S_{2u} \le \chi^2(G),\; i = 1,\, 2, \tag{6.22}$$

$$\mathcal{H}_3 \xrightarrow{L} \chi^2(G), \tag{6.23}$$

$$\mathcal{T}_1 \xrightarrow{L} F(G,\, k - G), \quad \mathcal{T}_2 \xrightarrow{L} \frac{1}{G}\chi^2(G), \quad \mathcal{T}_4 \xrightarrow{L} \chi^2(G), \tag{6.24}$$

$$\mathcal{T}_3 \xrightarrow{L} \frac{1}{\bar{\sigma}_u^2} S_{2u}' \Sigma_A S_{2u} \le \chi^2(G)\,, \tag{6.25}$$

$$\mathcal{RH} \xrightarrow{L} \frac{1}{k_2}\chi^2(k_2)\,, \tag{6.26}$$

where

$$\bar{\sigma}_u^2 = \sigma_u^2 + S_{2u}'\Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})\Psi_V^{-1}\Sigma_V\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}, \tag{6.27}$$

$$\Sigma_A \equiv \Sigma_A(S_{2V}) = \Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}, \tag{6.28}$$

$$\Psi_V = (\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})\,. \tag{6.29}$$

In the above theorem, the statistics $\mathcal{T}_1,\ \mathcal{T}_2,\ \mathcal{T}_4$ and $\mathcal{H}_3$ are asymptotically pivotal irrespective of whether the instruments are weak or not, i.e. their asymptotic distributions under the null hypothesis do not involve any nuisance parameters even when β is not identified or close not to being identifiable. So, all exogeneity tests are asymptotically valid even in presence of weak instruments. Furthermore, Theorem **6.1** provides an upper bound for $\mathcal{T}_3,\ \mathcal{H}_1$ and $\mathcal{H}_2$. This result clearly indicates that the conclusion in Hahn et al. (2008) is inaccurate. In fact, under the assumption (3) of Hahn et al. (2008), and if further $\rho_v = 0$ in eq. (5), the distributions of $\mathcal{H}_1$ and $\mathcal{H}_2$ in Theorem 1 [see Hahn et al. (2008)] are bounded by a chi-square distribution with K degrees of freedom. This result is omitted by the authors who conclude (wrongly) that $\mathcal{H}_1$ and $\mathcal{H}_2$ are invalid in presence of weak IV. When IV are weak, $\mathcal{T}_3,\ \mathcal{H}_1$ and $\mathcal{H}_2$ are valid but conservative.

The following theorem studies the asymptotic power of the tests.

Theorem 6.2 ASYMPTOTIC POWER. *Suppose that the assumptions* (2.1) - (2.4) *and* (6.3) - (6.12) *hold, and let* $\delta \neq 0$. (A) *If* $\Pi_2 = \Pi_0$ *where* Π_0 *is a* $k_2 \times G$ *constant matrix with rank* G, *then, for* $i = 1,\ 2,\ 3$ *, and* $l = 1,\ 2,\ 3,\ 4$,

$$\mathcal{H}_i, \xrightarrow{L} +\infty,\ \mathcal{T}_l \xrightarrow{L} +\infty\,, \mathcal{RH}, \xrightarrow{L} +\infty\,. \tag{6.30}$$

(B) If $\Pi_2 = \Pi_0/\sqrt{T}$ where Π_0 is a $k_2 \times G$ constant matrix, then

$$\mathcal{H}_i \xrightarrow{L} \frac{1}{\sigma_{i*}^2}(\Pi_0 a - \Sigma_{\bar{Z}_2}^{-1} S_{2\varepsilon})' \Delta_V (\Pi_0 a - \Sigma_{\bar{Z}_2}^{-1} S_{2\varepsilon}),\ i = 1,\, 2,\, 3\,, \tag{6.31}$$

$$\mathcal{T}_l \xrightarrow{L} \frac{\bar{\kappa}_l}{\tilde{\sigma}_{l*}^2}(\Pi_0 a - \Sigma_{\bar{Z}_2}^{-1} S_{2\varepsilon})' \Delta_V (\Pi_0 a - \Sigma_{\bar{Z}_2}^{-1} S_{2\varepsilon}),\ l = 1,\, 2,\, 3,\, 4\,, \tag{6.32}$$

$$\mathcal{RH} \quad \xrightarrow{L} \quad \frac{1}{k_2 \sigma_\varepsilon^2}(S_{2\varepsilon} - \Sigma_{\bar{Z}_2}\Pi_0 a)' \Sigma_{\bar{Z}_2}^{-1} (S_{2\varepsilon} - \Sigma_{\bar{Z}_2}\Pi_0 a) \sim \frac{1}{k_2}\chi^2(k_2,\, \mu_R)\,, \tag{6.33}$$

where $a = \Sigma_V^{-1}\delta,\, \mu_R = a'\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0 a$ *and*

$$\bar{\kappa}_1 = (k_2 - G)/G,\ \bar{\kappa}_2 = 1/G,\ \ \bar{\kappa}_3 = \bar{\kappa}_4 = 1\,, \tag{6.34}$$

$$\Delta_V = (\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})', \tag{6.35}$$

$$\Psi_V = (\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})\,, \tag{6.36}$$

$$\sigma_{1*}^2 = \sigma_{2*}^2 = \tilde{\sigma}_*^2,\, \sigma_{3*}^2 = \sigma_\varepsilon^2,\, \tilde{\sigma}_{2*}^2 = \tilde{\sigma}_{4*}^2 = \sigma_\varepsilon^2,\, \tilde{\sigma}_{3*}^2 = \tilde{\sigma}_*^2, \tag{6.37}$$

$$\tilde{\sigma}_{1*}^2 = (\Sigma_{\bar{Z}_2}^{-1/2} S_{2V} a + \Sigma_{\bar{Z}_2}^{-1/2} S_{2\varepsilon})' \Delta_V^* (\Sigma_{\bar{Z}_2}^{-1/2} S_{2V} a + \Sigma_{\bar{Z}_2}^{-1/2} S_{2\varepsilon}), \tag{6.38}$$

$$\Delta_V^* = I_{k_2} - \Sigma_{\bar{Z}_2}^{-1/2}\Delta_V \Sigma_{\bar{Z}_2}^{-1/2}, \tag{6.39}$$

$$\tilde{\sigma}_*^2 = \sigma_u^2 - 2\delta'\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1} S_{2u} \tag{6.40}$$

$$+S_{2u}'\Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})\Psi_V^{-1}\Sigma_V\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}\,. \tag{6.41}$$

Further,

$$\mathcal{H}_3 | S_{2V} \xrightarrow{L} \chi^2(G,\, \mu_V), \tag{6.42}$$

$$\mathcal{T}_1 | S_{2V} \xrightarrow{L} F(G,\, k - G;\, \mu_V,\, \lambda_V)\,, \tag{6.43}$$

$$\mathcal{T}_2 | S_{2V} \xrightarrow{L} \frac{1}{G}\chi^2(G,\, \mu_V), \quad \mathcal{T}_4 | S_{2V} \xrightarrow{L} \chi^2(G,\, \mu_V)\,, \tag{6.44}$$

where

$$\mu_V = \frac{1}{\sigma_\varepsilon^2} a'\Pi_0'\Delta_V\Pi_0 a = \delta'\Sigma_V^{-1}\Pi_0'\Delta_V\Pi_0\Sigma_V^{-1}\delta\,, \tag{6.45}$$

$$\lambda_V = \frac{1}{\sigma_\varepsilon^2} a' S_{2V}' \Sigma_{\bar{Z}_2}^{-1/2} \Delta_V^* \Sigma_{\bar{Z}_2}^{-1/2} S_{2V} a = \frac{1}{\sigma_\varepsilon^2} a' S_{2V}' (\Sigma_{\bar{Z}_2}^{-1} - \Sigma_{\bar{Z}_2}^{-1} \Delta_V \Sigma_{\bar{Z}_2}^{-1}) S_{2V} a \,. \tag{6.46}$$

Part (A) of Theorem **6.2** shows that DWH and $\mathcal{RH}$ tests are consistent in presence of strong identification. Part (B) establishes that the tests are non consistent when parameters are not identified or nearly so. More precisely, all exogeneity tests converge to finite non-degenerate distributions in presence of weak instruments. The conditional limiting distributions of $\mathcal{H}_3,\ \mathcal{T}_2, \mathcal{T}_4$ and $\mathcal{RH}$ given S_{2V} are noncentral chi-square distributions while $\mathcal{T}_1$ has a double noncentral F-distribution. However, the conditional limiting distributions of $\mathcal{H}_1,\ \mathcal{H}_2$ and $\mathcal{T}_3$, given S_{2V} cannot be characterized by standard distributions.

Overall, our results indicate that exogeneity tests may have power even in presence of weak IV. We now characterize in the following corollary the situation where the tests have no power.

Corollary 6.3 ASYMPTOTIC POWER. *Under the assumption of Theorem 6.***2**, *if $\Pi_2 = 0$, then*

$$\mathcal{H}_i \xrightarrow{L} \frac{1}{\sigma_{0*}^2} S_{2\varepsilon}' \Sigma_A^0 S_{2\varepsilon} \leq \chi^2(G),\ i = 1,\ 2,\ \ \mathcal{H}_3 \xrightarrow{L} \chi^2(G)\,, \tag{6.47}$$

$$\mathcal{T}_1 \xrightarrow{L} F(G,\, k - G), \quad \mathcal{T}_2 \xrightarrow{L} \frac{1}{G}\chi^2(G), \quad \mathcal{T}_4 \xrightarrow{L} \chi^2(G), \tag{6.48}$$

$$\mathcal{T}_3 \xrightarrow{L} \frac{1}{\sigma_{0*}^2} S_{2\varepsilon}' \Sigma_A^0 S_{2\varepsilon} \leq \chi^2(G),\ \ \mathcal{RH} \xrightarrow{L} \frac{1}{k_2}\chi^2(k_2)\,, \tag{6.49}$$

where

$$\Sigma_A^0 = \Sigma_{\bar{Z}_2}^{-1} S_{2V} (S_{2V}' \Sigma_{\bar{Z}_2}^{-1} S_{2V})^{-1} S_{2V}' \Sigma_{\bar{Z}_2}^{-1}\,, \tag{6.50}$$

$$\sigma_{0*}^2 = \sigma_\varepsilon^2 + S_{2\varepsilon}' \Sigma_{\bar{Z}_2}^{-1} S_{2V} (S_{2V}' \Sigma_{\bar{Z}_2}^{-1} S_{2V})^{-1} \Sigma_V (S_{2V}' \Sigma_{\bar{Z}_2}^{-1} S_{2V})^{-1} S_{2V}' \Sigma_{\bar{Z}_2}^{-1} S_{2\varepsilon} \geq \sigma_\varepsilon^2\,. \tag{6.51}$$

In the above corollary, the instruments are irrelevant, i.e. $\Pi_2 = 0$. So, all the non-centrality parameters mentioned above in Theorem **6.2** vanish so that the statistics $\mathcal{H}_3,\ \mathcal{T}_2,\ \mathcal{T}_4$ and $\mathcal{RH}$ have chi-square limiting distributions while $\mathcal{T}_1$ is asymptotically distributed as a Fisher with $(k_2 - G, G)$ degrees of freedom. Furthermore, the asymptotic distributions of the statistics $\mathcal{H}_1,\ \mathcal{H}_2$ and $\mathcal{T}_3$ are bounded by a chi-square distribution with G degrees of freedom. These results mean that the asymptotic power of $\mathcal{H}_3,\ \mathcal{T}_2,\ \mathcal{T}_4, \mathcal{T}_1$ and $\mathcal{RH}$ equals the nominal levels while those of $\mathcal{H}_1,\ \mathcal{H}_2$ and $\mathcal{T}_3$ cannot exceed the nominal level, a finding consistent with our finite-sample theory.

7. Simulation experiments

This section presents the Monte Carlo experiments. First, we study the performance of OLS, 2SLS and two-stage estimators with possibly weak-IV. Second, we assess the effects of weak instruments on DWH and RH tests: size and power.

7.1. Performance of OLS, 2SLS and two-stage estimators with possibly weak-IV

For this experiment, we consider a single simultaneous equations system described by the following DGP:

$$y = Y\beta + u, \quad Y = Z_2 \mathit{\Pi}_2 + V, \tag{7.1}$$

where y and Y are $T \times 1$ random vectors $(G = 1)$, Z_2 is a $T \times k_2$ matrix of instruments such that $Z_{2t} \overset{i.i.d}{\sim} N(0,\, I_{k_2}),\ t = 1,\, \ldots,\, T,$ and $\mathit{\Pi}_2$ is a vector of dimension k_2 such that $\mathit{\Pi}_2 = \sqrt{\frac{\mu^2}{T\|Z_2 C\|}}C,$ where C is a $k_2 \times 1$ vector of ones and μ^2 is a concentration parameter. As in Guggenberger (2008), we cover several values of $\mu^2 : \mu^2 \in \{0;\, 13;\, 200;\, 613;\, 2,000;\, 1,000,000\}$ where the values of μ^2 less than 613 correspond to those in Hansen, Hausman and Newey (2008). In our framework, small values of μ^2 (say $\mu^2 \leq 613$) depicted cases where the IV are weak so that the parameter of interest β is not identified or weakly identified. The correlation between u and V is set at $\rho \in \{0,\, .05,\, .1,\, .5,\, .6,\, .95\}$ and the true value of β equals 1. We take $k_2 = 5$ instruments[3], so, both 2SLS and OLS estimators have finite moments. The sample size is $T = 500$ and the number of replications is $N = 10,000$. The results are presented in Tables 2.5 - 2.17 in Appendix C.

In the first column of the tables, we report the different estimators while in the second, we report the concentration parameters μ^2 which represents the quality of the IV. Finally, the other columns report the correlation ρ between the errors and (possibly) endogenous regressors.

[3]The choices of $k_2 = 10,\ 20$ lead to the same conclusions.

Our major findings can be summarized into two points: (1) over a wide range cases, including weak IV and moderate endogeneity, OLS performs better than 2SLS [finding similar to Kiviet and Niemczyk (2007)]; (2) pretest-estimators based on exogeneity have an excellent overall performance compared with usual IV estimator. This suggests that the practice of pretesting based on exogeneity tests is not to bad (at least in the viewpoint of estimation) as claimed by Guggenberger (2008).

We now analyze the properties (level and power) of the tests.

7.2. Size and power of DWH and RH exogeneity tests

In this subsection, we analyze the properties (size and power) of DWH and RH tests in presence of possibly weak instruments. We now consider the two endogenous variables model described by the following data generating process:

$$y = Y_1\beta_1 + Y_2\beta_2 + u, \quad (Y_1, Y_2) = (Z_2\mathit{\Pi}_{21}, Z_2\mathit{\Pi}_{22}) + (V_1, V_2), \tag{7.2}$$

where Z_2 is a $T \times k_2$ matrix of instruments such that Z_{2t} follow $i.i.d\ N(0,\ I_{k_2})$ for $t = 1,\ \ldots,\ T$, $\mathit{\Pi}_{21}$ and $\mathit{\Pi}_{22}$ are vectors of dimension k_2. We assume that

$$u = Va + \varepsilon = V_1a_1 + V_2a_2 + \varepsilon, \tag{7.3}$$

where a_1 and a_2 are 2×1 vectors and ε is independent with $V = (V_1, V_2)$, V_1 and V_2 are $T \times 1$ vectors. Through this experiment,

$$(V_{1t},\ V_{2t})' \overset{i.i.d}{\sim} N\left(0,\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) \quad \text{and} \quad \varepsilon_t \overset{i.i.d}{\sim} N\,(0,\ 1)\,, \quad \text{for all} \quad t = 1,\ \ldots,\ T\,. \tag{7.4}$$

The above setup allows us to take into account situations where $\beta = (\beta_1, \beta_2)'$ is partially identified. In particular, if $\mathit{\Pi}_{21} = 0$ and $\mathit{\Pi}_{22}'\mathit{\Pi}_{22} \neq 0$, the instruments Z_2 fail to identify the entire vector β but β_2 is identified. We define

$$\mathit{\Pi}_{21} = \eta_1 C_0,\ \ \mathit{\Pi}_{22} = \eta_2 C_1, \tag{7.5}$$

where η_1 and η_2 take the value 0 (design of complete non identification), $.01$ (design of weak identification) or $.5$ (design of strong identification), $[C_0, C_1]$ is a $k_2 \times 2$ matrix obtained by taking the first two columns of the identity matrix of order k_2. The number of instruments k_2 varies in $\{5, 10, 20\}$ and the true value of β is set at $\beta_0 = (2, 5)'$. It is worthwhile to note that when η_1 and η_2 belong to $\{0, .01\}$, the instruments Z_2 are weak and both ordinary least squares and two stage least squares estimators of β in (7.2) are biased and inconsistent unless $a_1 = a_2 = 0$. The simulations are run for different samples: $T = 50, 500$, and the number of replications for each sample is $N = 10,000$. The endogeneity vector a is chosen such that

$$a = (a_1, a_2)' \in \{(-20, 0)', (-5, 5)', (0, 0)', (.5, .2)', (100, 100)'\} . \tag{7.6}$$

From the above setup, the usual exogeneity hypothesis for Y can be expressed as

$$H_0 : a = (a_1, a_2)' = (0, 0)'. \tag{7.7}$$

The nominal level of the tests is 5%. For each value of the vector a, we compute the empirical rejection probability of exogeneity test statistics. When $a = 0$, the rejection frequencies are the empirical levels of the tests. However, if $a \neq 0$, the rejection frequencies represent the power of the tests.

The results are presented in Tables 2.1 - 2.2 below. In the first column of the tables, we report the statistics while in the second column, we report the values of k_2 (number of excluded instruments). Finally in the other columns, we report for each value of the endogeneity parameter a and the qualities of the instruments η_1 and η_2, the rejection frequencies at nominal level 5%.

First, we note that all exogeneity tests are valid whether the instruments are strong or weak. In particular, $\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_4$, $\mathcal{H}_3$ and RH control the level while $\mathcal{T}_3$, $\mathcal{H}_1$ and $\mathcal{H}_2$ are conservative for small-sample [see the column $(a_1, a_2)' = (0, 0)'$ in Table 2.1 below]. However, all $\mathcal{T}_3$, H_1 and $\mathcal{H}_2$ do not exhibit this problem in large-sample [see the column $(a_1, a_2)' = (0, 0)'$ in Table 2.2 below].

Second, all exogeneity tests have a low power when both β_1 and β_2 are not identified even in large-sample [see Tables 2.1 - 2.2 for $\eta_1 \in \{0, .01\}$ and $\eta_2 = 0$.] Nevertheless, when at least one component of β is identified [Table 2.1 (continued) and Table 2.2 (continued)], all exogeneity tests exhibit power. When $\eta_1 = \eta_2 = 0$ (irrelevant IV), all exogeneity tests have a low power (similar to Guggenberger (2008)).

Moreover, we observe that $\mathcal{H}_3$ dominates (hight power) $\mathcal{H}_2$ which itself dominates $\mathcal{H}_1$ irrespective of whether the instruments are strong or weak. The same argument applies to Wu-tests. Endeed, $\mathcal{T}_2$ dominates T_4, $\mathcal{T}_4$ dominates $\mathcal{T}_1$ and $\mathcal{T}_1$ dominates T_3.

8. Empirical illustrations

This section illustrates the behaviour of exogeneity tests through two empirical applications related to important issues in macroeconomics and labor economics literature: the relation between trade and growth [see, Dufour and Taamouti (2006), Irwin and Tervio (2002), Frankel and Romer (1999), Harrison (1996), Mankiw and al. (1992)] and the widely studied problem of returns to education [Dufour and Taamouti (2006), Angrist and Krueger (1991), Angrist and Krueger (1995), Angrist and al. (1999), Mankiw and al. (1992)].

8.1. Trade and growth

The trade and growth model studies the relationship between standards of living and openness. The recent studies in this issue include Irwin and Tervio (2002), Frankel and Romer (1999), Harrison (1996), Mankiw and al. (1992) and the survey of Rodrik (1995). Even if many studies conclude that openness is conductive to higher growth, there is no evidence concerning the effect of openness on income. Estimating the impact of openness on income through cross-country regression often raises the problem of finding a good proxy for openness. Frankel and Romer (1999) argue that trade share (ratio of imports or exports to GDP) which is the commonly used indicator of openness should be viewed as endogenous variable, and similarly for the other indicators such as trade policies. So,

Table 2.1. Power of exogeneity tests at nominal level 5%; $G = 2$, $T = 50$

	k_2	$(a_1, a_2)' = (-20, 0)'$			$(a_1, a_2)' = (-5, 5)'$			$(a_1, a_2)' = (0, 0)'$			$(a_1, a_2)' = (.5, .2)'$			$(a_1, a_2)' = (100, 100)'$		
		$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$
$\mathcal{T}_1$	5	4.98	4.6	65.81	5.26	4.92	70.9	4.87	5.06	5.24	5.09	4.84	19.85	4.94	4.18	70.09
$\mathcal{T}_2$	5	4.98	24.92	100	5.04	6.77	100	4.96	5.38	5.26	4.87	4.61	53.19	4.91	76.71	100
$\mathcal{T}_3$	5	0	0.19	97.93	0.02	0.05	97.85	0.02	0.03	0.59	0.03	0	29.02	0.01	5.83	97.93
$\mathcal{T}_4$	5	4.64	24.07	100	4.67	6.29	100	4.63	4.91	4.93	4.51	4.42	52	4.62	76.25	100
$\mathcal{H}_1$	5	0	0.09	92.53	0.01	0.02	91.83	0.01	0.02	0.26	0	0	17.97	0	3.59	92.48
$\mathcal{H}_2$	5	0.01	0.25	98.09	0.03	0.05	98.02	0.02	0.04	0.74	0.04	0	31.42	0.02	6.89	98.14
$\mathcal{H}_3$	5	5.34	25.73	100	5.33	7.19	100	5.27	5.72	5.56	5.18	4.92	54.41	5.31	77.11	100
$\mathcal{RH}$	5	4.84	45.25	100	5.36	7.83	100	5.04	5.2	4.9	4.88	4.73	41.31	5.02	100	100
$\mathcal{T}_1$	10	4.9	3.95	98.38	4.92	5.34	98.93	4.82	4.81	5.25	4.88	5.22	34.18	4.91	3.28	99.23
$\mathcal{T}_2$	10	5.01	17.5	100	5.19	6.2	100	5.16	4.88	5.07	4.77	5.45	54.24	4.8	50.74	100
$\mathcal{T}_3$	10	0.35	1.88	100	0.38	0.29	100	0.3	0.33	1.47	0.36	0.3	43.01	0.22	14.7	100
$\mathcal{T}_4$	10	4.65	16.77	100	4.75	5.73	100	4.78	4.55	4.72	4.45	5.02	52.81	4.46	50.05	100
$\mathcal{H}_1$	10	0.16	1.05	99.31	0.18	0.14	99.22	0.2	0.14	0.49	0.14	0.14	28.92	0.1	9.88	99.25
$\mathcal{H}_2$	10	0.46	2.3	100	0.48	0.42	100	0.38	0.43	1.76	0.46	0.39	45.54	0.33	16.85	100
$\mathcal{H}_3$	10	5.32	18.11	100	5.43	6.56	100	5.46	5.18	5.41	5.06	5.75	55.31	5.12	51.25	100
$\mathcal{RH}$	10	5.17	57.58	100	4.83	7.62	100	4.83	5.34	4.97	4.93	5.41	34.5	4.57	100	100
$\mathcal{T}_1$	20	4.93	2.26	99.8	4.94	4.64	99.78	4.9	5.02	5.07	5.02	4.93	39.4	5.02	1.5	99.96
$\mathcal{T}_2$	20	4.75	8.97	100	4.9	5.54	100	5.09	5.32	4.99	4.95	4.94	49.34	4.92	17.32	100
$\mathcal{T}_3$	20	1.95	3.73	100	1.82	2.01	100	2.1	2.02	2.79	2.01	1.95	44.9	1.94	9.2	100
$\mathcal{T}_4$	20	4.43	8.42	100	4.51	5.21	100	4.74	5.04	4.61	4.63	4.57	47.89	4.52	16.45	100
$\mathcal{H}_1$	20	1.08	2.43	99.89	1.13	1.08	99.82	1.13	1.2	1.03	1.08	1.21	29.88	1.15	6.44	99.7
$\mathcal{H}_2$	20	2.32	4.37	100	2.26	2.6	100	2.67	2.57	3.28	2.46	2.48	47.46	2.33	10.39	100
$\mathcal{H}_3$	20	5.15	9.36	100	5.25	5.73	100	5.4	5.68	5.41	5.23	5.18	50.31	5.23	17.76	100
$\mathcal{RH}$	20	4.88	79.08	100	5.03	8.36	100	5.38	5	5.21	5.07	5.04	24.88	5.3	100	100

Table 2.1 (continued). Power of exogeneity tests at nominal level 5%; $G = 2, T = 50$

	k_2	$(a_1, a_2)' = (-20, 0)'$			$(a_1, a_2)' = (-5, 5)'$			$(a_1, a_2)' = (0, 0)'$			$(a_1, a_2)' = (.5, .2)'$			$(a_1, a_2)' = (100, 100)'$		
		$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$
$\mathcal{T}_1$	5	4.73	15.16	81.58	69.69	68.76	78.22	4.91	5.26	5	8.01	7.48	24.2	63.6	65.14	78.04
$\mathcal{T}_2$	5	5.1	37.9	100	100	100	100	5.51	5.29	5.2	12.95	12.42	64.31	100	100	100
$\mathcal{T}_3$	5	0.63	18.25	98.68	98.15	98.26	98.50	0.75	0.85	0.83	3.82	3.47	42.79	97.43	97.09	98.52
$\mathcal{T}_4$	5	4.77	36.89	100	100	100	100	5.06	4.98	4.78	12.24	11.72	63.06	100	100	100
$\mathcal{H}_1$	5	0.27	10.48	90.44	92	92.3	92.20	0.39	0.29	0.32	1.93	1.69	24.39	92.4	91.95	92.12
$\mathcal{H}_2$	5	0.77	20.16	98.82	98.33	98.43	98.52	0.87	0.96	0.99	4.44	4.08	45.64	97.59	97.31	98.64
$\mathcal{H}_3$	5	5.48	38.88	100	100	100	100	5.83	5.64	5.41	13.39	12.95	65.44	100	100	100
$\mathcal{RH}$	5	5.13	28.27	100	100	100	100	4.77	5.13	5.17	9.81	10.28	50.59	100	100	100
$\mathcal{T}_1$	10	5.18	26.81	99.76	98.81	99.17	99.56	5.26	5.3	4.86	11.05	11.61	43.71	99.12	99.28	99.74
$\mathcal{T}_2$	10	5.29	41.58	100	100	100	100	4.92	5.19	5.07	13.49	14.75	66.24	100	100	100
$\mathcal{T}_3$	10	1.7	31.1	99.98	99.97	99.99	100	1.58	1.6	1.88	7.75	8.29	57.52	100	100	100
$\mathcal{T}_4$	10	4.96	40.35	100	100	100	100	4.57	4.87	4.67	12.81	14	65.15	100	100	100
$\mathcal{H}_1$	10	0.73	18.21	98.22	99.08	98.98	98.9	0.55	0.5	0.48	3.34	3.88	32.85	99.28	99.26	98.29
$\mathcal{H}_2$	10	2	33.67	99.98	99.98	100	100	1.88	2.03	2.31	8.65	9.3	60.4	100	100	100
$\mathcal{H}_3$	10	5.61	42.64	100	100	100	100	5.3	5.53	5.38	14.05	15.32	67.3	100	100	100
$\mathcal{RH}$	10	5.24	24.16	100	100	100	100	4.92	5.07	5.11	8.55	8.94	43.87	100	100	100
$\mathcal{T}_1$	20	5.12	27.67	99.96	99.45	99.48	99.62	4.86	4.91	4.29	10.45	10.95	41.15	99.91	99.9	99.94
$\mathcal{T}_2$	20	5.06	34.7	100	100	100	100	4.93	4.77	4.3	11.85	12.03	51.76	100	100	100
$\mathcal{T}_3$	20	2.97	30.26	100	100	100	100	3.2	2.88	2.74	9.14	9.14	47.52	100	100	100
$\mathcal{T}_4$	20	4.7	33.32	100	100	100	100	4.57	4.45	3.97	11.13	11.34	50.35	100	100	100
$\mathcal{H}_1$	20	1.2	17.73	99.24	99.93	99.91	99.93	1.1	1.03	0.72	4.51	4.53	27.81	99.77	99.81	98.75
$\mathcal{H}_2$	20	3.59	32.57	100	100	100	100	3.65	3.39	3.27	10.24	10.25	50.07	100	100	100
$\mathcal{H}_3$	20	5.32	35.69	100	100	100	100	5.25	5.06	4.55	12.42	12.55	52.91	100	100	100
$\mathcal{RH}$	20	5.46	16.17	100	100	100	100	5.2	4.64	4.82	7.45	7.45	26.62	100	100	100

Table 2.2. Power of exogeneity tests at nominal level 5%; $G = 2$, $T = 500$

	k_2	$(a_1, a_2)' = (-20, 0)'$			$(a_1, a_2)' = (-5, 5)'$			$(a_1, a_2)' = (0, 0)'$			$(a_1, a_2)' = (.5, .2)'$			$(a_1, a_2)' = (100, 100)'$		
		$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$	$\eta_1 = 0$ $\eta_2 = 0$	$\eta_1 = .01$ $\eta_2 = 0$	$\eta_1 = .5$ $\eta_2 = 0$
$\mathcal{T}_1$	5	5.24	6.23	100	5.12	5.35	100	5.06	4.76	4.73	4.8	4.98	94.91	4.91	5.96	100
$\mathcal{T}_2$	5	4.66	91.92	100	5.11	27.86	100	5.11	4.91	4.43	5.35	5.09	100	4.92	98.13	100
$\mathcal{T}_3$	5	0.02	13.61	100	0.04	0.51	99.98	0	0	0.99	0.02	0.03	99.45	0.01	19.26	99.99
$\mathcal{T}_4$	5	4.64	91.89	100	5.06	27.79	100	5.03	4.89	4.38	5.29	5.09	100	4.88	98.13	100
$\mathcal{H}_1$	5	0.02	13.26	99.93	0.04	0.45	99.86	0	0	0.64	0.02	0.03	98.25	0.01	18.87	99.88
$\mathcal{H}_2$	5	0.02	13.72	100	0.05	0.53	99.98	0	0	1.01	0.02	0.03	99.46	0.01	19.39	99.99
$\mathcal{H}_3$	5	4.68	91.94	100	5.14	27.96	100	5.12	4.94	4.44	5.39	5.12	100	4.98	98.13	100
$\mathcal{RH}$	5	4.76	100	100	5.04	45.45	100	5.02	5.02	4.74	5.05	5.59	100	5.34	100	100
$\mathcal{T}_1$	10	5.26	6.71	100	5.46	6.32	100	5	5.37	4.96	5.16	5.15	100	5.23	7.52	100
$\mathcal{T}_2$	10	4.63	86.64	100	4.75	30.49	100	4.84	5.6	4.91	4.74	5.53	100	4.91	95.81	100
$\mathcal{T}_3$	10	0.16	46.63	100	0.17	4.49	100	0.14	0.2	1.7	0.12	0.24	100	0.19	64.18	100
$\mathcal{T}_4$	10	4.62	86.63	100	4.7	30.45	100	4.84	5.57	4.9	4.68	5.48	100	4.91	95.81	100
$\mathcal{H}_1$	10	0.15	45.96	100	0.17	4.26	100	0.14	0.2	0.92	0.12	0.23	99.99	0.19	63.68	100
$\mathcal{H}_2$	10	0.16	46.97	100	0.17	4.62	100	0.15	0.2	1.72	0.15	0.25	100	0.19	64.5	100
$\mathcal{H}_3$	10	4.68	86.67	100	4.77	30.55	100	4.87	5.65	4.93	4.78	5.56	100	4.96	95.83	100
$\mathcal{RH}$	10	4.7	100	100	4.5	67.61	100	5.01	5.44	4.89	4.78	5.69	100	4.85	100	100
$\mathcal{T}_1$	20	5.07	10.67	100	5.27	8.1	100	4.84	5.15	5.03	4.82	5.45	100	4.99	11	100
$\mathcal{T}_2$	20	5.07	86.47	100	5.17	31.8	100	4.79	5.3	5.07	5.16	5.51	100	4.87	93.16	100
$\mathcal{T}_3$	20	1.2	79.4	100	1.38	17.44	100	1.1	1.46	2.87	1.22	1.52	100	1.28	89.05	100
$\mathcal{T}_4$	20	5.03	86.43	100	5.13	31.71	100	4.78	5.23	5.06	5.14	5.46	100	4.87	93.16	100
$\mathcal{H}_1$	20	1.16	79.11	100	1.28	17.08	100	1.03	1.42	1.44	1.11	1.43	100	1.2	88.91	100
$\mathcal{H}_2$	20	1.21	79.52	100	1.43	17.58	100	1.13	1.48	2.91	1.26	1.56	100	1.32	89.1	100
$\mathcal{H}_3$	20	5.08	86.49	100	5.22	31.83	100	4.83	5.33	5.13	5.17	5.54	100	4.88	93.16	100
$\mathcal{RH}$	20	5.27	100	100	5.06	86.37	100	5.01	5.07	4.99	4.97	5.84	100	5.26	100	100

Table 2.2 (continued). Power of exogeneity tests at nominal level 5%; $G = 2$, $T = 500$

	k_2	$(a_1, a_2)' = (-20, 0)'$			$(a_1, a_2)' = (-5, 5)'$			$(a_1, a_2)' = (0, 0)'$			$(a_1, a_2)' = (.5, .2)'$			$(a_1, a_2)' = (100, 100)'$		
		$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$	$\eta_1 = 0$ $\eta_2 = .5$	$\eta_1 = .01$ $\eta_2 = .5$	$\eta_1 = .5$ $\eta_2 = .5$
$\mathcal{T}_1$	5	4.57	86.42	100	100	100	100	4.79	5.15	4.94	38.92	42.3	98.29	100	100	100
$\mathcal{T}_2$	5	4.79	100	100	100	100	100	5.12	5.07	4.8	84.67	87.89	100	100	100	100
$\mathcal{T}_3$	5	1.06	98.99	100	99.98	99.98	99.98	1.07	1.09	0.97	70.68	75.19	99.67	99.98	100	99.97
$\mathcal{T}_4$	5	4.77	100	100	100	100	100	5.1	5.06	4.79	84.58	87.83	100	100	100	100
$\mathcal{H}_1$	5	0.7	96.47	99.88	99.91	99.92	99.95	0.77	0.74	0.58	60.63	65.43	97.97	99.9	99.95	99.86
$\mathcal{H}_2$	5	1.06	99.01	100	99.98	99.98	98.99	1.09	1.11	0.98	70.85	75.34	99.68	99.98	100	99.97
$\mathcal{H}_3$	5	4.83	100	100	100	100	100	5.14	5.08	4.82	84.75	87.9	100	100	100	100
$\mathcal{RH}$	5	4.94	99.99	100	100	100	100	4.99	4.94	4.74	71.72	76.17	100	100	100	100
$\mathcal{T}_1$	10	5.24	99.95	100	100	100	100	4.57	4.8	5.28	77.23	81.26	100	100	100	100
$\mathcal{T}_2$	10	4.94	100	100	100	100	100	4.91	4.99	5.13	92.25	94.65	100	100	100	100
$\mathcal{T}_3$	10	1.6	100	100	100	100	100	1.65	1.82	1.76	89.62	92.87	100	100	100	100
$\mathcal{T}_4$	10	4.91	100	100	100	100	100	4.88	4.94	5.1	92.2	94.62	100	100	100	100
$\mathcal{H}_1$	10	0.86	99.87	100	100	100	100	0.89	1.03	0.74	82.77	86.64	99.9	100	100	100
$\mathcal{H}_2$	10	1.62	100	100	100	100	100	1.7	1.83	1.79	89.7	92.96	100	100	100	100
$\mathcal{H}_3$	10	4.96	100	100	100	100	100	4.96	5.06	5.15	92.25	94.65	100	100	100	100
$\mathcal{RH}$	10	4.5	100	100	100	100	100	5.06	4.88	4.89	72.21	77.47	100	100	100	100
$\mathcal{T}_1$	20	5.05	100	100	100	100	100	4.74	5.05	5.31	90.63	92.31	100	100	100	100
$\mathcal{T}_2$	20	5.18	100	100	100	100	100	4.77	4.88	5.16	95.23	96.41	100	100	100	100
$\mathcal{T}_3$	20	2.93	100	100	100	100	100	2.9	3.02	3.01	94.69	96.02	100	100	100	100
$\mathcal{T}_4$	20	5.15	100	100	100	100	100	4.74	4.86	5.12	95.19	96.39	100	100	100	100
$\mathcal{H}_1$	20	1.49	100	100	100	100	100	1.54	1.65	1.23	91.87	93.55	100	100	100	100
$\mathcal{H}_2$	20	2.98	100	100	100	100	100	2.96	3.06	3.04	94.72	96.04	100	100	100	100
$\mathcal{H}_3$	20	5.24	100	100	100	100	100	4.83	4.91	5.19	95.24	96.41	100	100	100	100
$\mathcal{RH}$	20	5.18	99.98	100	100	100	100	4.55	5.32	5.7	63.54	69.11	100	100	100	100

instrumental variables method should be applied for estimating the income-trade relationship. The equation studied is

$$y_i = \alpha + \beta Tr_i + \gamma_1 N_i + \gamma_2 Ar_i + u_i, \tag{8.1}$$

where y_i is log of income per capita in country i, Tr_i the trade share (measured as a ratio of imports and exports to GDP), N_i the logarithm of population, and Ar_i the logarithm of country area. Since the trade share Tr_i may be endogenous, Frankel and Romer (1999) used an instrument constructed on the basis of geographic characteristics. The first stage equation is given by

$$Tr_i = a + bX_i + c_1 N_i + c_2 Ar_i + v_i, \tag{8.2}$$

where X_i is a constructed instrument from geographic characteristics. In this paper, we use the sample of 150 countries and the data include for each country: the trade share in 1985, the area and population (1985), per capita income (1985), and the fitted trade share (instrument)[4]. In this application, we focus on testing whether trade share is exogenous in (8.1). However, it is not clear how "weak "instruments are in this model. In fact, the F-statistic in the first stage regression (8.2) is around 13 [see Frankel and Romer (1999, Table 2, p.385)], which may indicate a possible weak identification problem [Staiger-Stock(1997)]. Dufour and Taamouti (2006) proposed to use directly identification-robust procedures to draw inference on the coefficients of the model (5.5). The projection approach shows that there is a slight difference between the usual 95 % IV-type confidence sets and the 95 % AR-based confidence sets of the coefficients of the structural equation (8.1). The 95 % IV-type confidence interval for the trade share coefficient is $[-.01\,, 3.95]$, while the corresponding 95 % AR-based confidence set is $[.284\,, 4.652]$. However, since all the confidence sets are bounded, we do not have a serious problem of identification in this model. We provide an alternative way to access whether the instrument used is weak by examining the behaviour of DWH and $\mathcal{RH}$ statistics. For example, if the test for exogeneity based on these statistics does not

[4]The data set and its sources are given in the Appendix of Frankel and Romer (1999)

reject trade share exogeneity, this may indicate that instrument are not "very poor ". Note that the model contains only one endogenous and one excluded instrument, hence $k_2 = G$, and the statistic T_1 is not considered in this application because it is identically zero. Table 2.3 below summarizes the results. In the first column of the table, we report the statistics while in the second and third columns, we report the sample values and the sample p-value of these tests. In the other columns, we report the Monte Carlo tests p-values for two data generating process where the disturbances u are drawn from normal and Cauchy distributions.

Table 2.3. Tests for exogeneity of trade share in trade-income relation

Statistics	Sample value	Sample p-value (%)	MC-test p-value (normal distribution)	MC-test p-value (Cauchy distribution)
$\mathcal{RH}$	3.9221	4.95^*	5.02^*	2.74^*
$\mathcal{H}_1$	2.3883	12.23	6.15	2.93^*
$\mathcal{H}_2$	2.4269	11.93	6.12	2.94^*
$\mathcal{H}_3$	3.9505	4.67^*	5.49	2.85^*
$\mathcal{T}_2$	3.9221	4.95^*	5.49	2.85^*
$\mathcal{T}_3$	2.3622	12.43	6.12	2.94^*
$\mathcal{T}_4$	3.8451	4.99^*	5.49	2.85^*

Note –$*$: H_0 is rejected at nominal level $\alpha = 5\%$.

First, we note from Table 2.3 that $\mathcal{H}_3$, $\mathcal{T}_2$, $\mathcal{T}_4$ and $\mathcal{RH}$, reject trade share exogeneity while $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{T}_3$, cannot reject the null hypothesis. When we run exact Monte Carlo tests (for Gaussian and Cauchy type errors), we see that all statistics strongly reject trade share exogeneity at level 5 %, which means that the quality of the instrument is not too poor in this model as noted by Dufour and Taamouti (2006) . Our results also underscore the difference between exact Monte Carlo exogeneity procedures and earlier procedures.

8.2. Education and earnings

This application considers the well known problem of estimating returns to education. The literature in this issue includes Angrist and Krueger (1991), Angrist and Krueger (1995), Angrist and al.

(1999), Bound (1995). The equation studies is a relationship where the log weekly earning is explained by the number of years of education and several other covariates (age, age squared, year of birth, . . .). Since education can be viewed as an endogenous variable, Angrist and Krueger (1991) used the birth quarter as an instrument. The basic idea is that individuals born in the first quarter of the year start school at an older age, and can therefore drop out after completing less schooling than individuals born near the end of the year. Consequently, individuals born at the beginning of the year are likely to earn less than those born during the rest of the year[5]. However, it is well known that the instruments used by Angrist and Krueger (1991) are weak and explains very little of the variation in education; see Bound (1995). So, standard IV-based inference is quite unreliable. As showed in this paper, DWH or RH tests for the exogeneity of education will lead to accept the null hypothesis of exogeneity of this variable. The model considered is specified as:

$$y = \beta_0 + \beta_1 E + \sum_{i=1}^{k_1} \gamma_i X_i + u\,, \tag{8.3}$$

$$E = \pi_0 + \sum_{i=1}^{k_2} \pi_i Z_i + \sum_{i=1}^{k_1} \phi_i X_i + v\,, \tag{8.4}$$

where y is log-weekly earnings, E is the number of years of education (possibly endogenous), X contains the exogenous covariates (age, age squared, 10 dummies for birth of year). Z contains 40 dummies obtained by interacting the quarter of birth with the year of birth. In this model, β_1 measures the return to education. The data set consists of the 5% public-use sample of the 1980 US census for men born between 1930 and 1939. The sample size is 329 509 observations. We test the exogeneity of education in this model using DWH and RH statistics. The results are summarized in Table 2.4. As showed in this table, all exogeneity tests cannot reject the exogeneity of “education ”even at level 15%. This is true for earlier versions of the tests or the MCE-tests.

The results can be interpreted as follow: (a) either the instruments are strong and education

[5]Other versions of the IV regression take as instruments interactions between the birth quarter and regional and/or birth year dummies.

is effectively exogenous, (b) or education is endogenous but the instruments are too poor and the tests fail to detect that education is endogenous. Moreover, it is well documented that the generated instruments obtained by interacting the quarter of birth with the year of birth are weak, see e.g., Bound (1995). So, our interpretation in (b) matter with these observations.

Table 2.4. Tests for exogeneity of education in income-education equation.

Statistics	Sample value	Sample p-value	MC-test p-value (normal distribution)	MC-test p-value (Cauchy distribution)
$\mathcal{RH}$	.6783	.93986	.6590	.9451
$\mathcal{H}_1$	1.337	.24757	.2474	.2488
$\mathcal{H}_2$	1.337	.24756	.2474	.2488
$\mathcal{H}_3$	1.3492	.24542	.2474	.2488
$\mathcal{T}_1$	2.0406	.16111	.2302	.2308
$\mathcal{T}_2$	1.3491	.24543	.2474	.2488
$\mathcal{T}_3$	1.3369	. 224757	.2474	.2488
$\mathcal{T}_4$	1.3491	.24543	.2474	.2488

9. Conclusion

In this paper, we focus on linear structural models and propose a finite-and large-sample analysis of the distribution of Durbin (1954), Wu (1973), Hausman (1978) and, Revankar and Hartley (1973) specification tests under the null hypothesis (level) and the alternative hypothesis (power), including when identification is deficient or weak (weak IV). Our finite-sample analysis provides several new insights and extensions of earlier procedures. First, the characterization of the finite-sample distribution of DWH and RH test statistics allows the construction of exact identification-robust exogeneity tests even with non-Gaussian errors (Monte Carlos tests). Second, we show that DWH- and RH-type tests are typically robust to weak IV (level is controlled). We provide a characterization of the power of the tests, which clearly exhibits the factors which determine power. We show that the tests have no power when all IV are weak [similar to Guggenberger (2008)]. But, power does exist as soon as we have one strong IV. The conclusions of Guggenberger (2008) focus on the case where all IV are weak (a case of little practical interest).

Our asymptotic distributional theory under weaker assumptions confirms the finite- sample theory. We present simulation evidence indicating that: (1) Over a wide range cases, including weak IV and moderate endogeneity, OLS performs better than 2SLS [Similar to Kiviet and Niemczyk (2007)]; (2) pretest-estimators based on exogeneity tests have an excellent overall performance compared with OLS and IV estimators. We illustrate our theoretical results through two empirical applications: the returns to education and the relation between trade and economic growth. We find that exogeneity tests cannot reject the exogeneity of schooling, *i.e.* the IV are weak in this model [Bound (1995)]. However, “trade share ”is endogenous, *i.e.* the IV are not too poor [similar to Dufour and Taamouti (2007)].

APPENDIX

A. Notes

1.1. Unified formulation of DWH test statistics

In this Appendix, we establish alternative formula of Durbin-Wu statistics in (3.1) - (3.13). From Wu (1973, Econemetrica, Eqs. (2.1), (2.18), (3.16), (3.20)), the statistics $T_l,\ l = 1,\,2,\,3,\,4$ are defined as

$$\mathcal{T}_1 = \frac{Q^*/G}{Q_1/(k_2-G)},\ \mathcal{T}_2 = \frac{Q^*/G}{Q_2/(T-k_1-2G)}, \tag{A.1}$$

$$\mathcal{T}_1 = \frac{Q^*/G}{Q_3/(T-k_1-G)},\ \mathcal{T}_2 = \frac{Q^*/G}{Q_4/(T-k_1-G)}, \tag{A.2}$$

where

$$Q^* = (b_1-b_2)'\left[(Y'A_2Y)^{-1}-(Y'A_1Y)^{-1}\right]^{-1}(b_1-b_2), \tag{A.3}$$

$$Q_1 = (y-Yb_2)'A_2(y-Yb_2),\ Q_2 = Q_4 - Q^*, \tag{A.4}$$

$$Q_4 = (y-Yb_1)'A_1(y-Yb_1),\ Q_3 = (y-Yb_2)'A_1(y-Yb_2), \tag{A.5}$$

$$b_i = (Y'A_iY)^{-1}Y'A_iy,\ i = 1,\,2, \tag{A.6}$$

$$A_1 = I - Z_1(Z_1'Z_1)^{-1}Z_1',\ A_2 = Z(Z'Z)^{-1}Z' - Z_1(Z_1'Z_1)^{-1}Z_1', \tag{A.7}$$

$$Z = [Z_1, Z_2]\,. \tag{A.8}$$

Clearly, in the above notations, b_1 is the ordinary least squares estimator of β, and b_2 is the instrumental variables method estimator of β. So, from the notations of this paper, $b_1 \equiv \hat{\beta}$ and $b_2 \equiv \tilde{\beta}$. It is also easy to see that $M_1 = A_1$ and $M_1 - M = A_2$.

So, from (3.5) - (3.13), we can observe that

$$Q^* = T\,(\tilde{\beta}-\hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta}-\hat{\beta}) = T\tilde{\sigma}^2(\tilde{\beta}-\hat{\beta})'\hat{\Sigma}_2^{-1}(\tilde{\beta}-\hat{\beta}), \tag{A.9}$$

$$Q_1 = T\,\tilde{\sigma}_1^2\,, \quad Q_3 = T\,\tilde{\sigma}^2\,, \quad Q_4 = T\hat{\sigma}^2\,, \tag{A.10}$$

$$Q_2 = Q_4 - Q^* = T\hat{\sigma}^2 - T\,(\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) = T\,\tilde{\sigma}_2^2\,. \tag{A.11}$$

Hence, the Durbin-Wu test statistics can be written as

$$\mathcal{T}_l \;=\; \kappa_l(\tilde{\beta} - \hat{\beta})'\tilde{\Sigma}_l^{-1}(\tilde{\beta} - \hat{\beta})\,, \quad l = 1,\,2,\,3,\,4\,, \tag{A.12}$$

where $\kappa_l,$ and $\tilde{\Sigma}_l,\ l = 1,\,2,\,3,\,4,$ are defined in (3.1) - (3.13).

1.2. Regression interpretation of DWH test statistics

Consider the equations (3.17) - (3.19). First, we note that H_0 and H_b can be written as

$$\begin{aligned} H_0 &: \; R\theta = 0, \textit{i.e.}\; Rb = a\,, \\ H_b &: \; R_*\theta_* = 0, \textit{i.e.}\; R_*\theta_* = \beta - a\,, \end{aligned}$$

where $R = [0,\,0,\,I_G]$ and $R_* = [I_G,\,0,\,-I_G]$. By definition, we have $\hat{\theta}_* = [\tilde{\beta}',\,\tilde{\gamma}',\,\tilde{b}']'$ and $\hat{\theta}_{*0} = [\hat{\beta}',\,\hat{\gamma}',\,\hat{\beta}']'$, where $\tilde{\beta}$ and $\tilde{\gamma}$ are the 2SLS estimates of β and γ and $\hat{\beta}$ and $\hat{\gamma}$ are the OLS estimates of β and γ based on the model

$$\begin{aligned} y &= Y\beta + Z_1\gamma + u \\ \hat{Y} &= Z\hat{\Pi}, \end{aligned}$$

with $\hat{\Pi} = (Z'Z)^{-1}Z'Y$. So, we can observe that

$$\begin{aligned} \hat{\theta}_{*0} &= \hat{\theta}_* + (\hat{X}'\hat{X})^{-1}R_*'\left[R_*(\hat{X}'\hat{X})^{-1}R_*'\right]^{-1}(-R_*\hat{\theta}_*) \\ S(\hat{\theta}_{*0}) - S(\hat{\theta}_*) &= (\hat{\theta}_{*0} - \hat{\theta}_*)'\hat{X}'\hat{X}(\hat{\theta}_{*0} - \hat{\theta}_*) \\ &= (R_*\hat{\theta}_*)'\left[R_*(\hat{X}'\hat{X})^{-1}R_*'\right]^{-1}(R_*\hat{\theta}_*). \end{aligned}$$

Furthermore, we have

$$\begin{aligned}
R_*\hat{\theta} &= \left[\begin{array}{ccc} I_G & 0 & -I_G \end{array}\right] \left[\begin{array}{c} \tilde{\beta} \\ \tilde{\gamma} \\ \tilde{b} \end{array}\right] = \tilde{\beta} - \tilde{b}\,, \\
\hat{X}'\hat{X} &= \left[\begin{array}{cc} (\hat{X}_1'\hat{X}_1) & 0 \\ 0 & (\hat{V}'\hat{V}) \end{array}\right],\ (\hat{X}'\hat{X})^{-1} = \left[\begin{array}{cc} (\hat{X}_1'\hat{X}_1)^{-1} & 0 \\ 0 & (\hat{V}'\hat{V})^{-1} \end{array}\right], \\
(\hat{X}_1'\hat{X}_1)^{-1} &= \left[\begin{array}{cc} \hat{Y}'\hat{Y} & \hat{Y}'Z_1 \\ Z_1'\hat{Y} & Z_1'Z_1 \end{array}\right]^{-1} = \left[\begin{array}{cc} M_{11} & M_{12} \\ M_{21} & M_{22} \end{array}\right],
\end{aligned}$$

where

$$\begin{aligned}
M_{11} &= \left[(\hat{Y}'\hat{Y}) - \hat{Y}'Z_1(Z_1'Z_1)^{-1}Z_1'\hat{Y}\right]^{-1} \\
&= \left[\hat{Y}'(I - Z_1(Z_1'Z_1)^{-1}Z_1')\hat{Y}\right]^{-1} = \left[\hat{Y}'M_1\hat{Y}\right]^{-1} = [Y'(M_1 - M)Y]^{-1}, \\
M_1 &= I - Z_1(Z_1'Z_1)^{-1}Z_1',\ M = I - Z(Z'Z)^{-1}Z'.
\end{aligned}$$

So, we have

$$\begin{aligned}
(\hat{X}'\hat{X})^{-1}R_*' &= \left[\begin{array}{ccc} M_{11} & M_{12} & 0 \\ M_{21} & M_{22} & 0 \\ 0 & 0 & (\hat{V}'\hat{V})^{-1} \end{array}\right] \left[\begin{array}{c} I_G \\ 0 \\ -I_G \end{array}\right] = \left[\begin{array}{c} M_{11} \\ M_{21} \\ -(\hat{V}'\hat{V})^{-1} \end{array}\right] \\
R_*(\hat{X}'\hat{X})^{-1}R_*' &= M_{11} + (\hat{V}'\hat{V})^{-1} \\
\hat{\theta}_{*0} - \hat{\theta}_* &= \left[\begin{array}{c} \hat{\beta} - \tilde{\beta} \\ \hat{\gamma} - \tilde{\gamma} \\ \hat{\beta} - \tilde{b} \end{array}\right] = \left[\begin{array}{c} M_{11} \\ M_{21} \\ -(\hat{V}'\hat{V})^{-1} \end{array}\right] \left[M_{11} + (\hat{V}'\hat{V})^{-1}\right]^{-1} (\tilde{b} - \tilde{\beta}).
\end{aligned}$$

Hence

$$\hat{\beta} - \tilde{\beta} = M_{11}\left[M_{11} + (\hat{V}'\hat{V})^{-1}\right]^{-1}(\tilde{b} - \tilde{\beta}) \quad \text{(A.13)}$$

$$= M_{11}\left[M_{11} + (\hat{V}'\hat{V})^{-1}\right]^{-1}\tilde{a}\,, \quad \text{(A.14)}$$

where $\tilde{a} = \tilde{b} - \tilde{\beta}$ is the OLS estimate of a from (3.17). We see from (A.14) that

$$\tilde{a} = \tilde{b} - \tilde{\beta} = \left[M_{11} + (\hat{V}'\hat{V})^{-1}\right] M_{11}^{-1}(\hat{\beta} - \tilde{\beta}) \quad \text{(A.15)}$$

$$= \left\{[Y'(M_1 - M)Y]^{-1} + (\hat{V}'\hat{V})^{-1}\right\} [Y'(M_1 - M)Y](\hat{\beta} - \tilde{\beta})\,. \quad \text{(A.16)}$$

So, we have

$$
\begin{aligned}
S(\hat{\theta}_{*0}) - S(\hat{\theta}_*) &= (R_*\hat{\theta}_*)' \left[R_*(\hat{X}'\hat{X})^{-1}R_*'\right]^{-1}(R_*\hat{\theta}_*) \\
&= (\tilde{b}-\tilde{\beta})' \left\{[Y'(M_1-M)Y]^{-1} + (\hat{V}'\hat{V})^{-1}\right\}^{-1}(\tilde{b}-\tilde{\beta}) \\
&= (\hat{\beta}-\tilde{\beta})'[Y'(M_1-M)Y]\left\{[Y'(M_1-M)Y]^{-1} + (\hat{V}'\hat{V})^{-1}\right\} \times \\
&\quad [Y'(M_1-M)Y](\hat{\beta}-\tilde{\beta}) \\
&= (\hat{\beta}-\tilde{\beta})'M_{11}^{-1}\left[M_{11} + (Y'MY)^{-1}\right]M_{11}^{-1}(\hat{\beta}-\tilde{\beta}) \\
&= (\hat{\beta}-\tilde{\beta})'M_{11}^{-1}\left[M_{11} + (Y'M_1Y - M_{11}^{-1})^{-1}\right]M_{11}^{-1}(\hat{\beta}-\tilde{\beta})\,. \qquad \text{(A.17)}
\end{aligned}
$$

We will now use the following lemma.

Lemma A.1 *Let A and B be two nonsingular $r \times r$ matrices. Then*

$$
\begin{aligned}
A^{-1} - B^{-1} &= B^{-1}(B-A)A^{-1} \\
&= A^{-1}(B-A)B^{-1} \\
&= A^{-1}(A - AB^{-1}A)A^{-1} \\
&= B^{-1}(BA^{-1}B - B)B^{-1}.
\end{aligned}
$$

Furthermore, if $B - A$ is nonsingular, then $A^{-1} - B^{-1}$ is nonsingular with

$$
\begin{aligned}
(A^{-1} - B^{-1})^{-1} &= A(B-A)^{-1}B = A + A(B-A)^{-1}A = A[A^{-1} + (B-A)^{-1}]A \\
&= B(B-A)^{-1}A = B(B-A)^{-1}B - B = B[(B-A)^{-1} - B^{-1}]B \\
&= A(A - AB^{-1}A)^{-1}A \\
&= B(BA^{-1}B - B)^{-1}B.
\end{aligned}
$$

The proof of Lemma **A.1** is not difficult and is omitted. Now, by setting $A = M_{11}^{-1}$ and $B =$

$Y'M_1Y$ in (A.17) and using Lemma **A.1**, we get

$$\begin{aligned} S(\hat{\theta}_{*0}) - S(\hat{\theta}_*) &= (\hat{\beta}-\tilde{\beta})'M_{11}^{-1}\left[M_{11} + (Y'M_1Y - M_{11}^{-1})^{-1}\right]M_{11}^{-1}(\hat{\beta}-\tilde{\beta}) \\ &= (\hat{\beta}-\tilde{\beta})'A\left[A^{-1} + (B-A)^{-1}\right]A(\hat{\beta}-\tilde{\beta}) = (\hat{\beta}-\tilde{\beta})'(B^{-1}-A^{-1})^{-1}(\hat{\beta}-\tilde{\beta}) \\ &= (\hat{\beta}-\tilde{\beta})'\{[Y'(M_1-M)Y]^{-1} - (Y'M_1Y)^{-1}\}^{-1}(\hat{\beta}-\tilde{\beta}) \\ &= \frac{1}{T}(\tilde{\beta}-\hat{\beta})'[\hat{\Omega}_{IV}^{-1} - \hat{\Omega}_{LS}^{-1}]^{-1}(\tilde{\beta}-\hat{\beta}) = \frac{1}{T}(\tilde{\beta}-\hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta}-\hat{\beta}), \end{aligned} \tag{A.18}$$

where $\hat{\Omega}_{IV} = \frac{1}{T}Y'\bar{M}_1Y$ and $\hat{\Omega}_{LS} = \frac{1}{T}Y'M_1Y$. Note also that

$$S(\hat{\theta}_{*0}) - S(\hat{\theta}_*) = S(\hat{\theta}_0) - S(\hat{\theta}) = \tilde{a}'[\hat{V}'M_X\hat{V}]\tilde{a}, \tag{A.19}$$

$M_X = I - P_X = I - X(X'X)^{-1}X'$, $X = [Y,\, Z_1,\, \hat{V}]$. Moreover, from (3.20), we have

$$S(\hat{\theta}) = T\tilde{\sigma}_2^2,\; S(\hat{\theta}_0) = T\hat{\sigma}^2,\; S_*(\hat{\theta}_*^0) = T\tilde{\sigma}^2, \tag{A.20}$$

where $\tilde{\sigma}^2$, $\hat{\sigma}^2$, $\tilde{\sigma}_2^2$ are defined in (3.9) - (3.13). So, except H_1, the other statistics $H_i,\ i = 2,\,3$ and $T_l,\ l = 1,\,2,\,3,\,4$ can be expressed as follow:

$$\mathcal{H}_2 = \frac{S(\hat{\theta}_0) - S(\hat{\theta})}{S_*(\hat{\theta}_*^0)/T}, \quad \mathcal{H}_3 = \frac{S(\hat{\theta}_0) - S(\hat{\theta})}{S(\hat{\theta}_0)/T}, \tag{A.21}$$

$$\mathcal{T}_1 = \frac{[S(\hat{\theta}_0) - S(\hat{\theta})]/G}{Q_1/(k_2 - G)}, \quad \mathcal{T}_2 = \frac{[S(\hat{\theta}_0) - S(\hat{\theta})]/G}{S(\hat{\theta})/(T - k_1 - 2G)}, \tag{A.22}$$

$$\mathcal{T}_3 = \frac{S(\hat{\theta}_0) - S(\hat{\theta})}{S_*(\hat{\theta}_*^0)/(T - k_1 - G)}, \quad \mathcal{T}_4 = \frac{S(\hat{\theta}_0) - S(\hat{\theta})}{S(\hat{\theta}_0)/(T - k_1 - G)}, \tag{A.23}$$

where $Q_1 = T\tilde{\sigma}_1^2$, $\tilde{\sigma}_1^2$ is also defined in (3.9) - (3.13). Equations (A.21) - (A.23) are the regression interpretation of DWH test statistics.

1.3. Reduced form model in terms of orthogonal pair

Let

$$\bar{Z} = [Z_1,\, \bar{Z}_2]\,, \quad \bar{Z}_2 = M_1 Z_2\,, \quad M_{\bar{Z}} = I - \bar{Z}(\bar{Z}'\bar{Z})^{-1}\bar{Z}'. \tag{A.24}$$

Then, from (2.1) - (2.2), the reduced form for $[y, Y]$ can be written in terms of the orthogonal pair $[Z_1, \bar{Z}_2]$ as

$$y = Z_1\pi_1 + \bar{Z}_2\pi_2 + v\,, \tag{A.25}$$

$$Y = Z_1\bar{\Pi}_1 + \bar{Z}_2\Pi_2 + V\,, \tag{A.26}$$

where $\pi_1 = \gamma + \bar{\Pi}_1\beta,\ \pi_2 = \Pi_2\beta,\ v = u + V\beta = [v_1, \ldots, v_T]'$, $\bar{\Pi}_1 = \Pi_1 + (Z_1'Z_1)^{-1}Z_1'Z_2\Pi_2$. Since $\bar{Z}_2$ is orthogonal with Z_1 (i.e., $Z_1'\bar{Z}_2 = 0$), we have:

$$M_{\bar{Z}} = I - Z_1(Z_1'Z_1)^{-1}Z_1' - \bar{Z}_2(\bar{Z}_2'\bar{Z}_2)^{-1}\bar{Z}_2'\,, \tag{A.27}$$

$$M_1 - M_{\bar{Z}} = \bar{Z}_2(\bar{Z}_2'\bar{Z}_2)^{-1}\bar{Z}_2' = M_1 M_{\bar{Z}} M_1. \tag{A.28}$$

By noting that the projection matrix M defined in (2.7) can be decomposed as

$$M = M_1 - P_{M_1 Z_2} = M_1 - P_{\bar{Z}_2} = M_1 - \bar{Z}_2(\bar{Z}_2'\bar{Z}_2)^{-1}\bar{Z}_2', \tag{A.29}$$

and using (A.27) - (A.28), we have

$$M = M_{\bar{Z}}. \tag{A.30}$$

So, if we replace Z by $\bar{Z}$ in (3.1) - (3.13), then the statistics H_i $(i = 1, 2, 3)$, T_l $(l = 1, 2, 3, 4)$ and RH do not change. Thus, using (A.25) - (A.26) instead of (2.1) - (2.2) does not restrict our results.

B. Proofs

PROOF OF THEOREM **4.1** Note first that

$$\begin{aligned} \tilde{\beta} &= \beta + [Y'(M_1 - M)Y]^{-1} Y'(M_1 - M)u = \beta + \bar{A}_1 u, \\ & \quad \bar{A}_1 = [Y'(M_1 - M)Y]^{-1} Y'(M_1 - M)\,, \qquad \text{(B.1)} \\ \hat{\beta} &= \beta + (Y'M_1Y)^{-1}Y'M_1u = \beta + A_1 u, \quad A_1 = (Y'M_1Y)^{-1}Y'M_1 \qquad \text{(B.2)} \end{aligned}$$

so that

$$\begin{aligned} \tilde{\beta} - \hat{\beta} &= (\bar{A}_1 - A_1)u, \quad \hat{\Delta} = \hat{\Omega}_{IV}^{-1} - \hat{\Omega}_{LS}^{-1}\,, \qquad \text{(B.3)} \\ (\tilde{\beta} - \hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta} - \hat{\beta}) &= u'C_0u, \quad C_0 = (\bar{A}_1 - A_1)'\hat{\Delta}^{-1}(\bar{A}_1 - A_1)\,, \qquad \text{(B.4)} \\ \hat{\Omega}_{IV} &= \frac{1}{T}Y'(M_1 - M)Y, \quad \hat{\Omega}_{LS} = \frac{1}{T}Y'M_1Y\,. \qquad \text{(B.5)} \end{aligned}$$

By proceeding as above, we also get

$$\begin{aligned} M_1(y - Y\tilde{\beta}) &= \bar{B}_1 u, \\ \bar{B}_1 &= M_1 - P_{(M_1-M)Y} = M_1(I - P_{(M_1-M)Y}) = M_1 M_{(M_1-M)Y}\,, \qquad \text{(B.6)} \\ M(y - Y\tilde{\beta}) &= Mu - MY\bar{A}_1u = Mu - MM_1Y\bar{A}_1u \qquad \text{(B.7)} \\ &= Mu - MP_{(M_1-M)Y}u = MM_{(M_1-M)Y}u\,, \end{aligned}$$

where for any matrix B, $P_B = B(B'B)^{-1}B'$ and $M_B = I - P_B$. So, we have

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{1}{T}u'M_1M_{(M_1-M)Y}u = u'\bar{D}_1u, \quad \hat{\sigma}^2 = \frac{1}{T}u'M_1M_{M_1Y}u = u'D_1u\,, \qquad \text{(B.8)} \\ \tilde{\sigma}_1^2 &= \tilde{\sigma}^2 - \hat{\sigma}^2 = u'(\bar{D}_1 - D_1)u = \frac{1}{T}u'(M_1 - M)M_{(M_1-M)Y}u\,, \qquad \text{(B.9)} \\ \tilde{\sigma}_2^2 &= \frac{1}{T}u'M_1M_{M_1Y}u - u'C_0u = u'(D_1 - C_0)u\,. \qquad \text{(B.10)} \end{aligned}$$

Now, it is easy to see from (B.1) - (B.10) that:

$$\mathcal{H}_2 = \frac{T}{\tilde{\sigma}^2}(\tilde{\beta}-\hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta}-\hat{\beta}) = T\frac{u'C_0u}{u'\bar{D}_1u} = T\frac{(u/\sigma_u)'C_0(u/\sigma_u)}{(u/\sigma_u)'\bar{D}_1(u/\sigma_u)}\,, \tag{B.11}$$

$$\mathcal{H}_3 = \frac{T}{\hat{\sigma}^2}(\tilde{\beta}-\hat{\beta})'\hat{\Delta}^{-1}(\tilde{\beta}-\hat{\beta}) = T\frac{u'C_0u}{u'D_1u} = T\frac{(u/\sigma_u)'C_0(u/\sigma_u)}{(u/\sigma_u)'D_1(u/\sigma_u)}\,, \tag{B.12}$$

$$\mathcal{T}_1 = \kappa_1\frac{u'C_0u}{u'(\bar{D}_1-D_1)u} = \kappa_1\frac{(u/\sigma_u)'C_0(u/\sigma_u)}{(u/\sigma_u)'(\bar{D}_1-D_1)(u/\sigma_u)}\,, \tag{B.13}$$

$$\mathcal{T}_2 = \kappa_2\frac{u'C_0u}{u'(D_1-C_0)u} = \kappa_2\frac{(u/\sigma_u)'C_0(u/\sigma_u)}{(u/\sigma_u)'(D_1-C_0)(u/\sigma_u)}\,, \tag{B.14}$$

$$\mathcal{T}_3 = (\kappa_3/T)\mathcal{H}_2 = \kappa_3\frac{(u/\sigma_u)'C_0(u/\sigma_u)}{(u/\sigma_u)'\bar{D}_1(u/\sigma_u)}\,, \tag{B.15}$$

$$\mathcal{T}_4 = (\kappa_4/T)\mathcal{H}_3 = \kappa_4\frac{(u/\sigma_u)'C_0(u/\sigma_u)}{(u/\sigma_u)'D_1(u/\sigma_u)}\,. \tag{B.16}$$

Suppose now that H_0 holds, i.e. Y is independent of u. Then, since the matrices $C_0,\, D_1,\, \bar{D}_1$ depend only on $\bar{X} = [Y\,,\, Z_1\,,\, Z_2]$, we see that the conditional distribution given $\bar{X}$ of the above statistics in (B.11) - (B.16) depend only on the distribution of u/σ_u irrespective of the rank of Π_2. It is also easy to see that the same arguments hold for the statistic H_1. For the statistic R, remark that under H_0^*, we have

$$\mathcal{RH} = \frac{u'(M_{X_1}-M_{\bar{X}})u/k_2}{u'M_{\bar{X}}u/(T-k-G)} = \frac{(u/\sigma_u)'(M_{X_1}-M_{\bar{X}})(u/\sigma_u)/k_2}{(u/\sigma_u)'M_{\bar{X}}(u/\sigma_u)/(T-k-G)}\,. \tag{B.17}$$

Let consider the following decomposition of the matrices M_{X_1} and $M_{\bar{X}}$:

$$\frac{1}{T}M_{X_1} = \frac{1}{T}(M_1 - P_{M_1Y}) = D_1,\; \frac{1}{T}M_{\bar{X}} = \frac{1}{T}(M_{X_1} - P_{M_{X_1}Z_2}) = D_1 - P_{D_1Z_2}\,, \tag{B.18}$$

where for any matrix B, $P_B = B(B'B)^{-1}B'$. Thus, we have

$$\frac{1}{T}(M_{X_1} - M_{\bar{X}}) = P_{D_1Z_2}\,, \tag{B.19}$$

and

$$\mathcal{RH} = \frac{(u/\sigma_u)'P_{D_1Z_2}(u/\sigma_u)/k_2}{(u/\sigma_u)'(D_1 - P_{D_1Z_2})(u/\sigma_u)/(T-k-G)}\,. \tag{B.20}$$

So, the conditional distribution of R given $\bar{X}$ only depends on u/σ_u whether Rank$(\varPi_2) < G$ or not. □

PROOF OF THEOREM **4.2** Let consider the expression of the statistics $[H_i,\, i = 1,\, 2,\, 3]$, $[T_l,\, l = 1,\, 2,\, 3,\, 4]$ and R in (B.11) - (B.17). It is important to note that these expressions follow only by algebraic arguments. By replacing u by $Va + \varepsilon$ in these expressions, we get

$$\mathcal{H}_1 = T(Va+\varepsilon)'(\bar{A}_1 - A_1)'\Sigma_1^{-1}(\bar{A}_1 - A_1)(Va+\varepsilon)\,, \tag{B.21}$$

$$\mathcal{H}_2 = T\frac{(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)}\,, \tag{B.22}$$

$$\mathcal{H}_3 = T\frac{(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'D_1(Va+\varepsilon)}\,, \tag{B.23}$$

$$\mathcal{T}_1 = \frac{\kappa_1(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'(\bar{D}_1 - D_1)(Va+\varepsilon)}\,, \tag{B.24}$$

$$\mathcal{T}_2 = \frac{\kappa_2(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'(D_1 - C_0)(Va+\varepsilon)}\,, \tag{B.25}$$

$$\mathcal{T}_3 = \frac{\kappa_3(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)}\,, \tag{B.26}$$

$$\mathcal{T}_4 = \frac{\kappa_4(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'D_1(Va+\varepsilon)}\,, \tag{B.27}$$

and

$$\mathcal{RH} = \frac{(Va+\varepsilon)'(M_{X_1} - M_{\bar{X}})(Va+\varepsilon)/k_2}{(Va+\varepsilon)'M_{\bar{X}}(Va+\varepsilon)/(T-k-G)}\,. \tag{B.28}$$

□

PROOF OF LEMMA **4.3** If $\varPi_2 = 0$, we have

$$M_1 Y = M_1 V,\; (M_1 - M)Y = (M_1 - M)V.$$

So, we can easily see from the proof of Theorem **4.2** above that

$$(\bar{A}_1 - A_1)Y \;=\; (\bar{A}_1 - A_1)V = 0,\; C_0 Y = C_0 V = 0,\; \bar{D}_1 Y = \bar{D}_1 V = 0,\; D_1 Y = D_1 V = 0$$

so that the non centrality parameters in (B.21) - (B.27) vanish. Furthermore, by considering the decomposition of the matrices M_{X_1} and $M_{\bar{X}}$ in (B.18) - (B.19), we find

$$M_{X_1} V \;=\; D_1 V = 0,\; M_{\bar{X}} V = D_1 V - P_{D_1 Z_2} V = 0$$

and the non centrality parameters in (B.28) also vanishes. □

PROOF OF THOEREM **4.4** First, we note from (B.9) that with probability one,

$$\bar{D}_1 - D_1 \;>\; 0\,.$$

So, we have $(Va+\varepsilon)'\bar{D}_1(Va+\varepsilon) > (Va+\varepsilon)'D_1(Va+\varepsilon)$ and from (B.22)-(B.23), it is clear that $H_2 \leq H_3$. Furthermore, we have

$$\begin{aligned} \varSigma_1 &= (Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)\hat{\varOmega}_{IV}^{-1} - (Va+\varepsilon)'D_1(Va+\varepsilon)\hat{\varOmega}_{LS}^{-1} \\ &\geq (Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)[\hat{\varOmega}_{IV}^{-1} - \hat{\varOmega}_{LS}^{-1}] = (Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)\hat{\varDelta}\,, \end{aligned}$$

hence, $(\bar{A}_1 - A_1)'\varSigma_1^{-1}(\bar{A}_1 - A_1) \leq (Va+\varepsilon)'\bar{D}_1(Va+\varepsilon)C_0$ and

$$0 \;\leq\; \mathcal{H}_1 \leq \mathcal{H}_2 \leq \mathcal{H}_3.$$

Second, we have

$$\mathcal{H}_3 \;=\; \frac{T(Va+\varepsilon)'C_0(Va+\varepsilon)}{(Va+\varepsilon)'D_1(Va+\varepsilon)}\,. \tag{B.29}$$

If $\varPi_2 \neq 0$, then when $\|a\| \to +\infty$, we have $Va+\varepsilon \approx Va$ and

$$\mathcal{H}_3 \;\equiv\; T\frac{a'V'C_0Va}{a'V'D_1Va}\,.$$

We note that C_0 and D_1 are symmetric idempotent and if V is a full columns rank matrix with probability one, then given Z and V, C_0V and $(D_1 - C_0)V$ are also full columns matrices. Thus, $V'C_0^2V = V'C_0V > 0$ and $V'D_1^2V = V'D_1V > 0$. So, for any $a \neq 0$, we have

$$0 \;<\; \lambda_1 \leq \frac{a'V'C_0Va}{a'V'D_1Va} \leq \lambda_2 < +\infty\,,$$

where λ_1 and λ_2 are the smallest and largest solution of

$$|V'C_0V - \lambda V'D_1V| \;=\; 0\,, \tag{B.30}$$

see, Anderson (2003, Theorem A.2.4). So, we have

$$0 \;\leq\; \lim_{\|a\|\to\infty} \mathcal{H}_1 \leq \lim_{\|a\|\to\infty} \mathcal{H}_2 \leq \lim_{\|a\|\to\infty} \mathcal{H}_3 \leq T\lambda_2 < +\infty\,. \tag{B.31}$$

If $\mathit{\Pi}_2 = 0$, the expressions of the statistics do not depend on a and the results hold. By following the same steps as before, we get the results for the statistics $T_l,\, l = 1, 2, 3, 4,$ and RH. □

Proof of Lemma **4.5** To simplify the proof, we show only the invariance of the statistic H_3 to the transformation P. The argument is similar for the other statistics. From (3.2) - (3.13)

$$\mathcal{H}_3 = T(\tilde{\beta} - \hat{\beta})'\hat{\mathit{\Sigma}}_3^{-1}(\tilde{\beta} - \hat{\beta})\,, \tag{B.32}$$

where $\hat{\beta} = (Y'M_1Y)^{-1}Y'M_1y$ is the ordinary least squares (OLS) estimator of β, $\tilde{\beta} = [Y'(M_1 - M)Y]^{-1}Y'(M_1 - M)y$ is the two-stage least squares (2SLS) estimator of β,

$$\hat{\mathit{\Sigma}}_3 = \hat{\sigma}^2\big[\,\big(Y'(M_1 - M)Y/T\big)^{-1} - \big(Y'M_1Y/T\big)^{-1}\,\big], \tag{B.33}$$

$$\hat{\sigma}^2 = (y - Y\hat{\beta})'M_1(y - Y\hat{\beta})/T\,. \tag{B.34}$$

Let us now replace y and Y by $y_* = yR_{11} + YR_{21}$ and $Y_* = YR_{22}$ in (B.32) - (B.34). This yields:

$$\mathcal{H}_{3*} = T(\tilde{\beta}_* - \hat{\beta}_*)'\hat{\mathit{\Sigma}}_{*3}^{-1}(\tilde{\beta}_* - \hat{\beta}_*) \tag{B.35}$$

where $\hat{\beta}_* = (Y_*' M_1 Y_*)^{-1} Y_*' M_1 y_*$ and $\tilde{\beta}_* = (Y_*'(M_1 - M)Y_*)^{-1} Y_*'(M_1 - M)y_*$,

$$\hat{\Sigma}_{3*} = \hat{\sigma}_*^2 \big[\big(Y_*'(M_1 - M)Y_*/T\big)^{-1} - \big(Y_*' M_1 Y_*/T\big)^{-1} \big],$$

$$\hat{\sigma}_*^2 = (y_* - \bar{Y}\hat{\beta}_*)' M_1 (y_* - \bar{Y}\hat{\beta}_*)/T\,.$$

Since $y_* = yR_{11} + YR_{21}$ and $Y_* = YR_{22}$, we have:

$$\begin{aligned} Y_*' M_1 Y_* &= R_{22}' Y' M_1 Y R_{22} = R_{22}' Y' M_1 Y R_{22}\,, \\ Y_*' M_1 y_* &= R_{22}'(Y' M_1 y R_{11} + Y' M_1 Y R_{21})\,, \end{aligned} \tag{B.36}$$

hence

$$\hat{\beta}_* = R_{22}^{-1}(Y' M_1 Y)^{-1}(R_{22}^{-1})' R_{22}'(Y' M_1 y R_{11} + Y' M_1 Y R_{21}) = R_{22}^{-1}(\hat{\beta} R_{11} + R_{21}).$$

Similarly for $\tilde{\beta}_*$, we get

$$\tilde{\beta}_* = (Y_*'(M_1 - M)Y_*)^{-1} Y_*'(M_1 - M)y_* = R_{22}^{-1}(\tilde{\beta} R_{11} + R_{21})$$

so that

$$\tilde{\beta}_* - \hat{\beta}_* = R_{22}^{-1}(\tilde{\beta} - \hat{\beta})R_{11}\,.$$

Furthermore,

$$(Y_*'(M_1 - M)Y_*/T)^{-1} - (Y_*' M_1 Y_*/T)^{-1} = R_{22}^{-1}\left[(Y'(M_1 - M)Y/T)^{-1} - (Y' M_1 Y/T)^{-1}\right](R_{22}^{-1})'\,,$$

and, since R_{11} is a positive scalar,

$$\begin{aligned} &(\tilde{\beta}_* - \hat{\beta}_*)'\left[(Y_*'(M_1 - M)Y_*/T)^{-1} - (Y_*' M_1 Y_*/T)^{-1}\right]^{-1}(\tilde{\beta}_* - \hat{\beta}_*) \\ &= R_{11}^2(\tilde{\beta} - \hat{\beta})'\left[(Y'(M_1 - M)Y/T)^{-1} - (Y' M_1 Y/T)^{-1}\right]^{-1}(\tilde{\beta} - \hat{\beta})\,. \end{aligned} \tag{B.37}$$

Moreover,

$$\begin{aligned} y_* - \bar{Y}\hat{\beta}_* &= yR_{11} + YR_{22} - YR_{22}\left[R_{22}Y'(M_1 - M)YR_{22}\right]^{-1}YR_{22}'M_1(yR_{11} + YR_{22}) \\ &= yR_{11} + YR_{22} - Y\hat{\beta}R_{11} - YR_{22} = (y - Y\hat{\beta})R_{11}\,. \end{aligned} \tag{B.38}$$

So, we get

$$\hat{\sigma}_*^2 = (y_* - \bar{Y}\hat{\beta}_*)'M_1(y_* - \bar{Y}\hat{\beta}_*)/T = R_{11}^2(y - Y\hat{\beta})'M_1(y - \bar{Y}\hat{\beta})/T = R_{11}^2\hat{\sigma}^2\,.$$

Hence from (B.35),

$$\begin{aligned} \mathcal{H}_{3*} &= \frac{TR_{11}^2(\tilde{\beta} - \hat{\beta})'\left[(Y_*'(M_1 - M)Y_*/T)^{-1} - (Y_*'M_1Y_*/T)^{-1}\right]^{-1}(\tilde{\beta} - \hat{\beta})}{R_{11}^2\hat{\sigma}^2} \\ &= T(\tilde{\beta} - \hat{\beta})'\left[\hat{\sigma}^2(Y_*'(M_1 - M)Y_*/T)^{-1} - \hat{\sigma}^2(Y_*'M_1Y_*/T)^{-1}\right]^{-1}(\tilde{\beta} - \hat{\beta}) \\ &= \mathcal{H}_3. \end{aligned} \tag{B.39}$$

□

PROOF OF THEOREM **4.6** In (3.2) - (B.41), replace y by $\bar{y}$ and Y by $\bar{Y}$. By Lemma **4.5**, we can write:

$$\mathcal{H}_i = T(\tilde{\beta}_* - \hat{\beta}_*)'\hat{\mathbf{\Sigma}}_{i*}^{-1}(\tilde{\beta}_* - \hat{\beta}_*)\,,\ i = 1,\ 2,\ 3\,, \tag{B.40}$$

$$\mathcal{T}_l = \kappa_l(\tilde{\beta}_* - \hat{\beta}_*)'\tilde{\mathbf{\Sigma}}_{l*}^{-1}(\tilde{\beta}_* - \hat{\beta}_*)\,,\ l = 1,\ 2,\ 3,\ 4\,, \tag{B.41}$$

$$\mathcal{RH} = \kappa_R\bar{y}'\hat{\mathbf{\Sigma}}_{*R}\bar{y}\,, \tag{B.42}$$

where $\hat{\beta}_*,\ \tilde{\beta}_*,\ \hat{\Sigma}_{*i},\ \tilde{\Sigma}_{*l}$ and $\hat{\Sigma}_{*R}$ are the corresponding of $\hat{\beta},\ \tilde{\beta},\ \hat{\Sigma}_i$ and $\tilde{\Sigma}_l$ defined in (3.2)-(3.13).

From (4.56) and noting that $M_1 - M = (M_1 - M),$ and $MZ_2 = 0,$ we have

$$\begin{aligned} M\bar{y} &= M\bar{v}\,,\ M\bar{Y} = M\bar{V} \\ M_1\bar{y} &= M_1(\mu_1 + \bar{v})\,,\ M_1\bar{Y} = M_1(\mu_2 + \bar{V})\,, \end{aligned} \tag{B.43}$$

where $\mu_1 = M_1 Z_2 \Pi_2 \zeta = \mu_2 P_{22}^{-1}\zeta$ and $\mu_2 = M_1 Z_2 \Pi_2 P_{22}$, where $\zeta = \beta P_{11} + P_{21}$. From (B.43), we get:

$$\bar{Y}' M_1 \bar{Y} = (\mu_2 + \bar{V})' M_1 (\mu_2 + \bar{V}) = \Omega_{LS}(\mu_2, \bar{V}), \tag{B.44}$$

$$\bar{Y}'(M_1 - M)\bar{Y} = (\mu_2 + \bar{V})'(M_1 - M)(\mu_2 + \bar{V}) = \Omega_{IV}(\mu_2, \bar{V}), \tag{B.45}$$

$$\bar{Y}' M_1 \bar{y} = (\mu_2 + \bar{V})' M_1 (\mu_1 + \bar{v}), \tag{B.46}$$

$$\bar{Y}'(M_1 - M)\bar{y} = (\mu_2 + \bar{V})'(M_1 - M)(\mu_1 + \bar{v}), \tag{B.47}$$

$$\hat{\beta}_* = \Omega_{LS}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})' M_1 (\mu_1 + \bar{v}), \tag{B.48}$$

$$\tilde{\beta}_* = \Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'(M_1 - M)(\mu_1 + \bar{v}), \tag{B.49}$$

hence

$$\tilde{\beta}_* - \hat{\beta}_* = C(\mu_1 + \bar{v}) \tag{B.50}$$

where

$$C = \Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'(M_1 - M) - \Omega_{LS}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})' M_1 . \tag{B.51}$$

Moreover,

$$\hat{\sigma}_*^2 = \frac{1}{T}(\bar{y} - \bar{Y}\hat{\beta}_*)' M_1 (\bar{y} - \bar{Y}\hat{\beta}_*) = \frac{1}{T}(\mu_1 + \bar{v})' C_* (\mu_1 + \bar{v}) \tag{B.52}$$

$$= \frac{1}{T}\omega_{LS}^2(\mu_1, \mu_2, \bar{V}, \bar{v}), \tag{B.53}$$

$$\tilde{\sigma}_*^2 = \frac{1}{T}(\bar{y} - \bar{Y}\tilde{\beta}_*)' M_1 (\bar{y} - \bar{Y}\tilde{\beta}_*) = \frac{1}{T}(\mu_1 + \bar{v})' D_*' D_* (\mu_1 + \bar{v}), \tag{B.54}$$

$$= \frac{1}{T}\omega_{IV}^2(\mu_1, \mu_2, \bar{V}, \bar{v}), \tag{B.55}$$

where

$$C_* = \left[I - M_1(\mu_2 + \bar{V})\Omega_{LS}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'\right] M_1 , \tag{B.56}$$

$$D_* = \left[I - M_1(\mu_2 + \bar{V})\Omega_{IV}(\mu_2, \bar{V})^{-1}(\mu_2 + \bar{V})'(M_1 - M)\right] M_1 . \tag{B.57}$$

So, we have

$$\hat{\mathbf{\Sigma}}_{1*} = \omega_{IV}^2(\mu_1,\mu_2,\bar{V},\bar{v})\Omega_{IV}(\mu_2,\bar{V})^{-1} - \omega_{LS}^2(\mu_1,\mu_2,\bar{V},\bar{v})\Omega_{LS}(\mu_2,\bar{V})^{-1},$$
$$\hat{\mathbf{\Sigma}}_{2*} = \frac{1}{T}\omega_{IV}^2(\mu_1,\mu_2,\bar{V},\bar{v})\Delta,\ \hat{\mathbf{\Sigma}}_{3*} = \frac{1}{T}\omega_{LS}^2(\mu_1,\mu_2,\bar{V},\bar{v})\Delta\,, \tag{B.58}$$

where

$$\Delta = C'C = \Omega_{IV}(\mu_2,\bar{V})^{-1} - \Omega_{LS}(\mu_2,\bar{V})^{-1}. \tag{B.59}$$

For $T-k_1-k_2>G$, the matrix Δ is positive definite, thus we have,

$$\mathcal{H}_i = T[\mu_1+\bar{v}]'\Gamma_i(\mu_1,\mu_2,\bar{v},\bar{V})[\mu_1+\bar{v}]\,, \quad i=1,\,2,\,3\,.$$

where $\Gamma_i(\mu_1,\mu_2,\bar{v},\bar{V})$, $i=1,\,2,\,3$ are defined in Theorem **4.6**. For the T-tests, remark that by the definition of these statistics, we have $T_4=(\kappa_4/T)H_3,$ so we easily get

$$\mathcal{T}_4 = \kappa_4[\mu_1+\bar{v}]'\Gamma_3(\mu_1,\mu_2,\bar{v},\bar{V})[\mu_1+\bar{v}]\,. \tag{B.60}$$

Since $\tilde{\sigma}_{*2}^2 = \hat{\sigma}_*^2 - \tilde{\sigma}_*^2(\bar{\beta}_*-\tilde{\beta}_*)'(\bar{\Gamma}_2)^{-1}(\bar{\beta}_*-\tilde{\beta}_*),$ from (B.50)-(B.55), we have

$$\tilde{\sigma}_{*2}^2 = \omega_{LS}^2(\mu_1,\mu_2,\bar{V},\bar{v}) - (\mu_1+\bar{v})'C'\Delta^{-1}C(\mu_1+\bar{v}) \tag{B.61}$$
$$= (\mu_1+\bar{v})'(C_* - C'\Delta^{-1}C)(\mu_1+\bar{v}) = \omega_2^2(\mu_1,\mu_2,\bar{V},\bar{v}) \equiv \omega_2^2. \tag{B.62}$$

So, we find

$$\mathcal{T}_2 = \frac{\kappa_2}{\omega_2^2}[\mu_1+\bar{v}]'C'\Delta^{-1}C[\mu_1+\bar{v}]\,. \tag{B.63}$$

By proceeding as above, we also get

$$\mathcal{T}_l = \frac{\kappa_l}{\omega_l^2}[\mu_1+\bar{v}]'C'\Delta^{-1}C[\mu_1+\bar{v}]\,, \quad l=1,\,3\,,$$
$$\mathcal{RH} = \frac{\kappa_R}{\omega_R^2}[\mu_1+\bar{v}]'P_{D1\bar{Z}_2}[\mu_1+\bar{v}]\,,$$

where $\omega_l^2\ \ l=1,\,3$ and ω_R^2 are defined in Theorem **4.6**. □

PROOF OF COROLLARY **4.7** The results of Corollary **4.7** follow by setting $\Pi_2\zeta = 0$ in the proof of Theorem **4.6** above. □

PROOF OF THEOREM **4.8** Under the assumptions of Theorem **4.6**, we have

$$\begin{aligned} \mathcal{T}_l &= \kappa_l[\mu_1 + \bar{v}]' \bar{\Gamma}_l(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}]\,, \quad l = 1,\, 2,\, 3,\, 4\,, \\ \mathcal{H}_i &= T[\mu_1 + \bar{v}]' \Gamma_i(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}]\,, \quad i = 1,\, 2,\, 3\,, \\ \mathcal{RH} &= \kappa_R[\mu_1 + \bar{v}]' \Gamma_R(\mu_1, \mu_2, \bar{v}, \bar{V})[\mu_1 + \bar{v}]\,, \end{aligned}$$

where $\bar{\Gamma}_l(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v}),\ \bar{\Gamma}_i(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v}),\ \Gamma_R(\mu_1,\, \mu_2,\, \bar{V},\, \bar{v}),\ \ \mu_1,\ \mu_2,\ \kappa_l$ and κ_R are defined in Theorem **4.6**. Under the normality assumption (4.43), $\mu_1 + \bar{v}$ is independent of $\bar{V}$ and $\mu_1 + \bar{v} \sim N(\mu_1, 1)$. Since $C'\Delta^{-1}C$ [where C and Δ are defined in Theorem **4.6**] is symmetric idempotent of rank G, provided $[Z_1, Z_2]$ is fixed, we have

$$(\mu_1 + \bar{v})' C' \Delta^{-1} C(\mu_1 + \bar{v})|\, \bar{V} \sim \chi^2(G,\, \nu_1), \tag{B.64}$$

where $\nu_1 = \mu_1' C' \Delta^{-1} C \mu_1$. By noting that E [also defined in Theorem **4.6**] is symmetric idempotent of rank $k_2 - G,$ we get

$$(\mu_1 + \bar{v})' E(\mu_1 + \bar{v})|\, \bar{V} \sim \chi^2(k_2 - G,\, \upsilon_1), \tag{B.65}$$

with $\upsilon_1 = \mu_1' E \mu_1$. Furthermore, we have $(C'\Delta^{-1}C)E = 0,$ hence

$$\mathcal{T}_1|\, \bar{V} \sim F(G,\, k_2 - G; \nu_1,\, \upsilon_1). \tag{B.66}$$

By applying the same arguments to T_2, we get

$$\mathcal{T}_2|\, \bar{V} \sim F(G,\, T - k_1 - 2G; \nu_1,\, \upsilon_2), \tag{B.67}$$

where $\upsilon_2 = \mu_1'(C_* - C'\Delta^{-1}C)\mu_1$. Moreover, from the notations in Theorem **4.6**, we can see that

$$\mathcal{T}_4 = \frac{\kappa_4}{1 + \frac{1}{\kappa_2}\frac{1}{\mathcal{T}_2}} \tag{B.68}$$

and since $T_2|\,\bar{V} \sim F(G, T - k_1 - 2G; \nu_1, \upsilon_2),$ we have $\frac{1}{\mathcal{T}_2}|\,\bar{V} \sim F(T - k_1 - 2G, G; \upsilon_2, \nu_1)$. So, we get

$$\mathcal{T}_4|\,\bar{V} \sim \frac{\kappa_4}{1 + \frac{1}{\kappa_2}F(T - k_1 - 2G,\, G; \upsilon_2,\, \nu_1)}. \tag{B.69}$$

Note also that because $\omega_{LS}^2 \geq \omega_2^2,$ hence, we have

$$\begin{aligned} \mathcal{T}_4|\,\bar{V} &\leq \frac{\kappa_4}{\omega_2^2}(\mu_1 + \bar{\upsilon})'C'\Delta^{-1}C(\mu_1 + \bar{\upsilon})|\,\bar{V} \\ &= \bar{\kappa}_2^*\mathcal{T}_2|\,\bar{V} \sim \bar{\kappa}_2^* F(G,\, T - k_1 - 2G; \nu_1,\, \upsilon_2) \end{aligned} \tag{B.70}$$

where $\kappa_2,\ \kappa_4,\ \bar{\kappa}_2^*$ are given in Theorem **4.8**. For the test T_3, we have

$$(\mu_1+\bar{\upsilon})'C'\Delta^{-1}C(\mu_1+\bar{\upsilon})|\,\bar{V} \sim \chi^2(G; \nu_1),\ \omega_{IV}^2 = (\mu_1+\bar{\upsilon})'D_*'D_*(\mu_1+\bar{\upsilon}) \sim \chi^2(T-k_1-G; \nu_3) \tag{B.71}$$

where $\nu_3 = \mu_1'D_*'D_*\mu_1$. However, since $D_*'D_*(C'\Delta^{-1}C) \neq 0,\ T_3$ does not follow necessary a F-type distribution.

By following the same steps as above, we get the results for H_2, H_3 and RH. We note that in contrast to other statistics, the distributions of $T_3,\ H_1$ and H_2 are not standard despite the normality assumption. □

PROOF OF COROLLARY **4.9** If $\Pi_2\zeta = 0$, which is typically the case under $H_0,$ we have $\mu_1 = 0$ so that $\nu_1 = \nu_2 = \nu_3 = \upsilon_1 = \nu_R = 0$ in the proof of Theorem **4.8** leading to the results in Corollary **4.9**. □

PROOF OF THEOREM **6.1** First, note that in the proofs of Theorem **6.1** and Theorem **6.2**, the reduced form system (A.25) - (A.26) is used instead of (2.1) - (2.2).

Now, since $\delta = 0,$ we have $a = \Sigma_V^{-1}\delta = 0.$

(A) We suppose first that $\Pi_2 = \Pi_0$ where Π_0 is a $k_2 \times G$ constant matrix with rank G. Then, we have:

$$\hat{\Omega}_{IV} \xrightarrow{p} \Pi_0' \Sigma_{\bar{Z}_2} \Pi_0,\ \hat{\Omega}_{LS} \xrightarrow{p} \Pi_0' \Sigma_{\bar{Z}_2} \Pi_0 + \Sigma_V\,, \tag{B.72}$$

$$\frac{Y'u}{T} \xrightarrow{p} \delta = 0\,, \tag{B.73}$$

$$Y' M_1 u/T \xrightarrow{p} \delta = 0\,, \tag{B.74}$$

$$\hat{\sigma}^2/T = \hat{u}'\hat{u}/T = u'u/T - (u'M_1Y/T)\hat{\Omega}_{LS}^{-1}(Y'M_1u/T) \xrightarrow{p} \sigma_u^2\,, \tag{B.75}$$

$$\begin{aligned} \tilde{\sigma}^2/T &= u'u/T - 2(u'M_1Y/T)\hat{\Omega}_{IV}^{-1}(Y'(M_1 - M)u/T) \\ &\quad + (u'(M_1 - M)Y/T)\hat{\Omega}_{IV}^{-1}(Y'(M_1 - M)u/T) \xrightarrow{p} \sigma_u^2\,, \end{aligned} \tag{B.76}$$

$$\begin{aligned} \frac{Y'u}{\sqrt{T}} &= \bar{\Pi}_1' \frac{Z_1'u}{\sqrt{T}} + \Pi_0' \frac{\bar{Z}_2'u}{\sqrt{T}} + \frac{V'V}{\sqrt{T}}(\Sigma_V^{-1}\delta) + \frac{V'\varepsilon}{\sqrt{T}} = \bar{\Pi}_1' \frac{Z_1'u}{\sqrt{T}} + \Pi_0' \frac{\bar{Z}_2'u}{\sqrt{T}} + \frac{V'\varepsilon}{\sqrt{T}} \\ &\xrightarrow{L} \bar{\Pi}_{01}' S_{1u} + \Pi_0' S_{2u} + S_{V\varepsilon}\,, \end{aligned} \tag{B.77}$$

where $\bar{\Pi}_1 \xrightarrow{p} \bar{\Pi}_{01} = \Pi_1 + \Sigma_{Z_1}\Sigma_{Z_1 Z_2}\Pi_0$ and $\delta = 0$. Hence

$$\begin{aligned} \frac{Y'M_1u}{\sqrt{T}} &= \frac{Y'u}{\sqrt{T}} - \left(\frac{Y'Z_1}{T}\right)\left(\frac{Z_1'Z_1}{T}\right)^{-1}\frac{Z_1'u}{\sqrt{T}} \\ &\xrightarrow{L} (\bar{\Pi}_{01}'S_{1u} + \Pi_0'S_{2u} + S_{V\varepsilon}) - \bar{\Pi}_{01}'S_{1u} = \Pi_0'S_{2u} + S_{V\varepsilon}, \end{aligned} \tag{B.78}$$

$$\frac{1}{\sqrt{T}}Y'(M_1 - M)u = \left(\frac{Y'\bar{Z}_2}{T}\right)\left(\frac{\bar{Z}_2'\bar{Z}_2}{T}\right)^{-1}\left(\frac{\bar{Z}_2'u}{\sqrt{T}}\right) \xrightarrow{L} \Pi_0' S_{2u}\,. \tag{B.79}$$

So, we see that

$$\sqrt{T}(\tilde{\beta} - \hat{\beta}) = \hat{\Omega}_{LS}^{-1}\frac{Y'M_1u}{\sqrt{T}} - \hat{\Omega}_{IV}^{-1}\frac{Y'(M_1 - M)u}{\sqrt{T}} \xrightarrow{L} \psi_\pi\,,$$

$$\hat{\mathbf{\Sigma}}_i \xrightarrow{p} \sigma_u^2 \Delta_\Pi,\ \Delta_\Pi = (\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0)^{-1} - (\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0 + \Sigma_V)^{-1},\ i = 1,\ 2,\ 3\,,$$

where

$$\psi_\pi = (\varPi_0'\varSigma_{\bar{Z}_2}\varPi_0 + \varSigma_V)^{-1}(\varPi_0'S_{2u} + S_{V\varepsilon}) - (\varPi_0'\varSigma_{\bar{Z}_2}\varPi_0)^{-1}\varPi_0'S_{2u}\ . \tag{B.80}$$

Thus

$$\mathcal{H}_i = \sqrt{T}(\tilde{\beta} - \hat{\beta})'\hat{\mathbf{\Sigma}}_i^{-1}\sqrt{T}(\tilde{\beta} - \hat{\beta}) \xrightarrow{L} \frac{1}{\sigma_u^2}\psi_\pi'\varDelta_\varPi^{-1}\psi_\pi,\ i = 1,\ 2,\ 3\,. \tag{B.81}$$

Since $a = 0$, we have $\sigma_u^2 = \sigma_\varepsilon^2$ and it is easy to see that

$$\begin{bmatrix} \varPi_0'S_{2u} + S_{V\varepsilon} \\ \varPi_0'S_{2u} \end{bmatrix} \sim N\left[0,\ \sigma_u^2\varSigma_0\right].$$

where

$$\varSigma_0 = \begin{bmatrix} \varPi_0'\varSigma_{\bar{Z}_2}\varPi_0 + \varSigma_V & \varPi_0'\varSigma_{\bar{Z}_2}\varPi_0 \\ \varPi_0'\varSigma_{\bar{Z}_2}\varPi_0 & \varPi_0'\varSigma_{\bar{Z}_2}\varPi_0 \end{bmatrix} \tag{B.82}$$

So, we have

$$\psi_\pi \sim N\left\{0,\ \sigma_u^2[(\varPi_0'\varSigma_{\bar{Z}_2}\varPi_0)^{-1} - (\varSigma_V + \varPi_0'\varSigma_{\bar{Z}_2}\varPi_0)^{-1}]\right\} \equiv N(0,\ \sigma_u^2\varDelta_\varPi), \tag{B.83}$$

hence

$$\mathcal{H}_i \xrightarrow{L} \chi^2(G),\ i = 1,\ 2,\ 3\,. \tag{B.84}$$

By the same arguments as above, we also get

$$\mathcal{T}_2 \xrightarrow{L} \frac{1}{G}\chi^2(G),\ \mathcal{T}_l \xrightarrow{L} \chi^2(G),\ l \ = \ 3,\ 4, \quad \text{and} \quad \mathcal{RH} \xrightarrow{L} \frac{1}{k_2}\chi^2(k_2)\,. \tag{B.85}$$

We now derive the distribution of T_1. We can write

$$\mathcal{T}_1 = \frac{k_2 - G}{G}\frac{T(\tilde{\beta} - \hat{\beta})'\varDelta^{-1}(\tilde{\beta} - \hat{\beta})}{T\tilde{\sigma}_1^2} \tag{B.86}$$

and $T(\tilde{\beta} - \hat{\beta})'\varDelta^{-1}(\tilde{\beta} - \hat{\beta}) \xrightarrow{L} \psi_\pi'\varDelta_\varPi^{-1}\psi_\pi \sim \sigma_u^2\chi^2(G)$. Furthermore, because Z_1 is orthogonal to $\bar{Z}_2$, we have

$$T\tilde{\sigma}_1^2 = u'((M_1 - M) - P_{\hat{Y}})u = u'(M_1 - M)u - u'P_{\hat{Y}}u\,,$$

where $\hat{Y} = (M_1 - M)Y$. Thus, if rank$(\Pi_0) = G$,

$$
\begin{aligned}
T\tilde{\sigma}_1^2 &\xrightarrow{L} S_{2u}'\Sigma_{\bar{Z}_2}^{-1}S_{2u} - S_{2u}'\Pi_0(\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0)^{-1}\Pi_0'S_{2u} \\
&= S_{2u}'\Sigma_{\bar{Z}_2}^{-1/2}[I_{k_2} - P(P'P)^{-1}P']\Sigma_{\bar{Z}_2}^{-1/2}S_{2u}\,, \qquad \text{(B.87)}
\end{aligned}
$$

where $P = \Sigma_{\bar{Z}_2}^{1/2}\Pi_0$ and the matrix $I_k - P(P'P)^{-1}P'$ is idempotent with rank $k_2 - G$. Further, we have $\frac{1}{\sigma_u}\Sigma_{\bar{Z}_2}^{-1/2}S_{2u} \sim N[0, I_{k_2}]$, hence $T\tilde{\sigma}_1^2 \xrightarrow{L} \sigma_u^2\chi^2(k_2 - G)$. In addition, by noting that $T\tilde{\sigma}_1^2 = u'(M_1 - M)M_{\hat{Y}}(M_1 - M)u$ and $T(\tilde{\beta} - \hat{\beta})'\Delta^{-1}(\tilde{\beta} - \hat{\beta}) = u'A_Z u,$ where

$$
\mathcal{A}_Z = \frac{1}{T}(M_1 Y\hat{\Omega}_{LS}^{-1} - \hat{Y}\hat{\Omega}_{IV}^{-1})\Delta^{-1}(\hat{\Omega}_{LS}^{-1}Y'M_1 - \hat{\Omega}_{IV}^{-1}\hat{Y}')
$$

is symmetric idempotent matrix with $A_Z((M_1 - M)M_{\hat{Y}}\bar{M}_1) = ((M_1 - M)M_{\hat{Y}}(M_1 - M))A_Z = 0$, we then have

$$
\mathcal{T}_1 \xrightarrow{L} F(G,\, k - G)\,. \qquad \text{(B.88)}
$$

(B) Suppose that $\Pi_2 = \Pi_0/\sqrt{T}$ where Π_0 is a $k \times G$ constant matrix ($\Pi_0 = 0$ is allowed), then we have

$$
T\hat{\Omega}_{IV} \xrightarrow{L} \Psi_V = (\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V}),\ \hat{\Omega}_{LS} \xrightarrow{p} \Sigma_V\,, \qquad \text{(B.89)}
$$

$$
\frac{1}{T}Y'M_1 u \xrightarrow{p} \delta = 0,\ Y'(M_1 - M)u \xrightarrow{L} (\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}\,, \qquad \text{(B.90)}
$$

$$
\hat{\sigma}^2 = u'u/T - (u'M_1Y/T)\hat{\Omega}_{LS}^{-1}(Y'M_1u/T) \xrightarrow{p} \sigma_u^2 - \delta'\Sigma_V^{-1}\delta = \sigma_u^2\,, \qquad \text{(B.91)}
$$

$$
\tilde{\beta} - \beta = (T\hat{\Omega}_{LS})^{-1}Y'(M_1 - M)u \xrightarrow{p} \Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}\,. \qquad \text{(B.92)}
$$

Thus,

$$
\tilde{\sigma}^2 = \frac{u'u}{T} - 2\frac{u'M_1Y}{T}(\tilde{\beta} - \beta) + (\tilde{\beta} - \beta)'\hat{\Omega}_{LS}(\tilde{\beta} - \beta) \xrightarrow{L} \bar{\sigma}_u^2,\ \hat{\Sigma}_i \xrightarrow{L} \frac{1}{\bar{\sigma}_u^2}\Psi_V^{-1},\ i = 1,\, 2,
$$

where $\bar{\sigma}_u^2 = \sigma_u^2 + S_{2u}'\Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})\Psi_V^{-1}\Sigma_V\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u},$ so that

$$
\mathcal{H}_i \xrightarrow{L} \frac{1}{\bar{\sigma}_u^2}S_{2u}'\Sigma_A S_{2u},\ i = 1,\ 2\,,
$$

where $\Sigma_A = \Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}$. Since $\bar{\sigma}_u^2 \geq \sigma_u^2,$ it is clear that

$$\mathcal{H}_i \leq \frac{1}{\sigma_u^2} S'_{2u}\Sigma_A S_{2u},\ i = 1,\ 2.$$

Since S_{2u} is independent with S_{2V}, we have:

$$\Sigma_{\bar{Z}_2}\Pi_0 + S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}|_{S_{2V}} \sim N(0,\ \sigma_u^2\Psi_V)\,, \tag{B.93}$$

$$\frac{1}{\sigma_u^2}S'_{2u}\Sigma_A S_{2u}|_{S_{2V}} \sim \chi^2(G)\,,$$

and $H_i \leq \chi^2(G),\ i = 1, 2$. However, because $\hat{\Sigma}_3 \xrightarrow{p} \frac{1}{\sigma_u^2}\Psi_V^{-1},$ it is easy to see that $H_3|_{S_{2V}} \xrightarrow{L} \chi^2(G)$ i.e. $H_3 \xrightarrow{L} \chi^2(G)$. By applying the same arguments to the T-tests, we also get:

$$\mathcal{T}_1 \xrightarrow{L} F(G,\ k_2 - G), \quad \mathcal{T}_2 \xrightarrow{L} \frac{1}{G}\chi^2(G), \quad \mathcal{T}_4 \xrightarrow{L} \chi^2(G)\ , \tag{B.94}$$

$$\mathcal{T}_3 \xrightarrow{L} \frac{1}{\bar{\sigma}_u^2}S'_{2u}\Sigma_A S_{2u} \leq \chi^2(G), \quad \text{and} \quad \mathcal{RH} \xrightarrow{L} \frac{1}{k_2}\chi^2(k_2)\,. \tag{B.95}$$

□

PROOF OF THEOREM **6.2** Suppose that $\delta \neq 0$. (A) If $\Pi_2 = \Pi_0$ with $\text{rank}(\Pi_0) = G$, as in the proof of Theorem **6.1**, we have from (2.5) - (2.6) and (6.12):

$$\hat{\Omega}_{IV} \xrightarrow{p} \Pi_0'\Sigma_{\bar{Z}_2}\Pi_0,\ \hat{\Omega}_{LS} \xrightarrow{p} \Pi_0'\Sigma_{\bar{Z}_2}\Pi_0 + \Sigma_V\ , \tag{B.96}$$

$$\frac{Y'M_1u}{T} \xrightarrow{p} \delta \neq 0,\ \frac{Y'(M_1 - M)u}{T} \xrightarrow{p} 0\,, \tag{B.97}$$

$$\hat{\sigma}^2 = u'u/T - (u'M_1Y/T)\hat{\Omega}_{LS}^{-1}(Y'M_1u/T) \xrightarrow{p} \sigma_u^2 - \delta'(\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0 + \Sigma_V)^{-1}\delta = \tilde{\sigma}_u^2\,, \tag{B.98}$$

$$\tilde{\sigma}^2 = u'u/T - 2(u'M_1Y/T)\hat{\Omega}_{IV}^{-1}(Y'(M_1 - M)u/T) + (u'(M_1 - M)Y/T)\hat{\Omega}_{IV}^{-1}(Y'(M_1 - M)u/T) \xrightarrow{p} \sigma_u^2\,. \tag{B.99}$$

Since we have

$$\mathcal{H}_i = T(\tilde{\beta} - \hat{\beta})'\hat{\mathbf{\Sigma}}_i^{-1}(\tilde{\beta} - \hat{\beta}) \tag{B.100}$$

and

$$(\tilde{\beta}-\hat{\beta}) = \hat{\Omega}_{LS}^{-1}(Y'M_1u/T) - \hat{\Omega}_{IV}^{-1}(Y'(M_1-M)u/T) \xrightarrow{p} (\Pi_0'\Sigma_Z\Pi_0+\Sigma_V)^{-1}\delta\,, \tag{B.101}$$

$$\hat{\mathbf{\Sigma}}_i \xrightarrow{p} \sigma_i^2\Delta_\Pi,\ \Delta_\Pi = (\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0)^{-1} - (\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0+\Sigma_V)^{-1},\ i=2,\,3, \tag{B.102}$$

$$\hat{\mathbf{\Sigma}}_1 \xrightarrow{p} \Sigma_{1\Pi},\ \Sigma_{1\Pi} = \sigma_u^2(\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0)^{-1} - \tilde{\sigma}_u^2(\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0+\Sigma_V)^{-1}\,, \tag{B.103}$$

where $\sigma_2^2 = \sigma_u^2$ and $\sigma_3^2 = \tilde{\sigma}_u^2$. Since $\delta \neq 0,$ we have:

$$\begin{aligned}(\tilde{\beta}-\hat{\beta})'\hat{\mathbf{\Sigma}}_i^{-1}(\tilde{\beta}-\hat{\beta}) &\xrightarrow{p} \frac{1}{\sigma_i^2}\delta'(\Pi_0'\Sigma_Z\Pi_0+\Sigma_V)^{-1}\Delta_\Pi^{-1}(\Pi_0'\Sigma_Z\Pi_0+\Sigma_V)^{-1}\delta > 0,\ i=2,\,3,\\ (\tilde{\beta}-\hat{\beta})'\hat{\mathbf{\Sigma}}_1^{-1}(\tilde{\beta}-\hat{\beta}) &\xrightarrow{p} \delta'(\Pi_0'\Sigma_Z\Pi_0+\Sigma_V)^{-1}\Sigma_{1\Pi}^{-1}(\Pi_0'\Sigma_Z\Pi_0+\Sigma_V)^{-1}\delta > 0\,.\end{aligned} \tag{B.104}$$

From (B.100), it clear that $H_i \xrightarrow{p} +\infty$ for all $i = 1, 2, 3$. By the same arguments, we also get $T_l \xrightarrow{p} +\infty$ for all $l = 1, 2, 3, 4$ and $RH \xrightarrow{p} +\infty$. Hence, the tests H_i, T_l and RH are consistent.

(B) Suppose now that $\Pi_2 = \Pi_0/\sqrt{T}$ where Π_0 is a $k \times G$ constant matrix ($\Pi_0 = 0$ is allowed), then we have

$$T\hat{\Omega}_{IV} \xrightarrow{L} \Psi_V = (\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V}),\ \hat{\Omega}_{LS} \xrightarrow{p} \Sigma_V\,, \tag{B.105}$$

$$\frac{1}{T}Y'M_1u \xrightarrow{p} \delta \neq 0,\ Y'(M_1-M)u \xrightarrow{L} (\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}\,, \tag{B.106}$$

$$\hat{\sigma}^2 = u'u/T - (u'M_1Y/T)\hat{\Omega}_{LS}^{-1}(Y'M_1u/T) \xrightarrow{p} \sigma_u^2 - \delta'\Sigma_V^{-1}\delta = \sigma_\varepsilon^2\,, \tag{B.107}$$

$$\tilde{\beta}-\beta = (T\hat{\Omega}_{IV})^{-1}Y'(M_1-M)u \xrightarrow{p} \Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}\,. \tag{B.108}$$

and

$$\tilde{\sigma}^2 = \frac{u'u}{T} - 2\frac{u'M_1Y}{T}(\tilde{\beta}-\beta) + (\tilde{\beta}-\beta)'\hat{\Omega}_{LS}(\tilde{\beta}-\beta) \xrightarrow{L} \tilde{\sigma}_*^2,\ \frac{1}{T}\hat{\mathbf{\Sigma}}_i \xrightarrow{L} \sigma_{i*}^2\Psi_V^{-1},\ i=1,\,2,\,3$$

where $\sigma_{1*}^2 = \sigma_{2*}^2 = \tilde{\sigma}_*^2,\ \sigma_{3*}^2 = \sigma_\varepsilon^2$ and

$$\tilde{\sigma}_*^2 \;=\; \sigma_u^2 - 2\delta'\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}$$

$$+S'_{2u}\Sigma_{\bar{Z}_2}^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})\Psi_V^{-1}\Sigma_V\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}.$$

Furthermore, we have

$$\tilde{\beta}-\hat{\beta}=\hat{\Omega}_{LS}^{-1}(u'M_1Y/T)-(T\hat{\Omega}_{IV})^{-1}Y'(M_1-M)u\xrightarrow{L}\Sigma_V^{-1}\delta-\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2u}\,. \tag{B.109}$$

By noting that $S_{2u}=S_{2V}a+S_{2\varepsilon}=S_{2V}\Sigma_V^{-1}\delta-S_{2\varepsilon},$ we easily get

$$\begin{aligned}\tilde{\beta}-\hat{\beta}&\xrightarrow{L}\Sigma_V^{-1}\delta-\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2V}\Sigma_V^{-1}\delta-\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon}\\&=\Psi_V^{-1}[\Lambda_Va-(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon}].\end{aligned}$$

where $\Lambda_V=\Psi_V-(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Sigma_{\bar{Z}_2}^{-1}S_{2V}=(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'\Pi_0$ and $a=\Sigma_V^{-1}\delta$. So, we have

$$\mathcal{H}_i\xrightarrow{L}\frac{1}{\sigma_{i*}^2}(\Pi_0a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon})'\Delta_V(\Pi_0a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon}),\ i=1,\,2,\,3\,,$$

where $\sigma_{1*}^2=\sigma_{2*}^2=\tilde{\sigma}_*^2,\ \sigma_{3*}^2=\sigma_\varepsilon^2,$ and $\Delta_V=(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})\Psi_V^{-1}(\Sigma_{\bar{Z}_2}\Pi_0+S_{2V})'$. Moreover, from (2.5) - (2.6) and (6.12), $S_{2\varepsilon}\sim N(0,\sigma_\varepsilon^2\Sigma_{\bar{Z}_2})$ and $S_{2\varepsilon}$ is independent with S_{2V}. So, we have

$$\mathcal{H}_3|S_{2V}\xrightarrow{L}\chi^2(G,\,\mu_V),\ \mu_V=\frac{1}{\sigma_\varepsilon^2}a'\Pi_0'\Delta_V\Pi_0a\,. \tag{B.110}$$

Since $T_3=\frac{\kappa_3}{T}H_2,\ T_3=\frac{\kappa_4}{T}H_3$ and $\frac{\kappa_3}{T}=\frac{\kappa_4}{T}=\frac{T-G}{T}\to 1$ as $T\to+\infty,$ it follows that

$$\begin{aligned}\mathcal{T}_3&\xrightarrow{L}\frac{1}{\tilde{\sigma}_*^2}(\Pi_0a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon})'\Delta_V(\Pi_0a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon}),\\\mathcal{T}_4&\xrightarrow{L}\frac{1}{\sigma_\varepsilon^2}(\Pi_0a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon})'\Delta_V(\Pi_0a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon})\,.\end{aligned} \tag{B.111}$$

So, conditional on S_{2V},

$$\mathcal{T}_4|S_{2V}\xrightarrow{L}\chi^2(G,\,\mu_V)\,. \tag{B.112}$$

Moreover, by noting that $\operatorname*{plim}_{T\to\infty}(\tilde{\sigma}_2^2)=\operatorname*{plim}_{T\to\infty}(\hat{\sigma}^2)=\sigma_\varepsilon^2,$ we also get

$$\mathcal{T}_2\xrightarrow{L}\frac{1}{\sigma_\varepsilon^2G}(\Pi_0a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon})'\Delta_V(\Pi_0a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon})\text{ and }\mathcal{T}_2|S_{2V}\xrightarrow{L}\frac{1}{G}\chi^2(G,\,\mu_V)\,.$$

Furthermore, we can see that

$$T\tilde{\sigma}_1^2 = u'(M_1 - M)M_{\hat{Y}}(M_1 - M)u \xrightarrow{L} S'_{2u}(\Sigma_{\bar{Z}_2}^{-1} - \Sigma_{\bar{Z}_2}^{-1}\Delta_V\Sigma_{\bar{Z}_2}^{-1})S_{2u}\,, \tag{B.113}$$

where the limit term in (B.113) can be written as

$$S'_{2u}(\Sigma_{\bar{Z}_2}^{-1} - \Sigma_{\bar{Z}_2}^{-1}\Delta_V\Sigma_{\bar{Z}_2}^{-1})S_{2u} = (\Sigma_{\bar{Z}_2}^{-1/2}S_{2V}a + \Sigma_{\bar{Z}_2}^{-1/2}S_{2\varepsilon})'\Delta_V^*(\Sigma_{\bar{Z}_2}^{-1/2}S_{2V}a + \Sigma_{\bar{Z}_2}^{-1/2}S_{2\varepsilon}),$$

where $\Delta_V^* = I_{k_2} - \Sigma_{\bar{Z}_2}^{-1/2}\Delta_V\Sigma_{\bar{Z}_2}^{-1/2}$ is symmetric idempotent with rank $k_2 - G$. So, we have $T\tilde{\sigma}_1^2|S_{2V} \xrightarrow{L} \sigma_\varepsilon^2\chi^2(k - G, \lambda_V)$, where

$$\lambda_V = \frac{1}{\sigma_\varepsilon^2}a'S'_{2V}\Sigma_{\bar{Z}_2}^{-1/2}\Delta_V^*\Sigma_{\bar{Z}_2}^{-1/2}S_{2V}a = \frac{1}{\sigma_\varepsilon^2}a'S'_{2V}(\Sigma_{\bar{Z}_2}^{-1} - \Sigma_{\bar{Z}_2}^{-1}\Delta_V\Sigma_{\bar{Z}_2}^{-1})S_{2V}a.$$

Further, we have $\Delta_V(\Sigma_{\bar{Z}_2}^{-1/2}\Delta_V^*\Sigma_{\bar{Z}_2}^{-1/2}) = \Delta_V\Sigma_{\bar{Z}_2}^{-1} - \Delta_V\Sigma_{\bar{Z}_2}^{-1}\Delta_V\Sigma_{\bar{Z}_2}^{-1}$ and since $\Delta_V\Sigma_{\bar{Z}_2}^{-1}\Delta_V = \Delta_V$, it follows that $\Delta_V(\Sigma_{\bar{Z}_2}^{-1/2}\Delta_V^*\Sigma_{\bar{Z}_2}^{-1/2}) = 0$. So, conditional on S_{2V}, the quadratic forms

$(\Sigma_{\bar{Z}_2}^{-1/2}S_{2V}a+\Sigma_{\bar{Z}_2}^{-1/2}S_{2\varepsilon})'\Delta_V^*(\Sigma_{\bar{Z}_2}^{-1/2}S_{2V}a+\Sigma_{\bar{Z}_2}^{-1/2}S_{2\varepsilon})$ and $(\Pi_0 a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon})'\Delta_V(\Pi_0 a-\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon})$

are independent and distributed as noncentral chi-squares. Hence, we have

$$\mathcal{T}_1|S_{2V} \xrightarrow{L} F(G,\, k - G;\, \mu_V,\, \lambda_V)\,. \tag{B.114}$$

For the statistic RH, the denominator is

$$\frac{1}{T}u'M_{\bar{X}}u = \frac{1}{T}(u'M_{X_1}u - u'M_{X_1}\bar{Z}_2(\bar{Z}_2'M_{X_1}\bar{Z}_2)^{-1}\bar{Z}_2'M_{X_1}u)\,, \tag{B.115}$$

where

$$\begin{aligned} \frac{1}{T}u'M_{X_1}u &= \frac{1}{T}u'M_1u - \frac{1}{T}u'M_1Y(Y'M_1Y)^{-1}Y'M_1u \xrightarrow{p} \sigma_u^2 - \delta'\Sigma_V^{-1}\delta = \sigma_\varepsilon^2\,, \\ \text{and} \quad & \frac{1}{T}(u'M_{X_1}\bar{Z}_2(\bar{Z}_2'M_{X_1}\bar{Z}_2)^{-1}\bar{Z}_2'M_{X_1}u) \xrightarrow{p} 0. \end{aligned}$$

So, we find $\frac{1}{T}u'M_{\bar{X}}u \xrightarrow{p} \sigma_\varepsilon^2$. For the numerator, we have

$$\frac{1}{k_2}u'(M_{X_1} - M_{\bar{X}})u = \frac{1}{k_2}\frac{u'M_{X_1}\bar{Z}_2}{\sqrt{T}}\left(\frac{\bar{Z}_2'M_{X_1}\bar{Z}_2}{T}\right)^{-1}\frac{\bar{Z}_2'M_{X_1}u}{\sqrt{T}}\,. \tag{B.116}$$

Moreover, $\frac{\bar{Z}_2'M_{X_1}\bar{Z}_2}{T} = \frac{\bar{Z}_2'M_1\bar{Z}_2}{T} - \frac{\bar{Z}_2'M_1Y}{T}\left(\frac{Y'M_1Y}{T}\right)^{-1}\frac{Y'M_1\bar{Z}_2}{T} \xrightarrow{p} \Sigma_{\bar{Z}_2}$ because $\frac{Y'M_1\bar{Z}_2}{T} \xrightarrow{p} 0$. Now, we have

$$\frac{\bar{Z}_2'M_{X_1}u}{\sqrt{T}} = \frac{\bar{Z}_2'u}{\sqrt{T}} - \frac{\bar{Z}_2'M_1Y}{\sqrt{T}}\left(\frac{Y'M_1Y}{T}\right)^{-1}\frac{Y'M_1u}{T}\,,$$

where $\frac{\bar{Z}_2'u}{\sqrt{T}} = \frac{\bar{Z}_2'V}{\sqrt{T}}\Sigma_V^{-1}\delta + \frac{\bar{Z}_2'\varepsilon}{\sqrt{T}} \xrightarrow{L} S_{2\varepsilon} + S_{2V}\Sigma_V^{-1}\delta$, $\left(\frac{Y'M_1Y}{T}\right)^{-1}\frac{Y'M_1u}{T} \xrightarrow{p} \Sigma_V^{-1}\delta$ and $\frac{\bar{Z}_2'M_1Y}{\sqrt{T}} \xrightarrow{L} \Sigma_{\bar{Z}_2}\Pi_0 + S_{2V}$. Hence we have

$$\frac{1}{k_2}u'(M_{X_1} - M_{\bar{X}})u \xrightarrow{L} \frac{1}{k_2}(S_{2\varepsilon} - \Sigma_{\bar{Z}_2}\Pi_0 a)'\Sigma_{\bar{Z}_2}^{-1}(S_{2\varepsilon} - \Sigma_{\bar{Z}_2}\Pi_0 a)\,,$$

thus

$$\begin{aligned}\mathcal{RH} &\xrightarrow{L} \frac{1}{k_2\sigma_\varepsilon^2}(S_{2\varepsilon} - \Sigma_{\bar{Z}_2}\Pi_0 a)'\Sigma_{\bar{Z}_2}^{-1}(S_{2\varepsilon} - \Sigma_{\bar{Z}_2}\Pi_0 a) \sim \frac{1}{k_2}\chi^2(k_2,\, \mu_R)\,,\\ \mu_R &= a'\Pi_0'\Sigma_{\bar{Z}_2}\Pi_0 a.\end{aligned}$$

□

PROOF OF OROLLARY **6.3** Let $\Pi_2 = 0$ in the proof of part (B) of Theorem **6.2** above. Then, we have $\mu_V = \lambda_V = \mu_R = 0$ and

$$\tilde{\sigma}_*^2 = \sigma_{0*}^2 = \sigma_\varepsilon^2 + S_{2\varepsilon}'\Sigma_{\bar{Z}_2}^{-1}S_{2V}(S_{2V}'\Sigma_{\bar{Z}_2}^{-1}S_{2V})^{-1}\Sigma_V(S_{2V}'\Sigma_{\bar{Z}_2}^{-1}S_{2V})^{-1}S_{2V}'\Sigma_{\bar{Z}_2}^{-1}S_{2\varepsilon} \geq \sigma_\varepsilon^2\,, \tag{B.117}$$

so that the results in Corollary **6.3** follow. □

C. Performance of OLS and 2SLS estimators: Tables

Table 2.5. Absolute bias of OLS, 2SLS and two-stage estimators for $\beta = 1$.

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
OLS		0	3.36E-06	4.97E-02	9.98E-02	5.00E-01	6.00E-01	9.51E-01
		13	-5.41E-04	4.91E-02	9.92E-02	5.00E-01	6.00E-01	9.50E-01
		200	-7.03E-05	4.99E-02	1.00E-01	4.99E-01	6.00E-01	9.49E-01
		613	1.48E-04	5.03E-02	1.00E-01	4.99E-01	5.99E-01	9.48E-01
		1000	-9.73E-05	5.03E-02	9.94E-02	4.97E-01	5.98E-01	9.47E-01
		2000000	-8.45E-05	5.40E-03	1.12E-02	5.53E-02	6.66E-02	1.05E-01
2SLS		0	3.40E-03	4.57E-02	1.08E-01	4.93E-01	5.95E-01	9.35E-01
		13	1.23E-03	5.71E-02	1.02E-01	5.09E-01	5.99E-01	9.45E-01
		200	4.81E-03	4.29E-02	9.24E-02	4.59E-01	5.58E-01	8.77E-01
		613	1.04E-03	4.41E-02	7.60E-02	4.01E-01	4.77E-01	7.49E-01
		1000	-3.41E-04	3.29E-02	6.69E-02	3.47E-01	4.15E-01	6.56E-01
		2000000	-5.85E-05	-8.96E-05	8.52E-05	2.43E-04	5.15E-04	4.53E-04
two-stage	Pre-tests $\mathcal{T}_1$	0	1.66E-04	4.95E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
		13	-4.47E-04	4.95E-02	9.94E-02	5.01E-01	6.00E-01	9.50E-01
		200	1.82E-04	4.96E-02	9.98E-02	4.97E-01	5.98E-01	9.45E-01
		613	1.94E-04	5.00E-02	9.92E-02	4.93E-01	5.92E-01	9.35E-01
		1000	-1.09E-04	4.94E-02	9.78E-02	4.88E-01	5.86E-01	9.26E-01
		2000000	-8.32E-05	4.70E-03	7.10E-03	2.43E-04	5.15E-04	4.53E-04
	$\mathcal{T}_2$	0	1.67E-04	4.95E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
		13	-4.52E-04	4.94E-02	9.94E-02	5.01E-01	6.00E-01	9.50E-01
		200	1.87E-04	4.96E-02	9.97E-02	4.97E-01	5.98E-01	9.44E-01
		613	1.93E-04	5.00E-02	9.90E-02	4.93E-01	5.90E-01	9.29E-01
		1000	-1.09E-04	4.93E-02	9.78E-02	4.86E-01	5.82E-01	9.08E-01
		2000000	-8.32E-05	4.40E-03	4.90E-03	2.43E-04	5.15E-04	4.53E-04
	$\mathcal{T}_3$	0	1.29E-05	4.97E-02	9.98E-02	5.00E-01	6.00E-01	9.51E-01
		13	-5.36E-04	4.91E-02	9.92E-02	5.00E-01	6.00E-01	9.50E-01
		200	-5.23E-05	4.99E-02	1.00E-01	4.99E-01	6.00E-01	9.48E-01
		613	1.51E-04	5.03E-02	1.00E-01	4.98E-01	5.98E-01	9.45E-01
		1000	-9.88E-05	5.01E-02	9.93E-02	4.95E-01	5.95E-01	9.40E-01
		2000000	-8.32E-05	4.40E-03	4.90E-03	2.43E-04	5.15E-04	4.53E-04
	$\mathcal{T}_4$	0	1.67E-04	4.95E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
		13	-4.52E-04	4.94E-02	9.94E-02	5.01E-01	6.00E-01	9.50E-01
		200	1.87E-04	4.96E-02	9.97E-02	4.97E-01	5.98E-01	9.44E-01
		613	1.92E-04	5.00E-02	9.90E-02	4.93E-01	5.90E-01	9.29E-01
		1000	-1.09E-04	4.93E-02	9.78E-02	4.86E-01	5.82E-01	9.08E-01
		2000000	-8.32E-05	4.40E-03	4.90E-03	2.43E-04	5.15E-04	4.53E-04

Table 2.5 (continued). Absolute bias of OLS, 2SLS and two-stage estimators for $\beta = 1$.

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	1.22E-05	4.97E-02	9.98E-02	5.00E-01	6.00E-01	9.51E-01
	13	-5.36E-04	4.91E-02	9.92E-02	5.00E-01	6.00E-01	9.50E-01
	200	-5.23E-05	4.99E-02	1.00E-01	4.99E-01	6.00E-01	9.48E-01
$\mathcal{H}_1$	613	1.51E-04	5.03E-02	1.00E-01	4.98E-01	5.98E-01	9.45E-01
	1000	-9.87E-05	5.01E-02	9.93E-02	4.95E-01	5.95E-01	9.40E-01
	2000000	-8.32E-05	4.50E-03	5.00E-03	2.43E-04	5.15E-04	4.53E-04
	0	1.29E-05	4.97E-02	9.98E-02	5.00E-01	6.00E-01	9.51E-01
	13	-5.36E-04	4.91E-02	9.92E-02	5.00E-01	6.00E-01	9.50E-01
	200	-5.23E-05	4.99E-02	1.00E-01	4.99E-01	6.00E-01	9.48E-01
$\mathcal{H}_2$	613	1.51E-04	5.03E-02	1.00E-01	4.98E-01	5.98E-01	9.45E-01
	1000	-9.88E-05	5.01E-02	9.92E-02	4.95E-01	5.95E-01	9.40E-01
	2000000	-8.32E-05	4.40E-03	4.90E-03	2.43E-04	5.15E-04	4.53E-04
	0	1.67E-04	4.95E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
	13	-4.51E-04	4.95E-02	9.94E-02	5.01E-01	6.00E-01	9.50E-01
	200	1.87E-04	4.96E-02	9.97E-02	4.97E-01	5.98E-01	9.44E-01
$\mathcal{H}_3$	613	1.93E-04	5.00E-02	9.90E-02	4.93E-01	5.90E-01	9.29E-01
	1000	-1.09E-04	4.93E-02	9.78E-02	4.86E-01	5.82E-01	9.08E-01
	2000000	-8.32E-05	4.40E-03	4.90E-03	2.43E-04	5.15E-04	4.53E-04
	0	1.78E-04	4.96E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
	13	-4.54E-04	4.95E-02	9.94E-02	5.01E-01	6.00E-01	9.50E-01
	200	1.84E-04	4.96E-02	9.98E-02	4.97E-01	5.98E-01	9.44E-01
$\mathcal{RH}$	613	1.91E-04	5.00E-02	9.92E-02	4.93E-01	5.90E-01	9.26E-01
	1000	-1.09E-04	4.94E-02	9.78E-02	4.86E-01	5.83E-01	9.05E-01
	2000000	-8.32E-05	4.90E-03	7.60E-03	2.43E-04	5.15E-04	4.53E-04

Table 2.6. MSE of OLS, 2SLS and two-stage estimators for $\beta = 1$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
OLS		0	2.00E-03	4.50E-03	1.20E-02	2.52E-01	3.62E-01	9.07E-01
		13	2.00E-03	4.50E-03	1.19E-02	2.52E-01	3.62E-01	9.04E-01
		200	2.00E-03	4.49E-03	1.20E-02	2.51E-01	3.62E-01	9.02E-01
		613	2.00E-03	4.54E-03	1.21E-02	2.51E-01	3.60E-01	9.00E-01
		1000	1.90E-03	4.60E-03	1.19E-02	2.49E-01	3.59E-01	8.99E-01
		2000000	2.21E-04	2.54E-04	3.53E-04	3.30E-03	4.70E-03	1.14E-02
2SLS		0	3.29E-01	3.33E-01	3.54E-01	5.68E-01	6.86E-01	1.21E+00
		13	3.30E-01	3.26E-01	3.46E-01	5.92E-01	7.06E-01	1.23E+00
		200	3.01E-01	3.08E-01	3.24E-01	5.31E-01	6.39E-01	1.09E+00
		613	2.67E-01	2.61E-01	2.91E-01	4.38E-01	5.16E-01	8.77E-01
		1000	2.33E-01	2.38E-01	2.35E-01	3.71E-01	4.38E-01	7.16E-01
		2000000	2.50E-04	2.47E-04	2.57E-04	3.24E-04	3.46E-04	4.79E-04
two-stage	Pre-tests	0	2.60E-03	5.00E-03	1.24E-02	2.41E-01	3.46E-01	8.65E-01
		13	2.80E-03	5.00E-03	1.21E-02	2.42E-01	3.46E-01	8.62E-01
		200	2.70E-03	5.10E-03	1.22E-02	2.39E-01	3.44E-01	8.56E-01
	$\mathcal{T}_1$	613	2.60E-03	4.90E-03	1.21E-02	2.36E-01	3.39E-01	8.35E-01
		1000	2.50E-03	4.80E-03	1.18E-02	2.31E-01	3.32E-01	8.19E-01
		2000000	2.11E-04	2.21E-04	2.30E-04	3.24E-04	3.46E-04	4.79E-04
		0	2.60E-03	5.10E-03	1.24E-02	2.41E-01	3.46E-01	8.63E-01
		13	2.70E-03	5.00E-03	1.22E-02	2.41E-01	3.46E-01	8.60E-01
		200	2.70E-03	5.10E-03	1.22E-02	2.38E-01	3.43E-01	8.46E-01
	$\mathcal{T}_2$	613	2.60E-03	5.00E-03	1.22E-02	2.35E-01	3.34E-01	8.06E-01
		1000	2.50E-03	4.90E-03	1.17E-02	2.27E-01	3.23E-01	7.59E-01
		2000000	2.11E-04	2.09E-04	2.05E-04	3.24E-04	3.46E-04	4.79E-04
		0	2.00E-03	4.40E-03	1.19E-02	2.51E-01	3.61E-01	9.05E-01
		13	2.00E-03	4.40E-03	1.18E-02	2.52E-01	3.61E-01	9.02E-01
		200	2.00E-03	4.50E-03	1.19E-02	2.50E-01	3.60E-01	8.98E-01
	$\mathcal{T}_3$	613	2.00E-03	4.50E-03	1.21E-02	2.49E-01	3.57E-01	8.85E-01
		1000	1.90E-03	4.50E-03	1.18E-02	2.46E-01	3.53E-01	8.70E-01
		2000000	2.11E-04	2.10E-04	2.05E-04	3.24E-04	3.46E-04	4.79E-04
		0	2.60E-03	5.10E-03	1.24E-02	2.41E-01	3.46E-01	8.63E-01
		13	2.70E-03	5.00E-03	1.22E-02	2.41E-01	3.46E-01	8.60E-01
		200	2.70E-03	5.10E-03	1.22E-02	2.38E-01	3.43E-01	8.47E-01
	$\mathcal{T}_4$	613	2.60E-03	5.00E-03	1.22E-02	2.35E-01	3.34E-01	8.06E-01
		1000	2.50E-03	4.90E-03	1.17E-02	2.27E-01	3.23E-01	7.59E-01
		2000000	2.11E-04	2.09E-04	2.05E-04	3.24E-04	3.46E-04	4.79E-04

Table 2.6 (Continued). MSE of OLS, 2SLS and two-stage estimators for $\beta = 1$

Estimators	$\mu^2 \downarrow, \ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	2.00E-03	4.40E-03	1.19E-02	2.51E-01	3.61E-01	9.05E-01
	13	2.00E-03	4.40E-03	1.18E-02	2.52E-01	3.61E-01	9.02E-01
	200	2.00E-03	4.50E-03	1.19E-02	2.50E-01	3.60E-01	8.98E-01
$\mathcal{H}_1$	613	2.00E-03	4.50E-03	1.21E-02	2.49E-01	3.57E-01	8.85E-01
	1000	1.90E-03	4.50E-03	1.18E-02	2.46E-01	3.53E-01	8.70E-01
	2000000	2.11E-04	2.10E-04	2.05E-04	3.24E-04	3.46E-04	4.79E-04
	0	2.00E-03	4.40E-03	1.19E-02	2.51E-01	3.61E-01	9.05E-01
	13	2.00E-03	4.40E-03	1.18E-02	2.52E-01	3.61E-01	9.01E-01
	200	2.00E-03	4.50E-03	1.19E-02	2.50E-01	3.60E-01	8.98E-01
$\mathcal{H}_2$	613	2.00E-03	4.50E-03	1.21E-02	2.49E-01	3.57E-01	8.85E-01
	1000	1.90E-03	4.50E-03	1.18E-02	2.46E-01	3.53E-01	8.70E-01
	2000000	2.11E-04	2.09E-04	2.05E-04	3.24E-04	3.46E-04	4.79E-04
	0	2.60E-03	5.10E-03	1.24E-02	2.41E-01	3.46E-01	8.63E-01
	13	2.70E-03	5.00E-03	1.22E-02	2.41E-01	3.46E-01	8.60E-01
	200	2.70E-03	5.10E-03	1.22E-02	2.38E-01	3.43E-01	8.46E-01
$\mathcal{H}_3$	613	2.60E-03	5.00E-03	1.22E-02	2.35E-01	3.34E-01	8.05E-01
	1000	2.50E-03	4.90E-03	1.18E-02	2.27E-01	3.23E-01	7.59E-01
	2000000	2.11E-04	2.09E-04	2.05E-04	3.24E-04	3.46E-04	4.79E-04
	0	2.70E-03	5.00E-03	1.23E-02	2.41E-01	3.44E-01	8.63E-01
	13	2.70E-03	5.00E-03	1.22E-02	2.41E-01	3.45E-01	8.62E-01
	200	2.70E-03	5.00E-03	1.22E-02	2.38E-01	3.43E-01	8.44E-01
$\mathcal{RH}$	613	2.50E-03	5.00E-03	1.21E-02	2.34E-01	3.31E-01	7.95E-01
	1000	2.40E-03	4.90E-03	1.18E-02	2.28E-01	3.24E-01	7.49E-01
	2000000	2.11E-04	2.27E-04	2.40E-04	3.24E-04	3.46E-04	4.79E-04

Table 2.7. Relative bias of OLS and two-stage estimators compared with 2SLS for $\beta = 1$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
OLS		0	0.00	1.09	0.92	1.01	1.01	1.01
		13	-0.44	0.86	0.97	0.98	1.00	1.00
		200	-0.01	1.16	1.08	1.09	1.07	1.07
		613	0.14	1.14	1.32	1.24	1.25	1.25
		1000	0.29	1.53	1.49	1.43	1.44	1.44
		2000000	1.44	-60.29	131.51	227.37	129.21	129.21
two-Stage	Pre-tests	0	0.05	1.08	0.93	1.01	1.01	1.01
		13	-0.36	0.87	0.98	0.98	1.00	1.00
		200	0.04	1.16	1.08	1.08	1.07	1.07
	$\mathcal{T}_1$	613	0.19	1.13	1.31	1.23	1.24	1.24
		1000	0.32	1.50	1.46	1.41	1.41	1.41
		2000000	1.42	-52.47	83.37	1.00	1.00	1.00
		0	0.05	1.08	0.93	1.01	1.01	1.01
		13	-0.37	0.87	0.98	0.98	1.00	1.00
		200	0.04	1.16	1.08	1.08	1.07	1.07
	$\mathcal{T}_2$	613	0.19	1.13	1.30	1.23	1.24	1.24
		1000	0.32	1.50	1.46	1.40	1.40	1.40
		2000000	1.42	-49.12	57.54	1.00	1.00	1.00
		0	0.00	1.09	0.92	1.01	1.01	1.01
		13	-0.44	0.86	0.97	0.98	1.00	1.00
		200	-0.01	1.16	1.08	1.09	1.07	1.07
	$\mathcal{T}_3$	613	0.15	1.14	1.32	1.24	1.25	1.25
		1000	0.29	1.52	1.48	1.43	1.43	1.43
		2000000	1.42	-49.12	57.54	1.00	1.00	1.00
		0	0.05	1.08	0.93	1.01	1.01	1.01
		13	-0.37	0.87	0.98	0.98	1.00	1.00
		200	0.04	1.16	1.08	1.08	1.07	1.07
	$\mathcal{T}_4$	613	0.19	1.13	1.30	1.23	1.24	1.24
		1000	0.32	1.50	1.46	1.40	1.40	1.40
		2000000	1.42	-49.12	57.54	1.00	1.00	1.00

Table 2.7 (Continued). Relative bias of OLS and two-stage estimators compared with 2SLS for $\beta = 1$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	0.00	1.09	0.92	1.01	1.01	1.01
	13	-0.44	0.86	0.97	0.98	1.00	1.00
	200	-0.01	1.16	1.08	1.09	1.07	1.07
$\mathcal{H}_1$	613	0.15	1.14	1.32	1.24	1.25	1.25
	1000	0.29	1.52	1.48	1.43	1.43	1.43
	2000000	1.42	-50.24	58.71	1.00	1.00	1.00
	0	0.00	1.09	0.92	1.01	1.01	1.01
	13	-0.44	0.86	0.97	0.98	1.00	1.00
	200	-0.01	1.16	1.08	1.09	1.07	1.07
$\mathcal{H}_2$	613	0.15	1.14	1.32	1.24	1.25	1.25
	1000	0.29	1.52	1.48	1.43	1.43	1.43
	2000000	1.42	-49.12	57.54	1.00	1.00	1.00
	0	0.05	1.08	0.93	1.01	1.01	1.01
	13	-0.37	0.87	0.98	0.98	1.00	1.00
	200	0.04	1.16	1.08	1.08	1.07	1.07
$\mathcal{H}_3$	613	0.19	1.13	1.30	1.23	1.24	1.24
	1000	0.32	1.50	1.46	1.40	1.40	1.40
	2000000	1.42	-49.12	57.54	1.00	1.00	1.00
	0	0.05	1.09	0.93	1.01	1.01	1.01
	13	-0.37	0.87	0.98	0.98	1.00	1.00
	200	0.04	1.16	1.08	1.08	1.07	1.07
$\mathcal{RH}$	613	0.18	1.13	1.31	1.23	1.24	1.24
	1000	0.32	1.50	1.46	1.40	1.40	1.40
	2000000	1.42	-54.70	89.24	1.00	1.00	1.00

Table 2.8. Relative MSE of OLS and two-stage estimators compared with 2SLS for $\beta = 1$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
OLS		0	0.01	0.01	0.01	0.44	0.53	0.75
		13	0.01	0.01	0.01	0.43	0.51	0.73
		200	0.01	0.01	0.01	0.47	0.57	0.83
		613	0.01	0.02	0.02	0.57	0.70	1.03
		1000	0.01	0.02	0.02	0.67	0.82	1.26
		2000000	0.89	1.03	1.03	10.19	13.57	23.78
two-Stage	Pre-tests	0	0.01	0.02	0.02	0.43	0.50	0.72
		13	0.01	0.02	0.02	0.41	0.49	0.70
		200	0.01	0.02	0.02	0.45	0.54	0.79
	$\mathcal{T}_1$	613	0.01	0.02	0.02	0.54	0.66	0.95
		1000	0.01	0.02	0.02	0.62	0.76	1.14
		2000000	0.84	0.89	0.89	1.00	1.00	1.00
		0	0.01	0.02	0.02	0.42	0.50	0.71
		13	0.01	0.02	0.02	0.41	0.49	0.70
		200	0.01	0.02	0.02	0.45	0.54	0.78
	$\mathcal{T}_2$	613	0.01	0.02	0.02	0.54	0.65	0.92
		1000	0.01	0.02	0.02	0.61	0.74	1.06
		2000000	0.84	0.85	0.85	1.00	1.00	1.00
		0	0.01	0.01	0.01	0.44	0.53	0.75
		13	0.01	0.01	0.01	0.42	0.51	0.73
		200	0.01	0.01	0.01	0.47	0.56	0.82
	$\mathcal{T}_3$	613	0.01	0.02	0.02	0.57	0.69	1.01
		1000	0.01	0.02	0.02	0.66	0.81	1.22
		2000000	0.84	0.85	0.85	1.00	1.00	1.00
		0	0.01	0.02	0.02	0.42	0.50	0.71
		13	0.01	0.02	0.02	0.41	0.49	0.70
		200	0.01	0.02	0.02	0.45	0.54	0.78
	$\mathcal{T}_4$	613	0.01	0.02	0.02	0.54	0.65	0.92
		1000	0.01	0.02	0.02	0.61	0.74	1.06
		2000000	0.84	0.85	0.85	1.00	1.00	1.00

Table 2.8 (Continued). Relative MSE of OLS and two-stage estimators compared with 2SLS for $\beta = 1$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	0.01	0.01	0.01	0.44	0.53	0.75
	13	0.01	0.01	0.01	0.42	0.51	0.73
	200	0.01	0.01	0.01	0.47	0.56	0.82
$\mathcal{H}_1$	613	0.01	0.02	0.02	0.57	0.69	1.01
	1000	0.01	0.02	0.02	0.66	0.81	1.22
	2000000	0.84	0.85	0.85	1.00	1.00	1.00
	0	0.01	0.01	0.01	0.44	0.53	0.75
	13	0.01	0.01	0.01	0.42	0.51	0.73
	200	0.01	0.01	0.01	0.47	0.56	0.82
$\mathcal{H}_2$	613	0.01	0.02	0.02	0.57	0.69	1.01
	1000	0.01	0.02	0.02	0.66	0.80	1.21
	2000000	0.84	0.85	0.85	1.00	1.00	1.00
	0	0.01	0.02	0.02	0.42	0.50	0.71
	13	0.01	0.02	0.02	0.41	0.49	0.70
	200	0.01	0.02	0.02	0.45	0.54	0.78
$\mathcal{H}_3$	613	0.01	0.02	0.02	0.54	0.65	0.92
	1000	0.01	0.02	0.02	0.61	0.74	1.06
	2000000	0.84	0.85	0.85	1.00	1.00	1.00
	0	0.01	0.02	0.02	0.42	0.50	0.71
	13	0.01	0.02	0.02	0.41	0.49	0.70
	200	0.01	0.02	0.02	0.45	0.54	0.77
$\mathcal{RH}$	613	0.01	0.02	0.02	0.53	0.64	0.91
	1000	0.01	0.02	0.02	0.61	0.74	1.05
	2000000	0.84	0.92	0.92	1.00	1.00	1.00

Table 2.9. Relative MSE of 2SLS and two-stage estimators compared with OLS for $\beta = 1$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
		0	164.35	73.93	29.53	2.25	1.89	1.33
		13	164.95	72.53	29.04	2.35	1.95	1.36
		200	150.60	68.53	26.97	2.11	1.77	1.21
2SLS		613	133.25	57.62	24.01	1.75	1.43	0.97
		1000	122.47	51.80	19.75	1.75	1.43	0.97
		2000000	1.13	0.98	0.73	0.10	0.07	0.04
	Pre-tests	0	1.30	1.11	1.03	0.96	0.95	0.95
two-Stage		13	1.40	1.11	1.02	0.96	0.95	0.95
		200	1.35	1.14	1.02	0.95	0.95	0.93
	$\mathcal{T}_1$	613	1.30	1.08	1.00	0.94	0.94	0.00
		1000	1.32	1.04	0.99	0.93	0.92	0.91
		2000000	0.95	0.87	0.65	0.10	0.07	0.00
		0	1.30	1.13	1.03	0.96	0.95	0.95
		13	1.35	1.11	1.03	0.96	0.95	0.94
		200	1.35	1.14	1.02	0.95	0.95	0.89
	$\mathcal{T}_2$	613	1.30	1.10	1.01	0.94	0.93	0.00
		1000	1.32	1.07 0.98	0.91	0.90	0.84	
		2000000	0.95	0.83	0.58	0.10	0.07	0.00
		0	1.00	0.98	0.99	1.00	1.00	0.09
		13	1.00	0.98	0.99	1.00	1.00	0.09
		200	1.00	1.00	0.99	1.00	1.00	0.09
	$\mathcal{T}_3$	613	1.00	0.99	1.00	0.99	0.99	0.00
		1000	1.00	0.98	0.99	0.99	0.98	0.97
		2000000	0.95	0.83	0.58	0.10	0.07	0.00
		0	1.30	1.13	1.03	0.96	0.95	0.95
		13	1.35	1.11	1.03	0.96	0.95	0.94
		200	1.35	1.14	1.02	0.95	0.95	0.89
	$\mathcal{T}_4$	613	1.30	1.10	1.01	0.94	0.93	0.00
		1000	1.32	1.07	0.98	0.91	0.90	0.84
		2000000	0.95	0.83	0.58	0.10	0.07	0.00

Table 2.9 (Continued). Relative MSE of 2SLS and two-stage estimators compared with OLS for $\beta = 1$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	1.00	0.98	0.99	1.00	1.00	0.99
	13	1.00	0.98	0.99	1.00	1.00	0.99
	200	1.00	1.00	0.99	1.00	1.00	0.98
$\mathcal{H}_1$	613	1.00	0.99	1.00	0.99	0.99	0.00
	1000	1.00	0.98	0.99	0.99	0.98	0.97
	2000000	0.95	0.83	0.58	0.10	0.07	0.00
	0	1.00	0.98	0.99	1.00	1.00	0.99
	13	1.00	0.98	0.99	1.00	1.00	0.99
	200	1.00	1.00	0.99	1.00	1.00	0.98
$\mathcal{H}_2$	613	1.00	0.99	1.00	0.99	0.99	0.00
	1000	1.00	0.98	0.99	0.99	0.98	0.97
	2000000	0.95	0.83	0.58	0.10	0.07	0.00
	0	1.30	1.13	1.03	0.96	0.95	0.95
	13	1.35	1.11	1.03	0.96	0.95	0.94
	200	1.35	1.14	1.02	0.95	0.95	0.89
$\mathcal{H}_3$	613	1.30	1.10	1.01	0.94	0.93	0.00
	1000	1.32	1.07	0.99	0.91	0.90	0.84
	2000000	0.95	0.83	0.58	0.10	0.07	0.00
	0	1.35	1.11	1.03	0.96	0.95	0.95
	13	1.35	1.11	1.03	0.96	0.95	0.93
	200	1.35	1.11	1.02	0.95	0.95	0.88
$\mathcal{RH}$	613	1.25	1.10	1.00	0.93	0.92	0.00
	1000	1.26	1.07	0.99	0.91	0.90	0.83
	2000000	0.95	0.89	0.68	0.10	0.07	0.00

Table 2.10. Absolute bias of OLS, 2SLS and two-stage estimators for $\beta = 0$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
		0	-1.91E-04	4.99E-02	9.96E-02	5.00E-01	6.01E-01	9.50E-01
		13	-6.46E-04	5.02E-02	1.01E-01	4.99E-01	6.00E-01	9.50E-01
		200	3.98E-05	4.95E-02	1.00E-01	4.99E-01	5.99E-01	9.49E-01
OLS		613	3.23E-05	4.95E-02	9.94E-02	4.98E-01	5.99E-01	9.47E-01
		1000	-3.06E-04	4.99E-02	9.95E-02	4.98E-01	5.97E-01	9.46E-01
		2000000	-1.44E-05	5.90E-03	1.13E-02	5.54E-02	6.63E-02	1.06E-01
		0	2.00E-03	4.73E-02	1.05E-01	5.02E-01	5.85E-01	9.47E-01
		13	-2.04E-04	5.41E-02	1.02E-01	5.03E-01	5.97E-01	9.49E-01
		200	-7.40E-03	4.41E-02	9.45E-02	4.63E-01	5.53E-01	8.75E-01
2SLS		613	1.20E-03	3.62E-02	8.42E-02	3.95E-01	4.70E-01	7.49E-01
		1000	-3.90E-03	3.34E-02	5.80E-02	3.47E-01	4.12E-01	6.58E-01
		2000000	-2.52E-05	3.29E-04	2.79E-04	4.07E-04	2.84E-04	8.94E-04
	Pre-tests	0	-7.43E-05	4.98E-02	9.99E-02	5.00E-01	6.00E-01	9.50E-01
two-stage		13	-6.23E-04	5.04E-02	1.01E-01	5.00E-01	6.00E-01	9.50E-01
		200	-3.58E-04	4.92E-02	9.99E-02	4.97E-01	5.97E-01	9.45E-01
	$\mathcal{T}_1$	613	9.12E-05	4.89E-02	9.86E-02	4.93E-01	5.91E-01	9.36E-01
		1000	-4.95E-04	4.91E-02	9.75E-02	4.89E-01	5.85E-01	9.26E-01
		2000000	-1.49E-05	5.10E-03	7.40E-03	4.07E-04	2.84E-04	8.94E-04
		0	-7.65E-05	4.98E-02	9.99E-02	5.00E-01	6.00E-01	9.50E-01
		13	-6.23E-01	5.04E-02	1.01E-01	5.00E-01	6.00E-01	9.50E-01
		200	-3.54E-04	4.92E-02	9.99E-02	4.97E-01	5.96E-01	9.44E-01
	$\mathcal{T}_2$	613	8.97E-05	4.89E-02	9.86E-02	4.91E-01	5.89E-01	9.29E-01
		1000	-4.92E-04	4.91E-02	9.74E-02	4.87E-01	5.82E-01	9.10E-01
		2000000	-1.50E-05	4.80E-03	5.20E-03	4.07E-04	2.84E-04	8.94E-04
		0	-1.85E-04	4.99E-02	9.96E-02	5.00E-01	6.01E-01	9.50E-01
		13	-6.45E-04	5.02E-02	1.01E-01	4.99E-01	6.00E-01	9.50E-01
		200	1.59E-05	4.95E-02	1.00E-01	4.99E-01	5.99E-01	9.49E-01
	$\mathcal{T}_3$	613	3.78E-05	4.94E-02	9.94E-02	4.97E-01	5.98E-01	9.45E-01
		1000	-3.28E-04	4.98E-02	9.93E-02	4.97E-01	5.95E-01	9.39E-01
		2000000	-1.50E-05	4.80E-03	5.20E-03	4.07E-04	2.84E-04	8.94E-04
		0	-7.65E-05	4.98E-02	9.99E-02	5.00E-01	6.00E-01	9.50E-01
		13	-6.23E-01	5.04E-02	1.01E-01	5.00E-01	6.00E-01	9.50E-01
		200	-3.54E-04	4.92E-02	9.99E-02	4.97E-01	5.96E-01	9.48E-01
	$\mathcal{T}_4$	613	8.97E-05	4.89E-02	9.86E-02	4.91E-01	5.89E-01	9.29E-01
		1000	-4.91E-04	4.91E-02	9.73E-02	4.87E-01	5.82E-01	9.10E-01
		2000000	-1.50E-05	4.80E-03	5.20E-03	4.07E-04	2.84E-04	8.94E-04
		2000000	-1.50E-05	5.30E-03	7.70E-03	4.07E-04	2.84E-04	8.94E-04

Table 2.10 (Continued). Absolute bias of OLS, 2SLS and two-stage estimators for $\beta = 0$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	-1.85E-04	4.99E-02	9.96E-02	5.00E-01	6.01E-01	9.50E-01
	13	-6.45E-04	5.02E-02	1.01E-01	4.99E-01	6.00E-01	9.50E-01
	200	1.74E-05	4.95E-02	1.00E-01	4.99E-01	5.99E-01	9.49E-01
$\mathcal{H}_1$	613	3.78E-05	4.94E-02	9.94E-02	4.97E-01	5.98E-01	9.45E-01
	1000	-3.27E-04	4.98E-02	9.93E-02	4.97E-01	5.95E-01	9.39E-01
	2000000	-1.49E-05	4.80E-03	5.20E-03	4.07E-04	2.84E-04	8.94E-04
	0	-1.85E-04	4.99E-02	9.96E-02	5.00E-01	6.01E-01	9.50E-01
	13	-6.45E-04	5.02E-02	1.01E-01	4.99E-01	6.00E-01	9.50E-01
	200	1.59E-05	4.95E-02	1.00E-01	4.99E-01	5.99E-01	9.44E-01
$\mathcal{H}_2$	613	3.78E-05	4.94E-02	9.93E-02	4.97E-01	5.98E-01	9.45E-01
	1000	-3.29E-04	4.98E-02	9.93E-02	4.97E-01	5.95E-01	9.39E-01
	2000000	-1.50E-05	4.80E-03	5.20E-03	4.07E-04	2.84E-04	8.94E-04
	0	-7.59E-05	4.98E-02	9.99E-02	5.00E-01	6.00E-01	9.50E-01
	13	-6.23E-01	5.04E-02	1.01E-01	5.00E-01	6.00E-01	9.50E-01
	200	-3.57E-04	4.92E-02	9.99E-02	4.97E-01	5.96E-01	9.48E-01
$\mathcal{H}_3$	613	9.00E-05	4.89E-02	9.86E-02	4.91E-01	5.89E-01	9.29E-01
	1000	4.92E-04	4.91E-02	9.74E-02	4.87E-01	5.82E-01	9.10E-01
	2000000	-1.50E-05	4.80E-03	5.20E-03	4.07E-04	2.84E-04	8.94E-04
	0	-8.01E-05	4.98E-02	9.99E-02	5.00E-01	6.00E-01	9.50E-01
	13	-6.24E-04	5.04E-02	1.01E-01	5.00E-01	6.00E-01	9.50E-01
	200	-3.43E-04	4.92E-02	9.99E-02	4.97E-01	5.96E-01	9.44E-01
$\mathcal{RH}$	613	9.43E-05	4.88E-02	9.87E-02	4.91E-01	5.89E-01	9.28E-01
	1000	-4.82E-04	4.91E-02	9.73E-02	4.87E-01	5.82E-01	9.06E-01
	2000000	-1.50E-05	5.30E-03	7.70E-03	4.07E-04	2.84E-04	8.94E-04

Table 2.11. MSE of OLS, 2SLS and two-stage estimators for $\beta = 0$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
		0	2.00E-03	4.50E-03	1.19E-02	2.52E-01	3.64E-01	9.04E-01
		13	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.62E-01	9.05E-01
		200	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.61E-01	9.02E-01
MCO		613	2.00E-03	4.40E-03	1.18E-02	2.50E-01	3.60E-01	8.99E-01
		1000	2.00E-03	4.50E-03	1.19E-02	2.50E-01	3.59E-01	8.98E-01
		2000000	2.19E-04	2.58E-04	3.55E-04	3.30E-03	4.70E-03	1.15E-02
		0	3.40E-01	3.40E-01	3.37E-01	5.81E-01	6.77E-01	1.23E+00
		13	3.33E-01	3.31E-01	3.56E-01	5.97E-01	6.94E-01	1.23E+00
		200	3.06E-01	3.11E-01	3.22E-01	5.22E-01	6.18E-01	1.10E+00
2SLS		613	2.73E-01	2.57E-01	2.69E-01	3.95E-01	5.09E-01	8.60E-01
		1000	2.29E-01	2.34E-01	2.40E-01	3.69E-01	4.22E-01	7.18E-01
		2000000	2.48E-04	2.51E-04	2.56E-04	3.17E-04	3.48E-04	4.77E-04
		0	2.80E-03	5.10E-03	1.21E-02	5.00E-01	3.47E-01	8.62E-01
two-stage	Pre-tests	13	2.80E-03	5.10E-03	1.23E-02	2.40E-01	3.45E-01	8.63E-01
		200	2.70E-03	5.10E-03	1.22E-02	2.39E-01	3.43E-01	8.56E-01
	$\mathcal{T}_1$	613	2.60E-03	4.80E-03	1.19E-02	2.36E-01	3.37E-01	8.40E-01
		1000	2.50E-03	4.80E-03	1.17E-02	2.33E-01	3.30E-01	8.20E-01
		2000000	2.09E-04	2.25E-04	2.33E-04	3.17E-04	3.48E-04	4.77E-04
		0	2.80E-03	5.10E-03	1.22E-02	5.00E-01	3.46E-01	8.61E-01
		13	2.80E-03	5.10E-03	1.24E-02	2.40E-01	3.45E-01	8.64E-01
		200	2.70E-03	5.00E-03	1.22E-02	2.39E-01	3.41E-01	8.46E-01
	$\mathcal{T}_2$	613	2.60E-03	4.80E-03	1.19E-02	2.33E-01	3.32E-01	8.09E-01
		1000	2.50E-03	4.80E-03	1.17E-02	2.29E-01	3.23E-01	7.66E-01
		2000000	2.08E-04	2.14E-04	2.06E-04	3.17E-04	3.48E-04	4.77E-04
		0	2.00E-03	4.50E-03	1.19E-02	5.00E-01	3.63E-01	9.02E-01
		13	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.61E-01	9.02E-01
		200	2.00E-03	4.50E-03	1.20E-02	2.50E-01	3.59E-01	8.98E-01
	$\mathcal{T}_3$	613	2.00E-03	4.40E-03	1.18E-02	2.48E-01	3.57E-01	8.85E-01
		1000	2.00E-03	4.50E-03	1.18E-02	2.47E-01	3.52E-01	8.68E-01
		2000000	2.08E-04	2.14E-04	2.06E-04	3.17E-04	3.48E-04	4.77E-04
		0	2.80E-03	5.10E-03	1.22E-02	5.00E-01	3.46E-01	8.61E-01
		13	2.80E-03	5.10E-03	1.24E-02	2.40E-01	3.45E-01	8.64E-01
		200	2.70E-03	5.00E-03	1.22E-02	2.39E-01	3.41E-01	8.46E-01
	$\mathcal{T}_4$	613	2.60E-03	4.80E-03	1.19E-02	2.33E-01	3.32E-01	8.09E-01
		1000	2.50E-03	4.80E-03	1.17E-02	2.29E-01	3.23E-01	7.66E-01
		2000000	2.08E-04	2.14E-04	2.06E-04	3.17E-04	3.48E-04	4.77E-04

Table 2.11 (Continued). MSE of OLS, 2SLS and two-stage estimators for $\beta = 0$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
$\mathcal{H}_1$	0	2.00E-03	4.50E-03	1.19E-02	5.00E-01	3.63E-01	9.02E-01
	13	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.61E-01	9.02E-01
	200	2.00E-03	4.50E-03	1.20E-02	2.50E-01	3.59E-01	8.98E-01
	613	2.00E-03	4.40E-03	1.18E-02	2.48E-01	3.57E-01	8.86E-01
	1000	2.00E-03	4.50E-03	1.18E-02	2.47E-01	3.53E-01	8.69E-01
	2000000	2.09E-04	2.14E-04	2.06E-04	3.17E-04	3.48E-04	4.77E-04
$\mathcal{H}_2$	0	2.00E-03	4.50E-03	1.19E-02	5.00E-01	3.63E-01	9.02E-01
	13	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.61E-01	9.02E-01
	200	2.00E-03	4.50E-03	1.20E-02	2.50E-01	3.59E-01	8.98E-01
	613	2.00E-03	4.40E-03	1.18E-02	2.48E-01	3.57E-01	8.85E-01
	1000	2.00E-03	4.50E-03	1.18E-02	2.47E-01	3.52E-01	8.68E-01
	2000000	2.08E-04	2.14E-04	2.06E-04	3.17E-04	3.48E-04	4.77E-04
$\mathcal{H}_3$	0	2.80E-03	5.10E-03	1.21E-02	5.00E-01	3.46E-01	8.61E-01
	13	2.80E-03	5.10E-03	1.24E-02	2.40E-01	3.45E-01	8.63E-01
	200	2.70E-03	5.00E-03	1.22E-02	2.39E-01	3.41E-01	8.46E-01
	613	2.60E-03	4.80E-03	1.19E-02	2.33E-01	3.32E-01	8.09E-01
	1000	2.50E-03	4.80E-03	1.17E-02	2.29E-01	3.23E-01	7.66E-01
	2000000	2.08E-04	2.14E-04	2.06E-04	3.17E-04	3.48E-04	4.77E-04
$\mathcal{RH}$	0	2.70E-03	5.10E-03	1.22E-02	5.00E-01	3.46E-01	8.61E-01
	13	2.80E-03	5.10E-03	1.23E-02	2.40E-01	3.45E-01	8.65E-01
	200	2.70E-03	5.00E-03	1.22E-02	2.38E-01	3.41E-01	8.42E-01
	613	2.70E-03	4.80E-03	1.19E-02	2.32E-01	3.32E-01	8.01E-01
	1000	2.50E-03	4.80E-03	1.17E-02	2.29E-01	3.23E-01	7.52E-01
	2000000	2.08E-04	2.31E-04	2.39E-04	3.17E-04	3.48E-04	4.77E-04

Table 2.12. Relative bias of OLS and two-stage estimators compared with 2SLS for $\beta = 0$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
		0	-0.10	1.05	0.95	1.00	1.03	1.00
		13	-0.08	0.93	0.99	0.99	1.00	1.00
		200	-0.01	1.12	1.06	1.08	1.08	1.08
MCO		613	0.03	1.37	1.18	1.26	1.27	1.26
		1000	0.08	1.49	1.72	1.43	1.45	1.44
		2000000	0.57	17.94	40.56	136.17	233.26	118.05
		0	-0.04	1.05	0.95	1.00	1.03	1.00
two-stage	Pre-tests	13	-0.03	0.93	0.99	0.99	1.00	1.00
		200	0.05	1.12	1.06	1.07	1.08	1.08
	$\mathcal{T}_1$	613	0.08	1.35	1.17	1.25	1.26	1.25
		1000	0.13	1.47	1.68	1.41	1.42	1.41
		2000000	0.59	15.51	26.56	1.00	1.00	1.00
		0	-0.04	1.05	0.95	1.00	1.03	1.00
		13	-0.03	0.93	0.99	0.99	1.00	1.00
		200	0.05	1.12	1.06	1.07	1.08	1.08
	$\mathcal{T}_2$	613	0.07	1.35	1.17	1.24	1.25	1.24
		1000	0.13	1.47	1.68	1.40	1.41	1.38
		2000000	0.59	14.60	18.67	1.00	1.00	1.00
		0	-0.09	1.05	0.95	1.00	1.03	1.00
		13	-0.08	0.93	0.99	0.99	1.00	1.00
		200	0.00	1.12	1.06	1.08	1.08	1.08
	$\mathcal{T}_3$	613	0.03	1.36	1.18	1.26	1.27	1.26
		1000	0.08	1.49	1.71	1.43	1.44	1.43
		2000000	0.59	14.60	18.67	1.00	1.00	1.00
		0	-0.04	1.05	0.95	1.00	1.03	1.00
		13	-0.03	0.93	0.99	0.99	1.00	1.00
		200	0.05	1.12	1.06	1.07	1.08	1.08
	$\mathcal{T}_4$	613	0.07	1.35	1.17	1.24	1.25	1.24
		1000	0.13	1.47	1.68	1.40	1.41	1.38
		2000000	0.59	14.60	18.67	1.00	1.00	1.00

Table 2.12 (Continued). Relative bias of OLS and two-stage estimators compared with 2SLS for $\beta = 0$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	-0.09	1.05	0.95	1.00	1.03	1.00
	13	-0.08	0.93	0.99	0.99	1.00	1.00
	200	0.00	1.12	1.06	1.08	1.08	1.08
$\mathcal{H}_1$	613	0.03	1.36	1.18	1.26	1.27	1.26
	1000	0.08	1.49	1.71	1.43	1.44	1.43
	2000000	0.59	14.60	18.67	1.00	1.00	1.00
	0	-0.09	1.05	0.95	1.00	1.03	1.00
	13	-0.08	0.93	0.99	0.99	1.00	1.00
	200	0.00	1.12	1.06	1.08	1.08	1.08
$\mathcal{H}_2$	613	0.03	1.36	1.18	1.26	1.27	1.26
	1000	0.08	1.49	1.71	1.43	1.44	1.43
	2000000	0.59	14.60	18.67	1.00	1.00	1.00
	0	-0.04	1.05	0.95	1.00	1.03	1.00
	13	-0.03	0.93	0.99	0.99	1.00	1.00
	200	0.05	1.12	1.06	1.07	1.08	1.08
$\mathcal{H}_3$	613	0.08	1.35	1.17	1.24	1.25	1.24
	1000	0.08	1.47	1.68	1.40	1.41	1.38
	2000000	0.59	14.60	18.67	1.00	1.00	1.00
	0	-0.04	1.05	0.95	1.00	1.03	1.00
	13	-0.03	0.93	0.99	0.99	1.00	1.00
	200	0.05	1.12	1.06	1.07	1.08	1.08
$\mathcal{RH}$	613	0.08	1.35	1.17	1.24	1.25	1.24
	1000	-0.13	1.47	1.68	1.40	1.41	1.38
	2000000	0.59	16.12	27.64	1.00	1.00	1.00

Table 2.13. Relative MSE of OLS and two-stage estimators compared with 2SLS for $\beta = 0$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
		0	0.01	0.01	0.04	0.43	0.54	0.73
		13	0.01	0.01	0.03	0.42	0.52	0.74
		200	0.01	0.01	0.04	0.48	0.58	1.08
MCO		613	0.01	0.02	0.04	1.26	0.71	1.05
		1000	0.01	0.02	0.05	0.68	0.85	1.25
		2000000	0.88	1.03	1.38	10.42	13.52	24.09
		0	0.01	0.02	0.04	0.86	0.51	0.70
two-stage	Pre-tests	13	0.01	0.02	0.03	0.40	0.50	0.70
		200	0.01	0.02	0.04	0.46	0.56	1.08
	$\mathcal{T}_1$	613	0.01	0.02	0.04	1.25	0.66	0.98
		1000	0.01	0.02	0.05	0.63	0.78	1.14
		2000000	0.84	0.90	0.91	1.00	1.00	1.00
		0	0.01	0.02	0.04	0.86	0.51	0.70
		13	0.01	0.02	0.03	0.40	0.50	0.70
		200	0.01	0.02	0.04	0.46	0.55	1.08
	$\mathcal{T}_2$	613	0.01	0.02	0.04	1.24	0.65	0.94
		1000	0.01	0.02	0.05	0.62	0.77	1.07
		2000000	0.84	0.85	0.80	1.00	1.00	1.00
		0	0.01	0.01	0.04	0.86	0.54	0.73
		13	0.01	0.01	0.03	0.42	0.52	0.73
		200	0.01	0.01	0.04	0.48	0.58	1.08
	$\mathcal{T}_3$	613	0.01	0.02	0.04	1.26	0.70	1.03
		1000	0.01	0.02	0.05	0.67	0.84	1.21
		2000000	0.84	0.85	0.80	1.00	1.00	1.00
		0	0.01	0.02	0.04	0.86	0.51	0.70
		13	0.01	0.02	0.03	0.40	0.50	0.70
		200	0.01	0.02	0.04	0.46	0.58	0.97
	$\mathcal{T}_4$	613	0.01	0.02	0.04	1.24	0.65	0.94
		1000	0.01	0.02	0.05	0.62	0.76	1.07
		2000000	0.84	0.85	0.80	1.00	1.00	1.00

Table 2.13 (Continued). Relative MSE of OLS and two-stage estimators compared with 2SLS for $\beta = 0$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
$\mathcal{H}_1$	0	0.01	0.01	0.04	0.86	0.54	0.73
	13	0.01	0.01	0.03	0.42	0.52	0.73
	200	0.01	0.01	0.04	0.48	0.58	1.08
	613	0.01	0.02	0.04	1.26	0.70	1.03
	1000	0.01	0.02	0.05	0.67	0.84	1.21
	2000000	0.84	0.85	0.81	1.00	1.00	1.00
$\mathcal{H}_2$	0	0.01	0.01	0.04	0.86	0.54	0.73
	13	0.01	0.01	0.03	0.42	0.52	0.73
	200	0.01	0.01	0.04	0.48	0.58	1.03
	613	0.01	0.02	0.04	1.26	0.70	1.03
	1000	0.01	0.02	0.05	0.67	0.84	1.21
	2000000	0.84	0.85	0.80	1.00	1.00	1.00
$\mathcal{H}_3$	0	0.01	0.02	0.04	0.86	0.51	0.70
	13	0.01	0.02	0.03	0.40	0.50	0.70
	200	0.01	0.02	0.04	0.46	0.55	1.08
	613	0.01	0.02	0.04	1.24	0.65	0.94
	1000	0.01	0.02	0.05	0.62	0.76	1.07
	2000000	0.84	0.85	0.80	1.00	1.00	1.00
$\mathcal{RH}$	0	0.01	0.02	0.04	0.86	0.51	0.70
	13	0.01	0.02	0.03	0.40	0.50	0.70
	200	0.01	0.02	0.04	0.46	0.55	1.08
	613	0.01	0.02	0.04	1.24	0.65	0.93
	1000	0.01	0.02	0.05	0.62	0.77	1.05
	2000000	0.84	0.92	0.93	1.00	1.00	1.00

Table 2.14. Absolute bias of OLS, 2SLS and two-stage estimators for $\beta = 10$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
MCO		0	-3.75E-04	5.02E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
		13	9.05E-04	5.12E-02	1.00E-01	5.00E-01	6.00E-01	9.50E-01
		200	2.24E-04	4.97E-02	1.01E-01	5.00E-01	5.99E-01	9.49E-01
		613	-7.08E-05	4.92E-02	1.01E-01	4.98E-01	6.00E-01	9.47E-01
		1000	-3.13E-04	4.92E-02	9.94E-02	4.98E-01	5.98E-01	9.47E+03
		2000000	-1.51E-04	5.40E-03	1.12E-02	5.55E-02	6.64E-02	1.05E-01
2SLS		0	5.10E-03	6.02E-02	9.40E-02	4.99E-01	6.02E-01	9.55E-01
		13	2.60E-03	3.46E-02	9.87E-02	4.91E-01	6.03E-01	9.55E-01
		200	2.10E-03	4.16E-02	8.03E-02	4.56E-01	5.53E-01	8.71E-01
		613	-5.18E-04	4.33E-02	7.02E-02	3.96E-01	4.77E-01	7.48E-01
		1000	3.10E-03	3.61E-02	6.28E-02	3.47E-01	4.12E-01	6.70E-01
		2000000	-1.83E-04	-6.49E-05	2.55E-04	4.19E-04	4.89E-04	1.00E-03
two-stage	Pre-tests	0	-1.20E-04	5.08E-02	9.97E-02	4.99E-01	6.00E-01	9.50E-01
		13	9.87E-04	5.04E-02	9.99E-02	4.99E-01	6.00E-01	9.51E-01
		200	3.18E-04	4.93E-02	1.00E-01	4.97E-01	5.97E-01	9.45E-01
	$\mathcal{T}_1$	613	-9.42E-05	4.89E-02	9.95E-02	4.93E-01	5.93E-01	9.35E-01
		1000	-1.39E-04	4.85E-02	9.75E-02	4.90E-01	5.85E-01	9.26E-01
		2000000	-1.53E-04	4.70E-03	7.30E-03	4.19E-04	4.89E-04	1.00E-03
		0	-1.02E-04	5.07E-02	9.98E-02	4.99E-01	6.00E-01	9.50E-01
		13	9.86E-04	5.04E-02	9.99E-02	4.99E-01	6.00E-01	9.51E-01
		200	3.21E-04	4.93E-02	1.00E-01	4.97E-01	5.96E-01	9.44E-01
	$\mathcal{T}_2$	613	-9.49E-05	4.89E-02	9.94E-02	4.92E-01	5.92E-01	9.28E-01
		1000	-1.42E-04	4.85E-02	9.74E-02	4.87E-01	5.81E-01	9.10E-01
		2000000	-1.53E-04	4.40E-03	5.20E-03	4.19E-04	4.89E-04	1.00E-03
		0	-3.62E-04	5.03E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
		13	9.09E-04	5.12E-02	1.00E-01	5.00E-01	6.00E-01	9.51E-01
		200	2.31E-04	4.97E-02	1.01E-01	5.00E-01	5.99E-01	9.49E-01
	$\mathcal{T}_3$	613	-7.35E-05	4.92E-02	1.01E-01	4.97E-01	5.99E-01	9.44E-01
		1000	-2.90E-04	4.91E-02	9.92E-02	4.96E-01	5.95E-01	9.39E-01
		2000000	-1.53E-04	4.40E-03	5.30E-03	4.19E-04	4.89E-04	1.00E-03
		0	-1.02E-04	5.07E-02	9.98E-02	4.99E-01	6.00E-01	9.50E-01
		13	9.86E-04	5.04E-02	9.99E-02	4.99E-01	6.00E-01	9.51E-01
		200	3.20E-01	4.93E-02	1.00E-01	4.97E-01	5.96E-01	9.44E-01
	$\mathcal{T}_4$	613	-9.48E-05	4.89E-02	9.94E-02	4.92E-01	5.92E-01	9.28E-01
		1000	-1.42E-04	4.85E-02	9.74E-02	4.87E-01	5.82E-01	9.11E-01
		2000000	-1.53E-04	4.40E-03	5.20E-03	4.19E-04	4.89E-04	1.00E-03

Table 2.14 (Continued). Absolute bias of OLS, 2SLS and two-stage estimators for $\beta = 10$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	-3.62E-04	5.03E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
	13	9.08E-04	5.12E-02	1.00E-01	5.00E-01	6.00E-01	9.51E-01
	200	2.31E-04	4.97E-02	1.01E-01	5.00E-01	5.99E-01	9.49E-01
$\mathcal{H}_1$	613	-7.34E-05	4.92E-02	1.01E-01	**4.97E-01**	5.99E-01	9.45E-01
	1000	-2.90E-04	4.91E-02	9.92E-02	4.96E-01	5.95E-01	9.39E-01
	2000000	-1.53E-04	4.40E-03	5.20E-03	4.19E-04	4.89E-04	1.00E-03
	0	-3.62E-04	5.03E-02	1.00E-01	4.99E-01	6.00E-01	9.50E-01
	13	9.09E-01	5.12E-02	1.00E-01	5.00E-01	6.00E-01	9.51E-01
	200	2.31E-04	4.97E-02	1.01E-01	5.00E-01	5.99E-01	9.49E-01
$\mathcal{H}_2$	613	-7.35E-05	4.92E-02	1.01E-01	**4.97E-01**	5.99E-01	9.44E-01
	1000	-2.90E-04	4.91E-02	9.92E-02	4.96E-01	5.95E-01	9.39E-01
	2000000	-1.53E-04	4.40E-03	5.20E-03	4.19E-04	4.89E-04	1.00E-03
	0	-1.00E-04	5.07E-02	9.98E-02	4.99E-01	6.00E-01	9.50E-01
	13	9.86E-01	5.04E-02	9.99E-02	4.99E-01	6.00E-01	9.51E-01
	200	3.21E-04	4.93E-02	1.00E-01	4.97E-01	5.96E-01	9.44E-01
$\mathcal{H}_3$	613	-9.49E-05	4.89E-02	9.94E-02	4.92E-01	5.92E-01	9.28E-01
	1000	-1.41E-04	4.85E-02	9.74E-02	4.87E-01	5.81E-01	9.10E-01
	2000000	-1.53E-04	4.40E-03	5.20E-03	4.19E-04	4.89E-04	1.00E-03
	0	-1.21E-04	5.07E-02	9.97E-02	4.99E-01	6.00E-01	9.50E-01
	13	9.85E-04	5.04E-02	9.99E-02	4.99E-01	6.00E-01	9.51E-01
	200	3.17E-04	4.93E-02	1.00E-01	4.97E-01	5.96E-01	9.44E-01
$\mathcal{RH}$	613	-9.35E-05	4.89E-02	9.94E-02	4.92E-01	5.92E-01	9.27E-01
	1000	-1.48E-04	4.85E-02	9.75E-02	4.86E-01	5.82E-01	9.07E-01
	2000000	-1.53E-04	4.90E-03	7.70E-03	4.19E-04	4.89E-04	1.00E-03

Table 2.15. MSE of OLS, 2SLS and two-stage estimators for $\beta = 10$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
		0	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.62E-01	9.05E-01
		13	2.00E-03	4.60E-03	1.20E-02	2.52E-01	3.62E-01	9.05E-01
		200	2.00E-03	4.50E-03	1.22E-02	2.52E-01	3.61E-01	9.03E-01
MCO		613	2.00E-03	4.40E-03	1.22E-02	2.50E-01	3.61E-01	8.99E-01
		1000	2.00E-03	4.40E-02	1.18E-02	2.50E-01	3.59E-01	8.98E-01
		2000000	2.21E-04	2.45E-04	3.50E-04	3.30E-03	4.70E-03	1.15E-02
		0	3.26E-01	3.35E-01	3.31E-01	5.80E-01	7.03E-01	1.26E+00
		13	3.44E-01	3.37E-01	3.25E-01	5.78E-01	6.94E-01	1.25E+00
		200	3.13E-01	3.12E-01	3.24E-01	5.21E-01	6.18E-01	1.08E+00
2SLS		613	2.80E-01	2.66E-01	2.73E-01	4.35E-01	5.08E-01	8.59E-01
		1000	2.33E-01	2.32E-01	2.40E-01	3.67E-01	4.31E-01	7.52E-01
		2000000	2.48E-04	2.43E-04	2.54E-04	3.04E-04	3.45E-04	4.74E-04
	Pre-tests	0	2.60E-03	5.20E-03	1.23E-02	2.40E-01	3.45E-01	8.60E-01
two-stage		13	2.70E-03	5.20E-03	1.21E-02	2.40E-01	3.45E-01	8.65E-01
		200	2.60E-03	5.00E-03	1.23E-02	2.39E-01	3.43E-01	8.53E-01
	$\mathcal{T}_1$	613	2.70E-03	4.90E-03	1.21E-02	2.36E-01	3.40E-01	8.35E-01
		1000	2.50E-03	4.70E-02	1.17E-02	2.33E-01	3.30E-01	8.18E-01
		2000000	2.10E-04	2.13E-04	2.30E-04	3.04E-04	3.45E-04	4.74E-04
		0	2.70E-03	5.10E-03	1.22E-02	2.40E-01	3.45E-01	8.60E-01
		13	2.70E-03	5.20E-03	1.21E-02	2.40E-01	3.46E-01	8.63E-01
		200	2.70E-03	5.00E-03	1.24E-02	2.38E-01	3.41E-01	8.47E-01
	$\mathcal{T}_2$	613	2.70E-03	4.90E-03	1.22E-02	2.34E-01	3.37E-01	8.03E-01
		1000	2.50E-03	4.70E-02	1.17E-02	2.28E-01	3.22E-01	7.64E-01
		2000000	2.09E-04	2.03E-04	2.05E-04	3.04E-04	3.45E-04	4.74E-04
		0	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.61E-01	9.02E-01
		13	2.00E-03	4.60E-03	1.19E-02	2.51E-01	3.61E-01	9.03E-01
		200	2.00E-03	4.40E-03	1.22E-02	2.51E-01	3.59E-01	8.98E-01
	$\mathcal{T}_3$	613	2.00E-03	4.40E-03	1.21E-02	2.48E-01	3.58E-01	8.84E-01
		1000	2.00E-03	4.40E-03	1.18E-02	2.46E-01	3.52E-01	8.67E-01
		2000000	2.09E-04	2.03E-04	2.05E-04	3.04E-04	3.45E-04	4.74E-04
		0	2.70E-03	5.10E-03	1.22E-02	2.40E-01	3.45E-01	8.60E-01
		13	2.70E-03	5.20E-03	1.21E-02	2.40E-01	3.46E-01	8.63E-01
		200	2.70E-03	5.00E-03	1.24E-02	2.38E-01	3.41E-01	8.47E-01
	$\mathcal{T}_4$	613	2.70E-03	4.90E-03	1.22E-02	2.34E-01	3.37E-01	8.04E-01
		1000	2.50E-03	4.70E-02	1.17E-02	2.28E-01	3.22E-01	7.64E-01
		2000000	2.09E-04	2.03E-04	2.05E-04	3.04E-04	3.45E-04	4.74E-04

Table 2.15 (Continued). MSE of OLS, 2SLS and two-stage estimators for $\beta = 10$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.61E-01	9.02E-01
	13	2.00E-03	4.60E-03	1.19E-02	2.51E-01	3.61E-01	9.03E-01
	200	2.00E-03	4.40E-03	1.22E-02	2.51E-01	3.60E-01	8.99E-01
$\mathcal{H}_1$	613	2.00E-03	4.40E-03	1.22E-02	2.48E-01	3.58E-01	8.84E-01
	1000	2.00E-03	4.40E-03	1.18E-02	2.46E-01	3.52E-01	8.67E-01
	2000000	2.10E-04	2.04E-01	2.05E-04	3.04E-04	3.45E-04	4.74E-04
	0	2.00E-03	4.50E-03	1.20E-02	2.51E-01	3.61E-01	9.02E-01
	13	2.00E-03	4.60E-03	1.19E-02	2.51E-01	3.61E-01	9.03E-01
	200	2.00E-03	4.40E-03	1.22E-02	2.51E-01	3.59E-01	8.98E-01
$\mathcal{H}_2$	613	2.00E-03	4.40E-03	1.21E-02	2.48E-01	3.58E-01	8.84E-01
	1000	2.00E-03	4.40E-03	1.18E-02	2.46E-01	3.52E-01	8.66E-01
	2000000	2.09E-04	2.03E-04	2.05E-04	3.04E-04	3.45E-04	4.74E-04
	0	2.70E-03	5.10E-03	1.22E-02	2.40E-01	3.45E-01	8.60E-01
	13	2.70E-03	5.20E-03	1.21E-02	2.40E-01	3.46E-01	8.63E-01
	200	2.70E-03	5.00E-03	1.24E-02	2.38E-01	3.41E-01	8.47E-01
$\mathcal{H}_3$	613	2.70E-03	4.90E-03	1.21E-02	2.34E-01	3.37E-01	8.03E-01
	1000	2.50E-03	4.70E-02	1.17E-02	2.28E-01	3.22E-01	7.64E-01
	2000000	2.09E-04	2.03E-04	2.05E-04	3.04E-04	3.45E-04	4.74E-04
	0	2.60E-03	5.00E-03	1.22E-02	2.40E-01	3.46E-01	8.60E-01
	13	2.70E-03	5.20E-03	1.21E-02	2.41E-01	3.46E-01	8.63E-01
	200	2.60E-03	5.00E-03	1.24E-02	2.38E-01	3.41E-01	8.43E-01
$\mathcal{RH}$	613	2.60E-03	4.80E-03	1.21E-02	2.34E-01	3.36E-01	8.00E-01
	1000	2.40E-03	4.70E-02	1.17E-02	2.27E-01	3.23E-01	7.54E-01
	2000000	2.10E-04	2.19E-01	2.38E-04	3.04E-04	3.45E-04	4.74E-04

Table 2.16. Relative bias of OLS and two-stage estimators compared with 2SLS for $\beta = 10$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
MCO		0	-0.07	0.83	1.06	1.00	1.00	1.00
		13	0.35	1.48	1.01	1.02	0.99	1.00
		200	0.11	1.19	1.26	1.09	1.08	1.09
		613	0.14	1.14	1.44	1.26	1.26	1.27
		1000	-0.10	1.36	1.58	1.44	1.45	1.41
		2000000	0.83	-83.23	43.87	132.32	135.86	105.40
two-stage	Pre-tests	0	-0.02	0.84	1.06	1.00	1.00	1.00
		13	0.38	1.46	1.01	1.02	1.00	1.00
		200	0.15	1.19	1.25	1.09	1.08	1.08
	$\mathcal{T}_1$	613	0.18	1.13	1.42	1.24	1.24	1.25
		1000	-0.04	1.34	1.55	1.41	1.42	1.38
		2000000	0.83	-72.44	28.59	1.00	1.00	1.00
		0	-0.02	0.84	1.06	1.00	1.00	1.00
		13	0.38	1.46	1.01	1.02	1.00	1.00
		200	0.15	1.19	1.25	1.09	1.08	1.08
	$\mathcal{T}_2$	613	0.18	1.13	1.42	1.24	1.24	1.24
		1000	-0.05	1.34	1.55	1.40	1.41	1.36
		2000000	0.83	-67.81	20.37	1.00	1.00	1.00
		0	-0.07	0.84	1.06	1.00	1.00	1.00
		13	0.35	1.48	1.01	1.02	0.99	1.00
		200	0.11	1.19	1.26	1.09	1.08	1.09
	$\mathcal{T}_3$	613	0.14	1.14	1.44	1.25	1.25	1.26
		1000	-0.09	1.36	1.58	1.43	1.44	1.40
		2000000	0.83	-67.81	20.76	1.00	1.00	1.00
		0	-0.02	0.84	1.06	1.00	1.00	1.00
		13	0.38	1.46	1.01	1.02	1.00	1.00
		200	152.57	1.19	1.25	1.09	1.08	1.08
	$\mathcal{T}_4$	613	0.18	1.13	1.42	1.24	1.24	1.24
		1000	-0.05	1.34	1.55	1.40	1.41	1.36
		2000000	0.83	-67.81	20.37	1.00	1.00	1.00

Table 2.16 (Continued). Relative bias of OLS and two-stage estimators compared with 2SLS for $\beta = 10$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	-0.07	0.84	1.06	1.00	1.00	1.00
	13	0.35	1.48	1.01	1.02	0.99	1.00
	200	0.11	1.19	1.26	1.09	1.08	1.09
$\mathcal{H}_1$	613	0.14	1.14	1.44	1.25	1.25	1.26
	1000	-0.09	1.36	1.58	1.43	1.44	1.40
	2000000	0.83	-67.81	20.37	1.00	1.00	1.00
	0	-0.07	0.84	1.06	1.00	1.00	1.00
	13	349.42	1.48	1.01	1.02	0.99	1.00
	200	0.11	1.19	1.26	1.09	1.08	1.09
$\mathcal{H}_2$	613	0.14	1.14	1.44	1.25	1.25	1.26
	1000	-0.09	1.36	1.58	1.43	1.44	1.40
	2000000	0.83	-67.81	20.37	1.00	1.00	1.00
	0	-0.02	0.84	1.06	1.00	1.00	1.00
	13	379.38	1.46	1.01	1.02	1.00	1.00
	200	0.15	1.19	1.25	1.09	1.08	1.08
$\mathcal{H}_3$	613	0.18	1.13	1.42	1.24	1.24	1.24
	1000	-0.09	1.34	1.55	1.40	1.41	1.36
	2000000	0.83	-67.81	20.37	1.00	1.00	1.00
	0	-0.02	0.84	1.06	1.00	1.00	1.00
	13	0.38	1.46	1.01	1.02	1.00	1.00
	200	0.15	1.19	1.25	1.09	1.08	1.08
$\mathcal{RH}$	613	0.18	1.13	1.42	1.24	1.24	1.24
	1000	-0.05	1.34	1.55	1.40	1.41	1.36
	2000000	0.83	-75.52	30.16	1.00	1.00	1.00

Table 2.17. Relative MSE of OLS and two-stage estimators compared with 2SLS for $\beta = 10$

Estimators		$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
MCO		0	0.01	0.01	0.04	0.43	0.52	0.72
		13	0.01	0.01	0.04	0.44	0.52	0.73
		200	0.01	0.01	0.04	0.48	0.58	1.09
		613	0.01	0.02	0.04	1.26	0.71	1.05
		1000	0.01	0.19	0.05	0.68	0.83	1.19
		2000000	0.89	1.01	1.38	10.87	13.63	24.24
two-stage	Pre-tests	0	0.01	0.02	0.04	0.41	0.49	0.68
		13	0.01	0.02	0.04	0.41	0.50	0.69
		200	0.01	0.02	0.04	0.46	0.56	1.08
	$\mathcal{T}_1$	613	0.01	0.02	0.04	1.24	0.67	0.97
		1000	0.01	0.20	0.05	0.63	0.77	1.09
		2000000	0.85	0.88	0.91	1.00	1.00	1.00
		0	0.01	0.02	0.04	0.41	0.49	0.68
		13	0.01	0.02	0.04	0.42	0.50	0.69
		200	0.01	0.02	0.04	0.46	0.55	1.08
	$\mathcal{T}_2$	613	0.01	0.02	0.04	1.24	0.66	0.94
		1000	0.01	0.20	0.05	0.62	0.75	1.02
		2000000	0.84	0.84	0.81	1.00	1.00	1.00
		0	0.01	0.01	0.04	0.43	0.51	0.71
		13	0.01	0.01	0.04	0.43	0.52	0.72
		200	0.01	0.01	0.04	0.48	0.58	1.09
	$\mathcal{T}_3$	613	0.01	0.02	0.04	1.25	0.71	1.03
		1000	0.01	0.02	0.05	0.67	0.82	1.15
		2000000	0.84	0.84	0.81	1.00	1.00	1.00
		0	0.01	0.02	0.04	0.41	0.49	0.68
		13	0.01	0.02	0.04	0.42	0.50	0.69
		200	0.01	0.02	0.04	0.46	0.58	0.97
	$\mathcal{T}_4$	613	0.01	0.02	0.04	1.24	0.66	0.94
		1000	0.01	0.20	0.05	0.62	0.75	1.02
		2000000	0.84	0.84	0.81	1.00	1.00	1.00

Table 2.17 (Continued). Relative MSE of OLS and two-stage estimators compared with 2SLS for $\beta = 10$

Estimators	$\mu^2 \downarrow,\ \rho \rightarrow$	0	.05	.1	.5	.6	.95
	0	0.01	0.01	0.04	0.43	0.51	0.71
	13	0.01	0.01	0.04	0.43	0.52	0.72
	200	0.01	0.01	0.04	0.48	0.58	1.09
$\mathcal{H}_1$	613	0.01	0.02	0.04	1.25	0.71	1.03
	1000	0.01	0.02	0.05	0.67	0.82	1.15
	2000000	0.84	837.48	0.81	1.00	1.00	1.00
	0	0.01	0.01	0.04	0.43	0.51	0.71
	13	0.01	0.01	0.04	0.43	0.52	0.72
	200	0.01	0.01	0.04	0.48	0.58	1.03
$\mathcal{H}_2$	613	0.01	0.02	0.04	1.25	0.70	1.03
	1000	0.01	0.02	0.05	0.67	0.82	1.15
	2000000	0.84	0.84	0.81	1.00	1.00	1.00
	0	0.01	0.02	0.04	0.41	0.49	0.68
	13	0.01	0.02	0.04	0.42	0.50	0.69
	200	0.01	0.02	0.04	0.46	0.55	1.09
$\mathcal{H}_3$	613	0.01	0.02	0.04	1.24	0.66	0.94
	1000	0.01	0.20	0.05	0.62	0.75	1.02
	2000000	0.84	0.84	0.81	1.00	1.00	1.00
	0	0.01	0.01	0.04	0.41	0.49	0.68
	13	0.01	0.02	0.04	0.42	0.50	0.69
	200	0.01	0.02	0.04	0.46	0.55	1.08
$\mathcal{RH}$	613	0.01	0.02	0.04	1.24	0.66	0.93
	1000	0.01	0.20	0.05	0.62	0.75	1.00
	2000000	0.85	900.79	0.94	1.00	1.00	1.00

Chapter 3

Wald tests for error-regressors covariances, partial exogeneity tests and partial IV estimation

1. Introduction

An important problem in econometrics is testing whether a subset of stochastic explanatory variables is exogenous in a linear regression model. In many applied work, researchers often wants to assess the exogeneity of a subset of variables without imposing the exogeneity assumption on other model variables. For example, in the wage equation, one would like to test the partial exogeneity of mother's education without imposing restrictions on ability and schooling. In the New Keynesian Phillips Curve, one often needs to assess whether interest rate and unemployment rate are exogenous without restraining inflation rate and the other variables. Standard exogeneity tests of the type proposed by Durbin (1954), Wu (1973), Hausman (1978) (DWH) cannot solve such a problem. The difficulty of course is that covariance estimates and their standard errors are not typically produced from DWH-type tests. Consequently, these tests are not usable when testing the exogeneity of a subset of variables. As a result, testing linear restrictions on covariances using DWH-type tests is in general difficult and unpracticable.

The generalized Wald (GW)-type procedures proposed in Dufour (1987) alleviates such a difficulty. However, the Wald-type procedures assume that the errors are Gaussian. This raises the following question: how robust is the Wald-type procedures to error specification? In other words, are these procedures valid when the errors are non Gaussian?

This paper extends the Wald-type procedures to non Gaussian errors. We develop a new version of earlier tests which is typically valid (size is controlled) even for non normal errors. The version of test-statistic proposed differs from earlier statistic through the covariance matrix of the covariance estimate. We provide a correction of the covariance estimate which clearly depends on the excess kurtosis of the distribution of the errors. We show that when the errors are Gaussian, the excess kurtosis is zero and the modified covariance collapses to those in Dufour (1987).

We present the Monte Carlo experiment which confirms our theoretical results. In particular, when the instruments are strong and the errors have a chi square distribution, the earlier GW-test is size distorted with maximal rejection frequencies as greater as 99.14 % for the sample size $T = 300$. However, the modified test (MGW) is still valid (*i.e.*, level is controlled).

Moreover, we provide an analysis of the performance of different pretest-estimators based on Wald-type tests which allows us to develop two new pretest-estimators of structural parameter. Both estimators combine 2SLS and partial IV estimators. Two partial IV-estimators are used. The first estimator is obtained by treating the subset of supposedly endogenous regressors as exogenous, but is not included in the set of available instruments. The second estimator includes this set as additional instruments. The Monte Carlo simulations indicate that: (1) like OLS [Doko and Dufour (2009a)], partial IV-estimator outperform 2SLS when the instruments are weak; (2) pretest-estimators have a good performance (bias and MSE) over a wide range cases (including weak instruments and moderate endogeneity) compared with 2SLS. Therefore, this may be viewed as a variable selection procedure where a GW-test is used in the first stage to decide which variables should be instrumented and which ones are valid instruments.

The paper is organized as follows. Section 2 formulates the model considered. Section 3 presents the modified generalized Wald (MGW)-test and Section 4 studies through a Monte Carlo experiment: (1) the properties (level and power) of GW-type tests; (2) the performance (bias and MSE) of pretest-estimators–including OLS and 2SLS estimators. Section 5 illustrates our theoretical results through two empirical applications: the well known wage model and the returns to scale in electricity supply. We conclude in Section 9. Proofs are presented in the Appendix.

2. Framework

We consider the model describeb by

$$y = Y\beta + Z_1\gamma + u\,, \tag{2.1}$$

where y is a $T \times 1$ random vector; Y is a $T \times G$ matrix of explanatory variables $(G \geq 1)$, $u = [u_1, \ldots, u_T]'$ is a $T \times 1$ vector of disturbances, β and γ are $G \times 1$ and $k_1 \times 1$ vectors of unknown structural coefficients and Z_1 is a $T \times k_1$ matrix of included instruments.

We assume that

$$Y = Z\Pi + V\,, \tag{2.2}$$

where Z is $T \times k$ matrix of rank k, Π is a $k \times G$ matrix of unknown coefficients and V is a $T \times G$ matrix of disturbances. Let partition Z and Π as

$$Z = [Z_{11},\, Z_2] \quad \text{and} \quad \Pi = [\Pi_{11}',\, \Pi_2']'\,, \tag{2.3}$$

where Z_2 is a $T \times k_2$ matrix excluded from (2.1), Z_{11} is a set of variable included in Z_1, Π_2 and Π_{11} are respectively $k_2 \times G$ and $k_{11} \times G$ matrices of unknown coefficients. So, (2.2) can be written as

$$Y = Z_{11}\Pi_{11} + Z_2\Pi_2 + V\,. \tag{2.4}$$

We assume strong identification of model parameters, *i.e.*

$$T > 2G + k_1 \;\; \text{and} \;\; \operatorname{rank}(\Pi_2) = G\,. \tag{2.5}$$

Now, let also partition Y as

$$Y = [Y_1,\, Y_2]\,, \tag{2.6}$$

where Y_1 and Y_2 are respectively $T \times G_1$ and $T \times G_2$ matrices, $G_1 + G_2 = G.$ We assume by convention that if $G_1 = 0$ or $G_2 = 0,$ the corresponding variable Y_1 or Y_2 drops out of the model.

Define

$$\beta = [\beta_1',\, \beta_2']', \quad \Pi_{11} = [\Pi_{11}^1,\, \Pi_{11}^2],\;\; \Pi_2 = [\Pi_{21},\, \Pi_{22}], \quad \text{and} \;\; V = [V_1,\, V_2]\,, \tag{2.7}$$

where $\beta_1,\ \beta_2$ are respectively $G_1 \times 1$ and $G_2 \times 1$ vectors, Π_{11} and Π_2 are defined in (2.3), V_1 and

V_2 are reduced form errors.

From (2.6)-(2.7), model (2.1)-(2.2) can be written as

$$y = Y_1\beta_1 + Y_2\beta_2 + Z_1\gamma + u\,, \tag{2.8}$$

$$Y_1 = Z_{11}\Pi_{11}^1 + Z_2\Pi_{21} + V_1,$$

$$Y_2 = Z_{11}\Pi_{11}^2 + Z_2\Pi_{22} + V_2\,. \tag{2.9}$$

Let

$$V = [\omega_1,\, \ldots,\, \omega_G]\,, \tag{2.10}$$

where $\omega_k,\ 1 \le k \le G$ is the k-th column of V and

$$U = [u,\, V] = [U_1,\, \ldots\,,\, U_T]'\,. \tag{2.11}$$

We assume that

$$U_t \quad \text{are i.i.d for all } t = 1,\, \ldots,\, T \tag{2.12}$$

and

$$E[U_t] = 0, \quad E[U_tU_t'] = \Sigma = \begin{bmatrix} \sigma_u^2 & \delta' \\ \delta & \Sigma_V \end{bmatrix} > 0\,, \quad t = 1,\, \ldots\,,\, T, \tag{2.13}$$

$$E[U_{it}U_{jt}U_{kt}U_{lt}] < \infty,\ \forall\, t = 1,\, \ldots,\, T;\ \forall\, i, j, k, l = 1,\, \ldots\,,\, G+1\,, \tag{2.14}$$

where Σ_V has dimension G. Equations (2.13)-(2.14) assume finite fourth moments of model errors.

Let us define δ and Σ_V as

$$\delta = [\delta_1,\, \delta_2]', \quad \Sigma_V = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \equiv [\sigma_{jk}]_{j,\,k=1,\,\ldots,\,G}\,, \tag{2.15}$$

where δ and Σ_V are partitioned according to the partition of Y.

Let regress u on the columns of V, *i.e.*

$$u = Va + \varepsilon\,, \tag{2.16}$$

where a is a $G \times 1$ vector of unknown coefficients, and ε is independent of V with mean zero and variance σ_ε^2. From (2.13), we have

$$\delta = \Sigma_V a,\ \ \sigma_u^2 = \sigma_\varepsilon^2 + a'\Sigma_V a = \sigma_\varepsilon^2 + \delta'\Sigma_V^{-1}\delta\,. \tag{2.17}$$

We make the following generic assumptions on the asymptotic behaviour of model variables [where $A > 0$ for a matrix A means that A is positive definite (p.d.), and $\rightarrow$ refers to limits as $T \rightarrow \infty$]:

$$\frac{1}{T}\left[\begin{array}{cc} V & \varepsilon \end{array}\right]'\left[\begin{array}{cc} V & \varepsilon \end{array}\right] \xrightarrow{p} \left[\begin{array}{cc} \Sigma_V & 0' \\ 0 & \sigma_\varepsilon^2 \end{array}\right] > 0\,, \tag{2.18}$$

$$\frac{1}{T}\left[\begin{array}{cc} Z & Z_1 \end{array}\right]'\left[\begin{array}{cc} Z & Z_1 \end{array}\right] \xrightarrow{p} \Omega_Z = \left[\begin{array}{cc} \Sigma_Z & \Sigma_{ZZ_1}' \\ \Sigma_{ZZ_1} & \Sigma_{Z_1} \end{array}\right] > 0\,, \tag{2.19}$$

$$\frac{1}{T}\left[\begin{array}{cc} Z & Z_1 \end{array}\right]'\left[\begin{array}{cc} V & \varepsilon \end{array}\right] \xrightarrow{p} 0\,. \tag{2.20}$$

Now, substituting (2.16) in (2.1) yields

$$y = Y\beta + Z_1\gamma + Va + \varepsilon\,, \tag{2.21}$$

where ε is independent of all the regressors. If the matrix V were observed, we would test any set of linear restrictions on β, γ and a in equation (2.21) using standard Wald-type tests. In particular, linear hypotheses regarding the parameter vector a could be tested by using the least squares estimate $\hat{a}$ obtained from (2.21). Furthermore, if Σ_V were also known, the transformation $\delta = \Sigma_V a$ would allow one to test linear restriction on δ. Unfortunately, neither V nor Σ_V are known in practice and standard Wald-type tests are unusable in (2.21).

This paper has two main objectives. The first objective is to test linear restrictions on covariances, *i.e.* the hypothesis of the form

$$H_0 : Q\delta = d_0 \,, \tag{2.22}$$

where the errors U may follow any non Gaussian distribution. Q is an $r \times G$ fixed matrix of rank r and d_0 is a fixed $r \times 1$ vector. The second objective is related to pretesting practice where a GW-type test is used as pre-test in the first stage. Our main goal here is to analyze the relative performance –bias and mean square errors (MSE)– of pretest-estimators allowing for the presence of weak IV. However, the reader may note that the focus of this paper is not to provide formal theory on weak IV effects on Wald-type tests.

The following section derives the modified pre-test statistic (MGW).

3. Test statistics

Let us consider the equation

$$y = Y\beta + Z_1\gamma + \hat{V}a + \varepsilon^* = X\phi + \varepsilon^* \,, \tag{3.1}$$

where

$$X = [Y,\, Z_1,\, \hat{V}],\ \phi = (\beta',\, \gamma',\, a')',\ \hat{V} = Y - Z\hat{\Pi},\ \hat{\Pi} = (Z'Z)^{-1}Z'Y \,. \tag{3.2}$$

$\hat{\Pi}$ is the OLS estimate of Π obtained from equation (2.2) and $\hat{V}$ is the corresponding OLS residuals, and $\varepsilon^* = Z(\hat{\Pi} - \Pi)a + \varepsilon$.

The OLS estimate of ϕ obtained from (3.1) is given by

$$\hat{\phi} = (\hat{\beta}',\, \hat{\gamma}',\, \hat{a}')' = (X'X)^{-1}X'y \,. \tag{3.3}$$

Under the assumptions (2.18) - (2.20) we have

$$\Sigma_X = \underset{T\to\infty}{\text{plim}} \frac{X'X}{T} = \begin{bmatrix} \Pi'\Sigma_Z\Pi + \Sigma_V & \Pi'\Sigma_{ZZ_1} & \Sigma_V \\ \Sigma'_{ZZ_1}\Pi & \Sigma_{Z_1} & 0' \\ \Sigma_V & 0 & \Sigma_V \end{bmatrix}, \tag{3.4}$$

$$\Sigma_{ZX} = \underset{T\to\infty}{\text{plim}} \frac{Z'X}{T} = \begin{bmatrix} \Sigma_Z\Pi & \Sigma_{ZZ_1} & 0 \end{bmatrix}. \tag{3.5}$$

From (2.5), we have rank$(\Sigma_X) = 2G + k_1$, *i.e.* Σ_X is non-singular. Following Dufour (1987, Theorem 1), we show under the assumptions (2.1) - (2.2), (2.14) - (2.13), (2.5), and (6.3) - (6.5), that

(A) Consistency :

$$\underset{T\to\infty}{\text{plim}}\, \hat{\phi} = \phi, \quad \underset{T\to\infty}{\text{plim}}\, \hat{a} = a\,, \tag{3.6}$$

(B) Normality :

$$\sqrt{T}(\hat{\phi} - \phi) \overset{L}{\sim} N\,[0,\ \Sigma_\phi]\,, \quad \sqrt{T}(\hat{a} - a) \overset{L}{\sim} N\,[0,\ \Sigma_a]\,, \tag{3.7}$$

where

$$\Sigma_\phi = \sigma_\varepsilon^2\Sigma_X^{-1} + \tau\Sigma_X^{-1}\Sigma'_{ZX}\Sigma_Z^{-1}\Sigma_{ZX}\Sigma_X^{-1}\,, \tau = a'\Sigma_V a = \delta'\Sigma_V^{-1}\delta\,, \tag{3.8}$$

$$\Sigma_a = A_2[\sigma_\varepsilon^2\Sigma_X + \tau\Sigma'_{ZX}\Sigma_Z^{-1}\Sigma_{ZX}]A'_2, \quad A_2 = \underset{T\to\infty}{\text{plim}}\,(C_2) \tag{3.9}$$

and C_2 is a $G \times (2G + k_1)$ matrix such that:

$$\hat{\Sigma}_X^{-1} = \left(\frac{X'X}{T}\right)^{-1} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix},$$

[see proof in Appendix A]. Furthermore, let $\hat{\Sigma}_{XZ} = \left(\frac{X'Z}{T}\right)^{-1}$ and $\hat{\Sigma}_Z = \left(\frac{Z'Z}{T}\right)^{-1}$. Then, we have

$$\hat{\Sigma}_\phi = \hat{\Sigma}_X^{-1}[\hat{\sigma}_\varepsilon^2\hat{\Sigma}_X + \hat{\tau}\hat{\Sigma}_{XZ}\hat{\Sigma}_Z^{-1}\hat{\Sigma}'_{XZ}]\hat{\Sigma}_X^{-1}\,, \tag{3.10}$$

$$\hat{\Sigma}_a = C_2[\hat{\sigma}_\varepsilon^2 \hat{\Sigma}_X + \hat{\tau}\hat{\Sigma}_{XZ}\hat{\Sigma}_Z^{-1}\hat{\Sigma}'_{XZ}]C'_2\,, \tag{3.11}$$

$$\hat{\sigma}_\varepsilon^2 = (y - X\hat{\phi})'(y - X\hat{\phi})/T\,, \hat{\tau} = \hat{a}'\hat{\Sigma}_V\hat{a}\,, \hat{\Sigma}_V = \hat{V}'\hat{V}/T \tag{3.12}$$

and

$$\operatorname*{plim}_{T\to\infty} \hat{\Sigma}_\phi = \Sigma_\phi, \quad \operatorname*{plim}_{T\to\infty} \hat{\Sigma}_a = \Sigma_a, \quad \operatorname*{plim}_{T\to\infty} \hat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 \tag{3.13}$$

$$\operatorname*{plim}_{T\to\infty} \hat{\tau} = \tau, \quad \operatorname*{plim}_{T\to\infty} \hat{\Sigma}_V = \Sigma_V\,. \tag{3.14}$$

So, one can test any linear restrictions on ϕ, *i.e* the hypotheses of the form $H(M, m_0) : M\phi = m_0$, where M is a $\nu \times (2G + k_1)$ matrix of rank $\nu \leq 2G + k_1$ and m_0 is a $\nu \times 1$ fix vector, by using a critical region of the form $\{\mathcal{W}(M, m_0) \geq c\}$, where

$$\mathcal{W}(M, m_0) = T(M\hat{\phi} - m_0)'[M\hat{\Sigma}_\phi M']^{-1}(M\hat{\phi} - m_0)\,, \tag{3.15}$$

and c is a constant which depends on the level of the test [for more details, see Dufour (1987)].

Consider now the problem of testing linear restriction on covariances, *i.e* the hypothesis

$$H_0 : Q\delta = \delta_0\,. \tag{3.16}$$

To derive the test-statistic for this hypothesis, we shall exploit the following result in Anderson (2003, Theorem 3.6.2, p.102).

Define

$$\Sigma_V = [\sigma'_j]_{1\leq j\leq G}, \quad \hat{\Sigma}_V = [\hat{\sigma}'_j]_{1\leq j\leq G}\,, \tag{3.17}$$

where

$$\sigma'_j = (\sigma_{j1},\, \sigma_{j2},\, \ldots,\, \sigma_{jG}),\;\; \hat{\sigma}'_j = (\hat{\sigma}_{j1},\, \hat{\sigma}_{j2},\, \ldots,\, \hat{\sigma}_{jG}) \tag{3.18}$$

are the jth row of Σ_V and $\hat{\Sigma}_V$, $j = 1, \ldots, G$. Define

$$\sigma = (\sigma_1', \sigma_2', \ldots, \sigma_G')' = vec\Sigma_V, \ \hat{\sigma} = (\hat{\sigma}_1', \hat{\sigma}_2', \ldots, \hat{\sigma}_G')' = vec\hat{\Sigma}_V, \tag{3.19}$$

where for any matrix $B = [b_1, \ldots, b_G], \ vecB = vec[b_1, \ldots, b_G] = \begin{bmatrix} b_1 \\ \vdots \\ b_G \end{bmatrix}$. Then, we have

$$\sqrt{T}(\hat{\sigma} - \sigma) = \sqrt{T}(vec\hat{\Sigma}_V - vec\hat{\Sigma}_V) \overset{L}{\sim} N[0, \Sigma_\sigma], \tag{3.20}$$

where

$$\begin{aligned} \Sigma_\sigma &= (\kappa+1)(I_{G^2} + K_{GG})(\Sigma_V \otimes \Sigma_V) + \kappa vec\Sigma_V(vec\Sigma_V)' \equiv [\lambda_{jk}]_{1\le j, k\le G}, \\ \lambda_{jk} &= \sigma_{jk}\Sigma_V + \sigma_k\sigma_j' + \kappa(\sigma_{jk}\Sigma_V + \sigma_k\sigma_j' + \sigma_j\sigma_k') & (3.21) \\ \kappa &= \frac{1}{G(G+2)}E[V_t\Sigma_V^{-1}V_t']^2 - 1, \quad V_t = (\omega_{1t}, \omega_{2t}, \ldots, \omega_{Gt}). & (3.22) \end{aligned}$$

therefore, from Dufour (1987, Lemma 1), we have

$$\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)\mu \overset{L}{\sim} N[0, \Sigma_\mu], \tag{3.23}$$

where

$$\Sigma_\mu = (\mu'\Sigma_V\mu)\Sigma_V + (\Sigma_V\mu)(\Sigma_V\mu)' + \kappa[(\mu'\Sigma_V\mu)\Sigma_V + 2(\Sigma_V\mu)(\Sigma_V\mu)'], \tag{3.24}$$

μ is any fixed $G \times 1$ vector. Furthermore,

$$\hat{\kappa} = \frac{1}{G(G+2)}\frac{1}{T}\sum_{t=1}^{T}[\hat{V}_t'\hat{\Sigma}_V^{-1}\hat{V}_t]^2 - 1 \tag{3.25}$$

is a consistent estimator of κ. The parameter κ defined in (3.22) is the excess kurtosis of the distribution of V_t. Notice that (3.20) is identical to Dufour (1987, Lemma 1). The difference comes from the extra term in the expressions of the covariance matrices. The extra terms compared with the expressions in Dufour (1987, Lemma 1) are $\kappa(\sigma_{jk}\Sigma_V + \sigma_k\sigma_j' + \sigma_j\sigma_k')$ for λ_{jk} and

$\kappa[(\mu'\Sigma_V\mu)\Sigma_V + 2(\Sigma_V\mu)(\Sigma_V\mu)']$ for Σ_μ. In particular if the errors are gaussian, $\kappa = 0$ and we obtained the results in Dufour (1987).

We now use the above results to prove the following theorem on the distributions of $\hat{\delta} = \hat{\Sigma}_V\hat{a}$.

Theorem 3.1 *Under the assumptions* (2.1) - (2.2), (2.14) - (2.13), (2.5), *and* (6.3) - (6.5) :

(A) *Consistency* :

$$\operatorname*{plim}_{T\to\infty} \hat{\delta} = \delta\,, \tag{3.26}$$

(B) *Normality* :

$$\sqrt{T}(\hat{\delta} - \delta) \overset{L}{\sim} N\,[0,\,\Sigma_\delta]\ , \tag{3.27}$$

where

$$\begin{aligned}\Sigma_\delta &= \Sigma_V\Sigma_a\Sigma_V + \tau(1+\kappa)\Sigma_V + (1+2\kappa)\delta\delta' \\ &= \Sigma_V\Sigma_a\Sigma_V + \tau\Sigma_V + \delta\delta' + \kappa(\tau\Sigma_V + 2\delta\delta')\,. \end{aligned} \tag{3.28}$$

Furthermore, the covariance matrix Σ_δ can be estimated consistently by the statistic

$$\hat{\Sigma}_\delta = \hat{\Sigma}_V\hat{\Sigma}_a\hat{\Sigma}_V + \hat{\tau}(1+\hat{\kappa})\hat{\Sigma}_V + (1+2\hat{\kappa})\hat{\delta}\hat{\delta}'\,, \tag{3.29}$$

where $\hat{\tau}$, $\hat{\Sigma}_a$ and $\hat{\Sigma}_V$, and $\hat{\kappa}$ are defined above.

From Theorem **3.1**, one can test (3.16) by using a critical region of the form $\{\mathcal{W}_T \geq c\}$, where

$$\mathcal{W}_T = T(Q\hat{\delta} - \delta_0)'[Q\hat{\Sigma}_\delta Q']^{-1}(Q\hat{\delta} - \delta_0)\,, \tag{3.30}$$

and c depends on the level of the test. Under H_0, $\mathcal{W}_T$ is asymptotically distributed as a chi-square with r degrees of freedom. Thus, confidence regions for $Q\delta$ can be obtained by considering a complement of the critical region $\{\mathcal{W}_T \geq c\}$. In particular, confidence intervals for any subvector of δ, can be obtained. Remark that unlike DWH-procedures, the GW-procedure allows for the construction of confidence sets (CS's) for the covariances. Furthermore, the GW-procedure can be used to test partial exogeneity of subvectors in contrast of DWH specification tests.

Section 4 below analyzes the performance of pretest-estimators where a GM-type test is used in the first stage for the partial exogeneity of a subset of regressors.

4. Pretest-estimators

This section studies numerically the properties [bias and mean squares errors (MSE)] of the pretest-estimators where a GW-test is used in the first stage to assess whether a subset of explanatory variables is exogenous or not. Suppose that we want to estimate β_1 and β_2 in (2.8) - (2.9). If both Y_1 and Y_2 are endogenous, an IV method should be applied. However, if we have additional information that Y_1 is exogenous, estimating β_1 by OLS and applying IV method for β_2 alone would lead to more efficient estimators than applying (naively) an IV method for both parameters. This suggests that the practice of pretesting can help to know which variables will be instrumented and which ones are valid IV.

The use of the Wald-type procedures here has an advantage to allow one to test partial exogeneity, in contrast of Durbin-Wu-Hausman-type tests.

Now, consider the following setup. We want to assess whether Y_1 is exogenous, *i.e.* test the hypothesis

$$H_{\delta_1} : \delta_1 = 0, \, , \tag{4.1}$$

from a Wald-type test. If H_{δ_1} is rejected, 2SLS estimation are applied to both β_1 and β_2. However, if there is no evidence to reject H_{δ_1}, two possibilities are offered: (1) Y_1 is treat as exogenous in the estimation and only Y_2 is instrumented by the available IV; (2) Y_1 is included as additional IV in the set of instruments for Y_2. With these two possibilities in mind, we propose the following two pretest-estimators:

$$\hat{\beta} = \hat{\beta}_{PIV}\mathbb{1}_{(\mathcal{W}_{1T}\leq\chi^2_{1,1-\xi})} + \hat{\beta}_{2SLS}\mathbb{1}_{(\mathcal{W}_{1T}>\chi^2_{1,1-\xi})}\,, \tag{4.2}$$

$$\tilde{\beta} = \tilde{\beta}_{PIV}\mathbb{1}_{(\mathcal{W}_{1T}\leq\chi^2_{1,1-\xi})} + \hat{\beta}_{2SLS}\mathbb{1}_{(\mathcal{W}_{1T}>\chi^2_{1,1-\xi})}\,, \tag{4.3}$$

where $\mathcal{W}_{1T}$ is the GW-test for H_{δ_1}, ξ is the nominal size of the pre-test, $\mathbb{1}_{(.)}$ is the indicator function and $\chi^2_{1,1-\xi}$ is the $1-\xi$ quantile of a chi-square random variable with one degree of freedom. The estimators $\hat{\beta}_{2SLS}$, $\hat{\beta}_{PIV}$ and $\tilde{\beta}_{PIV}$ are defined as

$$\hat{\beta}_{2SLS} = (\hat{Y}'M_{Z_1}\hat{Y})^{-1}\hat{Y}'M_{Z_1}y,\ \hat{\beta}_{PIV} = [\hat{\beta}'_{1OLS},\ \hat{\beta}'_{2IV}]',\ \tilde{\beta}_{PIV} = [\hat{\beta}'_{1OLS},\ \tilde{\beta}'_{2IV}]' \tag{4.4}$$

$$\hat{\beta}_{1OLS} = (Y_1'M_{[Z_1,\,Y_2]}Y_1)^{-1}Y_1'M_{[Z_1,\,Y_2]}y,\ \hat{\beta}_{2IV} = (\hat{Y}_2'M_{[Z_1,\,Y_1]}\hat{Y}_2)^{-1}\hat{Y}_2'M_{[Z_1,\,Y_1]}y\,, \tag{4.5}$$

$$\tilde{\beta}_{2IV} = (\tilde{Y}_2'M_{[Z_1,\,Y_1]}\tilde{Y}_2)^{-1}\tilde{Y}_2'M_{[Z_1,\,Y_1]}y,\ \hat{Y} = [\hat{Y}_1,\,\hat{Y}_2] = P_ZY,\ \tilde{Y}_2 = P_WY_2\,, \tag{4.6}$$

$$W = [Z,\,Y_1],\ \hat{Y}^* = [Y_1,\,\hat{Y}_2],\ \tilde{Y} = [Y_1,\,\tilde{Y}_2]\,, \tag{4.7}$$

where for any matrix B with T rows, $P_B = B(B'B)^{-1}B'$ and $M_B = 1 - P_B$. Note that $\hat{\beta}_{2SLS}$ is the standard two-stage least squares estimator of β where both Y_1 and Y_2 are treated as endogenous. $\hat{\beta}_{PIV}$ is the partial IV-estimator of β where only Y_2 is instrumented and Y_1 treated as exogenous. $\tilde{\beta}_{PIV}$ differs from $\hat{\beta}_{PIV}$ in the use of Y_1. Y_1 is treated as additional set of instruments in the expression of $\tilde{\beta}_{PIV}$ while it is not in those of $\hat{\beta}_{PIV}$. However, the basic idea behind both estimators is the same, *i.e.* if Y_1 is exogenous, the two estimators should be more efficient than the standard 2SLS estimator.

4.1. Monte Carlo experiment

This subsection provides numerical comparison between GW-and MGW-tests in two directions: (1) the behaviour (level and power); (2) the performance (bias and MSE) of pretest-estimators based on these procedures.

4.1.1. Power of GW and MGW tests when the instruments are strong

We consider the following two endogenous simultaneous equations model:

$$\begin{aligned} y &= Y_1\beta_1 + Y_2\beta_2 + u, \\ (Y_1,\, Y_2) &= (Z_2\mathit{\Pi}_{21},\, Z_2\mathit{\Pi}_{22}) + (V_1,\, V_2), \\ u &= V_1a_1 + V_2a_2 + \varepsilon = V(\mathit{\Sigma}_V^{-1}\delta) + \varepsilon\,. \end{aligned} \tag{4.8}$$

where y: $T \times 1$, $Y = [Y_1,\, Y_2]$: $T \times G$ with $G = G_1 + G_2$, where Y_1: $T \times G_1$, Y_2: $T \times G_2$, $\delta = (\delta_1,\, \delta_2)'$, δ_1 is the covariance between u and V_1, and δ_2 those between u and V_2. Through all the experiment, $a_1 = a_2 = 1$. Two setups for the DGP are considered:

$$\text{DGP1}: \quad (\varepsilon_t, V_t)' \overset{i.i.d}{\sim} N\left[0, \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma_1^2 & \tau \\ 0 & \tau & \sigma_2^2 \end{pmatrix}\right], \quad \text{for all } t = 1,\, \ldots,\, T\ ; \tag{4.9}$$

$$\begin{aligned} \text{DGP2}: \quad \varepsilon_t &= \rho_\varepsilon(\varepsilon_t^* - \mu_\varepsilon), \quad V_{jt} = \rho_j(V_{jt}^* - \mu_j) \quad \text{where} \\ \varepsilon_t^*,\, V_{jt}^* &\overset{i.i.d}{\sim} \chi^2(\nu), \quad \text{with covariance matrix} \end{aligned}$$

$$\mathit{\Sigma} = \begin{bmatrix} 2\nu\rho_\varepsilon^2 & 0 & 0 \\ 0 & 2\nu\rho_1^2 & \tau \\ 0 & \tau & 2\nu\rho_2^2 \end{bmatrix}, \quad \text{for all } t = 1,\, \ldots,\, T \text{ and } j = 1,\, 2\,. \tag{4.10}$$

In both setups, $Z_{2t} \overset{i.i.d}{\sim} N(0,\, I_{k_2}),\ t = 1,\, \ldots,\, T,$ and $\sigma_1^2 = \sigma_2^2 = 1\ \tau = .44$, $\rho_1^2 = \rho_1^2 = \frac{1}{2r}$, $\nu = 1$, $\mu_\varepsilon = \mu_j = 1$. $\mathit{\Pi}_{21}$ and $\mathit{\Pi}_{22}$ are vectors of dimension k_2 defined as

$$\mathit{\Pi}_{21} = \eta_1 C_0, \quad \mathit{\Pi}_{22} = \eta_2 C_1\,, \tag{4.11}$$

where $[C_0,\, C_1]$ is a fixed $k_2 \times 2$ matrix obtained by taking the first two columns of the identity matrix with dimension k_2, $\eta_1 \in \{.5,\, 1,\, 5\}$ and $\eta_2 = 1$. Note that in this subsection, consider cases where both β_1 and β_2 are identified. The true value of $\beta = (\beta_1,\, \beta_2)' = (2,\, 5)'$ and the number of instruments k_2 belongs to $\{5,\ 20,\ 40\}$. The null hypothesis is

$$H_{\delta_1} : \delta_1 = \delta_{01} \, , \tag{4.12}$$

where $\delta_{01} \in \{0,\ 5\}$. Notice that $\delta_{01} = 0$ corresponds to partial exogeneity test for Y_1. In the DGP2, model errors are non Gaussian: chi square distribution with one degree of freedom. For each setup, we generate $N = 10,000$ replications of samples with size $T = 50,\ 100,\ 300$. The results for both setups are presented in Tables 3.1- 3.2. The first column of the tables contains the test-statistics and the second column the number of instruments k_2. The other columns contain the empirical rejection frequencies for each null hypothesis [$H_{\delta_1} : \delta_1 = 0$ and $H_{\delta_1} : \delta_1 = 5$] for each value of η_1 (quality of the IV for Y_1). We vary δ_1 in $\{-2,\ 0,\ 5\}$. We fix the quality of the IV for Y_2 at $\eta_2 = 1$.

We observe for the **DGP1** (Gaussian errors), that when testing the hypothesis $H_{\delta_1} : \delta_1 = 0$ both tests are valid (level is controlled) even in small sample. However, for the null hypothesis $H_{\delta_1} : \delta_1 = 5$, the earlier version [Dufour (1987, GW)] is still valid but the modified version (MGW) is slightly size distorted with maximal rejection as greater as 29.35 % for $T = 50$, $\eta_1 = 5$ and about 21.85 % for $T = 100$, $\eta_1 = 5$ [see Table 3.1]. However, if the errors are non Gaussian (see **DGP2**), the modified test is valid while the initial test is very size distorted when considering the null hypothesis $H_{\delta_1} : \delta_1 = 5$. The maximal rejection is as greater as 99.99 % for $T = 100$, $\eta_1 = 5$ and 99.14 % for $T = 300$, $\eta_1 = 5$ [see Table 3.2]. We also note that the test has power when identification is strong.

We now examine situation where the IV are weak.

Table 3.1. Level and Power of GW and MGW with strong IV: DGP1

$H_{\delta_1}: \delta_1 = 0, T = 50$										
	$\delta_1 \rightarrow$	-2			0			5		
Statistics	$k_2 \downarrow$ $\eta_1 \rightarrow$	.5	1	5	.5	1	5	.5	1	5
GW	2	22.36	91.59	99.51	0.78	2.48	3.19	69.64	99.93	100
MGW		27.94	94.1	99.73	1.31	3.87	6.67	73.15	99.97	100
GW	5	34.35	91.07	99.41	1.31	2.5	3.76	85.63	99.99	99.99
MGW		39.85	93.09	99.66	2.2	4.09	6.58	88.23	99.99	99.99
GW	20	24.85	84.94	99.18	3.37	3.36	3.19	78.34	99.75	99.99
MGW		29.15	87.84	99.37	4.71	4.5	5.07	81.48	99.86	99.99
$H_{\delta_1}: \delta_1 = 5, T = 50$										
GW	2	99.43	100	100	98.77	100	100	7.62	6.46	6.07
MGW		99.47	100	100	98.8	100	100	10.05	10.62	16.32
GW	5	99.38	100	100	98.22	100	100	25.96	10.32	6.24
MGW		99.42	100	100	98.31	100	100	29.54	15.96	19.12
GW	20	100	100	100	100	100	100	82.31	16.22	4.36
MGW		100	100	100	100	100	100	90.56	35.51	29.35
$H_{\delta_1}: \delta_1 = 0, T = 100$										
GW	2	93.56	99.96	100	2.42	3.58	4.19	99.71	100	100
MGW		94.74	99.98	100	3.12	5.05	7.44	99.73	100	100
GW	5	91.73	99.96	100	2.79	3.93	4.08	99.9	100	100
MGW		93.01	99.98	100	3.55	5.52	7.13	99.91	100	100
GW	20	73.89	99.85	100	3.65	4.09	4.1	99.79	100	100
MGW		76.52	99.92	100	4.59	5.5	6.48	99.83	100	100
$H_{\delta_1}: \delta_1 = 5, \quad T = 100$										
GW	2	99.91	100	100	99.76	100	100	6.24	5.23	5.31
MGW		99.91	100	100	99.76	100	100	7.75	8.77	13.31
GW	5	100	100	100	100	100	100	11.74	6.52	5.16
MGW		100	100	100	100	100	100	14.94	11.48	14.28
GW	20	100	100	100	100	100	100	79.96	18.54	4.35
MGW		100	100	100	100	100	100	85.32	33.12	21.85

Table 3.2. Level and Power of GW and MGW with strong IV: DGP2

$H_{\delta_1}: \delta_1 = 0,\ T = 50$										
	$\delta_1 \rightarrow$	-2			0			5		
Statistics	$k_2 \downarrow\ \eta_1 \rightarrow$	.5	1	5	.5	1	5	.5	1	5
GW	5	0.09	0.13	92.31	0.07	0.16	1.97	0.09	0.61	93.96
MGW		0.05	0.1	88.45	0.05	0.08	1.28	0.07	0.41	91.16
GW	20	1.99	3.22	99.95	2.16	1.96	4.21	1.89	5.03	100
MGW		3.31	4.91	99.97	3.36	3.22	6.08	3.3	7.49	100
GW	40	3.05	2.87	73.54	3.22	3.08	4.12	1.65	1.7	88.83
MGW		5.75	5.64	82.1	6.39	5.87	5.21	3.92	4.56	93.81
$H_{\delta_1}: \delta_1 = 5,\ T = 50$										
GW	5	98.12	97.94	99.73	98.27	98.37	99.86	97.48	94.69	41.29
MGW		97.91	97.66	99.52	98.02	98.1	99.82	97.13	93.63	2.18
GW	20	100	100	100	100	100	100	100	100	41.26
MGW		100	100	100	100	100	100	100	100	9.39
GW	40	100	100	100	99.99	100	100	100	100	6.85
MGW		100	100	100	99.99	100	100	100	100	6.1
$H_{\delta_1}: \delta_1 = 0,\ T = 100$										
GW	5	0.14	0.34	94.42	0.11	0.14	1.45	0.21	1.03	95.66
MGW		0.05	0.15	91.68	0.05	0.04	0.7	0.05	0.55	93.68
GW	20	2.42	4.85	100	2.26	2.68	3.79	2.61	9.5	100
MGW		1.97	4.16	100	1.95	2.34	3.48	2.14	8.12	99.99
GW	40	3.53	6.98	100	3.68	3.71	4.72	3.69	13.28	100
MGW		4.22	8.07	100	4.33	4.36	5.5	4.31	15.78	100
$H_{\delta_1}: \delta_1 = 5,\ T = 100$										
GW	5	98.27	97.77	99.8	98.04	98.31	99.93	97.43	94.98	57.08
MGW		98.1	97.63	99.74	97.88	98.11	99.93	97.11	94.06	3.64
GW	20	100	100	100	100	100	100	100	100	77.43
MGW		100	100	100	100	100	100	100	100	5.08
GW	40	100	100	100	100	100	100	100	100	74.47
MGW		100	100	100	100	100	100	100	100	4.97
$H_{\delta_1}: \delta_1 = 0,\ T = 300$										
GW	5	0.19	0.47	95.51	0.13	0.11	0.68	0.23	1.34	96.63
MGW		0.08	0.32	94.68	0.07	0.07	0.39	0.15	0.9	96.07
GW	20	2.94	5.87	100	2.64	2.68	3.83	3.18	13.32	100
MGW		2.29	4.76	100	2.05	2.16	2.44	2.47	10.62	100
GW	40	3.59	10.09	100	4.05	3.83	4.52	4.81	22.9	100
MGW		3.03	8.81	100	3.64	3.28	3.79	4.19	19.63	100
$H_{\delta_1}: \delta_1 = 5,\ T = 300$										
GW	5	98.27	98.05	99.83	98.07	98.5	99.94	97.38	94.74	72.57
MGW		98.19	97.98	99.81	97.97	98.41	99.94	97.23	94.4	6.56
GW	20	100	100	100	100	100	100	100	100	97.69
MGW		100	100	100	100	100	100	100	100	5.73
GW	40	100	100	100	100	100	100	100	100	99.14
MGW		100	100	100	100	100	100	100	100	6.01

4.1.2. Power of GW and MGW tests when the instruments are weak

In this subsection, we examine the properties of the tests when the IV are weak. The parameter which controls the quality of the IV for Y_2 is fixed at $\eta_2 = .01$ whereas those of Y_1 varies in { 0, .01, 5}. When $\eta = 5$, β_1 is identified but β_2 is still not identified. The main observations from Table 3.3-Table 3.4 are that: (1) both tests are valid if all IV are weak when testing partial exogeneity, *i.e*, the hypothesis of the form $H_{\delta_1} : \delta_1 = 0$ [similar to DWH-type tests]. However, the test are invalid for the general null hypothesis of the form $H_{\delta_1} : \delta_1 = 5$, even for Gaussian errors; (2) the tests have no power when all IV are weak [similar to DWH-type tests, see Doko and Dufour (2009a)]. This suggests that pretest-estimators based on Wald-type procedures should have a good performance (in term of MSE) compared with the usual IV estimator.

Table 3.3. Level and Power of GW and MGW with weak IV: DGP1

	$\delta_1 \rightarrow$	-2			0			5		
Statistics	$k_2 \downarrow$ $\eta_1 \rightarrow$	0	.01	5	0	.01	5	0	.01	5
$H_{\delta_1} : \delta_1 = 0,\ T = 50$										
GW	2	0	0	59.73	0.01	0	0.78	0	0	62.24
MGW		0	0	61.79	0.01	0.01	0.99	0.01	0.01	64.17
GW	5	0.06	0.07	96.93	0.11	0.07	2.02	0.08	0.11	97.05
MGW		0.13	0.16	97.41	0.17	0.15	2.79	0.13	0.18	97.39
GW	20	2.58	2.57	99.97	2.86	2.4	4.24	2.43	2.64	100
MGW		3.41	3.5	99.98	3.51	3.13	5.03	3.35	3.35	100
$H_{\delta_1} : \delta_1 = 5,\ T = 50$										
GW	2	45.89	45.69	91.29	45.73	45.87	93.76	46.48	44.63	3.12
MGW		46	45.76	91.78	45.8	45.93	94.01	46.53	44.72	7.83
GW	5	96.82	96.46	100	96.77	96.93	100	96.55	96.29	5.67
MGW		96.84	96.52	100	96.81	96.98	100	96.58	96.33	16.22
GW	20	100	100	100	100	100	100	100	100	5.16
MGW		100	100	99.99	100	100	100	100	100	32.47
$H_{\delta_1} : \delta_1 = 0,\ T = 100$										
GW	2	0.01	0	69.2	0	0	0.68	0.01	0.01	71.22
MGW		0.01	0	70.24	0	0	0.76	0.01	0.01	72.14
GW	5	0.04	0.05	98.69	0.08	0.06	2.28	0.04	0.1	99.03
MGW		0.07	0.07	98.8	0.09	0.07	2.65	0.06	0.14	99.1
GW	20	2.27	2.4	100	2.53	2.23	4.08	2.38	2.61	100
MGW		2.55	2.7	100	2.83	2.62	4.57	2.65	2.93	100
$H_{\delta_1} : \delta_1 = 5,\ T = 100$										
GW	2	36.52	36.64	94.22	35.81	36.1	93.75	35.68	35.68	3.17
MGW		36.59	36.73	94.48	35.84	36.16	94	35.72	35.73	7.68
GW	5	94.66	94.65	100	94.92	94.69	100	94.68	94.85	4.76
MGW		94.68	94.7	100	94.95	94.73	100	94.73	94.86	13.02
GW	20	100	100	100	100	100	100	100	100	5.4
MGW		100	100	100	100	100	100	100	100	23.9

Table 3.4. Level and Power of GW and MGW with weak IV: DGP2

	$\delta_1 \rightarrow$	-2			0			5		
Statistics	$k_2 \downarrow$ $\eta_1 \rightarrow$	0	.01	5	0	.01	5	0	.01	5
$H_{\delta_1}: \delta_1 = 0,\ T = 50$										
GW	5	0.06	0.05	92.31	0.11	0.05	1.97	0.05	0.05	93.96
MGW		0.05	0.02	88.45	0.1	0.05	1.28	0.03	0.05	91.16
GW	20	1.93	1.78	99.95	1.85	1.78	4.21	1.89	1.78	100
MGW		3.17	3.08	99.97	3.22	3.1	6.08	3.28	3.09	100
GW	40	2.82	2.92	73.54	3.01	2.96	4.12	3.06	2.76	88.83
MGW		5.71	5.71	82.1	5.71	5.43	5.21	5.74	5.47	93.81
$H_{\delta_1}: \delta_1 = 5,\ T = 50$										
GW	5	98.08	97.98	99.73	98.27	98.32	99.86	98.13	98.29	41.29
MGW		97.9	97.68	99.52	98.03	98.05	99.82	97.89	98.08	2.18
GW	20	100	100	100	100	100	100	100	100	41.26
MGW		100	100	100	100	100	100	100	100	9.39
GW	40	100	100	100	100	100	100	100	100	6.85
MGW		100	100	100	100	100	100	100	100	6.1
$H_{\delta_1}: \delta_1 = 0,\ T = 100$										
GW	5	0.1	0.14	94.42	0.12	0.06	1.45	0.1	0.1	95.66
MGW		0.03	0.08	91.68	0.06	0.04	0.7	0.03	0.05	93.68
GW	20	2.18	2.26	100	2.36	2.37	3.79	2.48	2.32	100
MGW		1.97	1.97	100	2.15	2.1	3.48	2.21	1.96	99.99
GW	40	3.57	3.68	100	3.28	3.35	4.72	3.34	3.43	100
MGW		4.02	4.19	100	3.95	3.83	5.5	3.75	3.96	100
$H_{\delta_1}: \delta_1 = 5,\ T = 100$										
GW	5	98.24	98.13	99.8	98	98.36	99.93	98.2	98.01	57.08
MGW		98.09	97.92	99.74	97.87	98.15	99.93	98	97.78	3.64
GW	20	100	100	100	100	100	100	100	100	77.43
MGW		100	100	100	100	100	100	100	100	5.08
GW	40	100	100	100	100	100	100	100	100	74.47
MGW		100	100	100	100	100	100	100	100	4.97
$H_{\delta_1}: \delta_1 = 0,\ T = 300$										
GW	5	0.15	0.13	95.51	0.07	0.12	0.68	0.17	0.09	96.63
MGW		0.06	0.08	94.68	0.02	0.07	0.39	0.11	0.08	96.07
GW	20	2.71	2.59	100	2.71	2.4	3.83	2.52	2.7	100
MGW		2.01	1.95	100	2.12	1.92	2.44	1.8	1.95	100
GW	40	3.53	3.47	100	3.97	3.83	4.52	3.39	4.02	100
MGW		3.16	2.97	100	3.53	3.52	3.79	2.98	3.57	100
$H_{\delta_1}: \delta_1 = 5,\ T = 300$										
GW	5	98.05	98.05	99.83	98.01	98.05	99.94	98.16	98.37	72.57
MGW		97.98	97.91	99.81	97.87	98.02	99.94	98.11	98.3	6.56
GW	20	100	100	100	100	100	100	100	100	97.69
MGW		100	100	100	100	100	100	100	100	5.73
GW	40	100	100	100	100	100	100	100	100	99.14
MGW		100	100	100	100	100	100	100	100	6.01

4.1.3. Performance of OLS, 2SLS and partial pretest-estimators

We now analyze the performance (bias and MSE) of the pretest-estimators defined in (4.2)-(4.3). As in the precious subsection, the true value of $\beta = (\beta_1,\, \beta_2)'$ is kept at $(2,\, 5)'$. The quality of the IV for Y_2, η_2 takes the value $.01$ (weak IV) and 1 (strong IV). The quality of the IV for Y_1, η_1 belongs to $\{0,\, .01,\, 5\}$. The covariance between Y_1 and u is such that $\delta_1 \in \{-2,\, 0,\, 1,\, 5,\, 10\}$ and the number of instruments is $k_2 = 20$. We maintained $a_1 = a_2 = 1$ through the exercise. Notice that the choice of $k_2 = 20$ insures us the existence of the bias and mean square errors (MSE) of IV estimators. The sample size is $T = 300$ and the number of replication is $N = 10,000$. The two pretest-estimators combine (1) $\hat{\beta}_{2SLS}$ and $\hat{\beta}_{PIV}$; (2) $\hat{\beta}_{2SLS}$ and $\tilde{\beta}_{PIV}$; depending on the outcome the pre-test in the first stage. All these estimators are defined above in Section 4. For each estimator, we compute the bias and MSE under the null hypothesis ($\delta_1 = 0$) and the alternative ($\delta_1 \in \{-2,\, 1,\, 5,\, 10\}$). Tables 3.6- 3.9 in Appendix C present the bias of different estimators while Tables 3.10 contains the MSE.

Our major findings can be summarized as follows: (1) OLS estimator and partial IV-estimators are more efficient than usual 2SLS when the instruments are weak (small MSE); (2) the pretest-estimators have an overall performance over a wide range of cases (including weak instruments and moderate endogeneity). This suggests that GW-test may be use to select which explanatory variables will be instrumented and which ones will served as valid instruments in order to improve model parameters estimation. Over all, our results suggests that pretesting must be conducted before the estimation.

5. Empirical illustration

In this section, we illustrate the behaviour of GW-type tests through two empirical applications related to important issues in economics: “Returns to scale in electricity supply” Nerlove (1963) and the widely studied problem of returns to education [Dufour and Taamouti (2006), Angrist and Krueger (1991), Angrist and Krueger (1995), Angrist and al. (1999), Mankiw and al. (1992)].

5.1. Returns to scale in electricity supply

Consider the following simplified equation costs

$$ln(TC_i) = \alpha_0 + \beta_1 ln(Q_i) + \beta_2 ln(PF_i) + u_i, \tag{5.1}$$

where TC_i is total costs for firm i, Q_i is output in millions of dollars, PF_i is the price of fuels; α_0, β_1, β_2, are unknown coefficients to be estimated. In this model, β_1 is the returns to scale. Since the firm’s output is supplied on demand, output depends on the the price of electricity. If the price is set to cover the average cost, then the firm’s efficiency affects output through the effect of the electricity price on demand and output in this case is endogenous. So, the price of electricity must be a good determinant (instrument) of output. This paper assesses whether the price of electricity PK is a valid IV for output from GW-and MGW-tests or not. The reduced form of the model is formulated as

$$ln(Q_i) = \gamma_{10} + \gamma_{11} ln(PK_i) + \gamma_{21} ln(PL_i) + V_{1i}, \tag{5.2}$$

$$ln(PF_i) = \gamma_{20} + \gamma_{12} ln(PK_i) + \gamma_{22} ln(PL_i) + V_{2i}, \tag{5.3}$$

where PL_i is the price of labor, PK_i is the price of capital. The data used are from the Nerlove (1963) paper "Returns to scale in electricity supply", and consist of 145 firms.

First, our results suggest that the price of fuel (PF) is exogenous. Indeed, the P-values of GW- and MGW-tests for the partial exogeneity of PF in (5.1)-(5.3) [$H_0 : cov(PF_i, u_i) = 0$] are respectively .8941 and .9056. Which means that PF constitutes valid IV for output in this model. So, estimate β_1 in (5.1) by OLS, as done in Nerlove (1963), is misleading and inaccurate. In this application, the OLS estimate of returns to scale is .7242 while the 2SLS estimate is .9868. Thus, the variation between OLS and 2SLS and is about 36.26 %, this is large enough.

Second, when using the price of fuel as an additional IV and consider the hypothesis of the form

$$H_0 : cov(Q_i, u_i) = cov(V_{1i}, u_i) = \delta_{10} \,, \tag{5.4}$$

the sample value of GW is 22.83 with a P-value equal to .00000 and those of GMW is 15.51 with a P-value equal to .00008 when $\delta_{10} = 0$, *i.e* when testing the exogeneity of the output. Therefore, the output is exogenous. However, when $\delta_{10} \neq 0$, we find for example that $P - value(GW) \leq .05$ for $\delta_{10} \leq .9474$ and $P - value(MGW) \leq .05$ for $\delta_{10} \leq .8069$. Thus, when the true value of δ_{10} links to $].8069,\ .9474]$, we have $P - value(GW) \leq .05$ while $P - value(MGW) > .05$. In particular, for $\delta_{10} = .9474$, $P - value(GW) = .05$ whereas $P - value(MGW) = .1063$, which is 2 times more than the P-value of the GW-test. It worthwhile to note that the estimation of δ_1 in the above model is $\hat{\delta}_1 = 1.606$.

5.2. Wage equation

Consider the standard wage equation:

$$w_i = \beta_1 S_i + \beta_2 A_i + Z_{1i}\gamma + u_i, \tag{5.5}$$

where w is log of hourly wage, S is years of schooling, A represents ability (IQ), Z_1 is a set of included instruments. Z_1 contains 11 variables such as experience (EXPR), a dummy for residency in the southern states (RNS), tenure in year (TEN), a dummy for residency in metropolitan areas (SMSA) and seven dummies (YDUM) for YEAR $= 66, \ldots 73$. In this model, even conditionally on ability, schooling may be endogenous: that is correlated with u. More precisely, an increase in ambition may increase both schooling S_i and error u_i, leading to positive correlation between these two variables. So, we need instruments for both A_i and S_i. The common practice consists on taking mother education in year (MED), a dummy for marital status (MRT), the score on the “Knowledge of the World of Work ”test (KWW), Age (AGE) and Age square (AGE2) as a set of excluded instruments and using DWH-tests to assess whether both ability and schooling are exogenous or not. However, there are at two problems for which mother’s education (MED) which is viewed in practice as an important determinant of schooling may also be endogenous. First, MED will also reflect the mother’s ambition, which she may pass on to her children through socialization or genetically, so that MED$_i$ and u_i may be correlated. Second, MED may actually belong to (5.5). In fact, better educated mothers may create better employment opportunities for their children through good networks for example. So, use mother’s education as an instrument is questionable. This suggests that one should first test the exogeneity (partial exogeneity) of this variable before estimating model parameters.

This application reformulates the model by taking mother’s education as possibly endogenous and uses GW-and MGW-tests to assess whether it is exogenous or not.

The model is

$$\begin{aligned} w_i &= \beta_1 S_i + \beta_2 A_i + \beta_3 MED_i + Z_{1i}\gamma + u_i \\ Y_i &= Z_{1i}\phi_1 + Z_{2i}\phi_2 + V_i\,; \end{aligned} \tag{5.6}$$

where Z_{2i} is the above set of excluded except for MED. The data set consists of the 758 sample of Griliches 1980 NLS-Y Data. Table 3.5 below summarize the results.

The first part of the Table 3.5 presents the results of the tests for the joint exogeneity hypothesis of ability, schooling and mother’s education as well as the partial exogeneity of each variable. The second part presents the results when treating mother’s education as additional IV.

Our results indicate that when mother’s education is considered as possibly endogenous, the joint exogeneity hypothesis of ability, schooling and mother’s education as soon as individual ex-

Table 3.5. GW-tests in wage equation model

GW-and MGW-tests, MED is not used as an IV					
Satatistics ↓ Null hypotheses →	$cov(MED_i, u_i) = 0$	$cov(S_i, u_i) = 0$	$cov(A_i, u_i) = 0$	$cov(A_i, S_i, u_i) = 0$	$cov(A_i, S_i, MED_i, u_i) = 0$
sample value of GW	0.3845	0.3154	0.7489	0.7866	0.8528
P-value of GW	0.5352	0.5744	0.3868	0.6748	0.8368
sample value of MGW	0.3832	0.3127	0.5421	0.6388	0.6869
P-value of MGW	0.5359	0.5760	0.7626	0.7266	0.8763
GW-and MGW-tests, MED is used as an IV					
	–	$cov(S_i, u_i) = 0$	$cov(A_i, u_i) = 0$	$cov(A_i, S_i, u_i) = 0$	–
sample value of GW	–	1.7136	11.7133	14.5579	–
P-value of GW	–	0.1905	0.0006	0.0007	–
sample value of MGW	–	1.6522	9.9816	12.0499	–
P-value of MGW	–	0.1987	0.0016	0.0024	–

ogeneity hypotheses cannot be rejected even at level 30 %. This is probably the effect of weak instruments in this model. In fact in presence of weak instruments, GW-and MGW-tests have a low power and may fail to detect a possible endogeneity in the model. Now, when treating mother's education as additional instrument (second part of Table 3.5), we find that the joint exogeneity hypothesis of both ability and schooling is rejected even at level 1 % : the P-value is .0007 for GW-test and .0024 for the modified version (MGW). This suggests that at least one of both variables is endogenous. Testing partial exogeneity of both ability and schooling leads to the conclusion that ability is endogenous but there is no evidence to treat schooling as endogenous. Our results reinforce the question of the quality of the instruments for schooling as questioned by Bound (1995).

6. Conclusions

In this paper, we focus on linear simultaneous equations model and propose an extension of the generalized linear Wald partial exogeneity tests [see Dufour (1987)] to non Gaussian errors. We propose a modified test which is valid even if the errors are non Gaussian errors. We present simulation evidence indicating that the modified test (MGW) performs better than the initial test (GW).

Moreover, we provide new pretest-estimators based on GW-type tests which have an excellent overall performance compared with usual IV estimators. The results can then apply to select variables which should be instrumented and which are valid instruments (exogenous) in an IV regression.

We illustrate our theoretical results through two empirical applications: the wage equation and the returns to scale in electricity supply. The results indicate that the GW-tests cannot reject the exogeneity of mother's education, *i.e.* mother's education may constitute a valid IV. However, the output in cost equation is endogenous and the price of fuel is a valid IV for estimating the returns to scale.

APPENDIX

A. Useful results

(A) Consistency : From (3.1) - (3.3), we have

$$\hat{\phi} - \phi = \hat{\Sigma}_X^{-1}\left(\frac{X'\varepsilon^*}{T}\right), \tag{A.1}$$

where $\hat{\Sigma}_X \xrightarrow{p} \Sigma_X$ and $\frac{X'\varepsilon^*}{T} = \hat{\Sigma}_{XZ}(\hat{\Pi} - \Pi)a + \frac{X'\varepsilon}{T} \xrightarrow{p} 0$ because $\hat{\Sigma}_{XZ} \xrightarrow{p} \Sigma_{XZ}$, $\hat{\Pi} - \Pi \xrightarrow{p} 0$, and $\frac{X'\varepsilon}{T} \xrightarrow{p} 0$. So, we have $\hat{\phi} - \phi \xrightarrow{p} 0$ and $\hat{\phi}$ is a consistent estimator of ϕ. This implies that $\hat{a} - a \xrightarrow{p} 0$, *i.e.* $\hat{a}$ is a consistent estimator of a.

(B) Normality : we can write

$$\sqrt{T}(\hat{\phi} - \phi) = \hat{\Sigma}_X)^{-1}\varepsilon_T, \tag{A.2}$$

$$\varepsilon_T = \frac{1}{\sqrt{T}}X'\varepsilon + \hat{\Sigma}_{XZ}\sqrt{T}(\hat{\Pi} - \Pi)a \tag{A.3}$$

where $\sqrt{T}(\hat{\Pi} - \bar{\Pi})a = \hat{\Sigma}_Z^{-1}\frac{Z'Va}{\sqrt{T}}$. We know that $\hat{\Sigma}_X \xrightarrow{p} \Sigma_X$, however, to find the distribution of ε_T, we will first find the joint distribution of $\frac{1}{\sqrt{T}}X'\varepsilon$ and $\hat{\Sigma}_{XZ}\sqrt{T}(\hat{\bar{\Pi}} - \bar{\Pi})a$. To find the limiting distribution of $[\frac{1}{\sqrt{T}}X'\varepsilon, \hat{\Sigma}_{XZ}\sqrt{T}(\hat{\Pi} - \Pi)a]$, we will consider first the limiting distribution of $[\frac{1}{\sqrt{T}}X'\varepsilon, \frac{1}{\sqrt{T}}Z'Va]$. Let

$$\vartheta = \frac{1}{\sqrt{T}}(L_1'X'\varepsilon + L_2'Z'Va) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\vartheta_t\,,$$

$$L_1' = [l_{11},\, l_{12},\, \ldots,\, l_{1K}]\,, L_2' = [l_{21},\, l_{22},\, \ldots,\, l_{2\bar{k}}]\,, \tag{A.4}$$

where L_1 and L_2 are $1 \times K$ and $1 \times \bar{k}$ vectors of arbitrary constants; $K = 2G + k_{11}$, $\bar{k} = k_{11} + k_2$

and $\vartheta_t = L_1' X_t' \epsilon_t + L_2' Z_t' V_t a$. So, we have

$$E[\vartheta_t] = 0, \quad E[\vartheta_t^2] = \sigma_\varepsilon^2 L_1' E(X_t' X_t) L_1 + \tau L_2' (Z_t' Z_t) L_2 = \sigma_\varepsilon^2 L_1' Q_t L_1 + \tau L_2' (Z_t' Z_t) L_2 \quad \text{(A.5)}$$

where $\tau = a' \Sigma_V a$ and

$$Q_t = \begin{bmatrix} \Pi'(Z_t' Z_t)\Pi + \Sigma_V & \Pi' D_{12t} & \Sigma_V \\ D_{12t}' \Pi & (Z_{1t}' Z_{1t}) & 0' \\ \Sigma_V & 0 & \Sigma_V \end{bmatrix}, \quad D_{12t} = [Z_{1t}' Z_{11t}, Z_{1t}' Z_{2t}]' . \quad \text{(A.6)}$$

Thus, we have from (A.4)

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} E(\vartheta_t) = 0, \text{ and } \operatorname*{plim}_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} Var(\vartheta_t) = \sigma_\varepsilon^2 L_1' \Sigma_X L_1 + \tau L_2' \Sigma_Z L_2 . \quad \text{(A.7)}$$

Under the assumption (2.14)-(2.13), (6.3)-(6.5), we can show [see Basmann (1960)] that $E(|\vartheta_t|^{2+\tau}) < \infty$ for some $\tau > 0$ for all t and all L_1, L_2. So, the Lindeberg condition is satisfied. Therefore, ϑ is asymptotically distributed as normal with mean zero and variance $\sigma_\varepsilon^2 L_1' \Sigma_X L_1 + \tau L_2' \Sigma_Z L_2$. Therefore, by the multivariate central limit theorem, we have

$$\left[\frac{1}{\sqrt{T}} X' \varepsilon, \frac{1}{\sqrt{T}} Z' V a\right] \xrightarrow{L} N(0, \Omega), \quad \Omega = \begin{bmatrix} \sigma_\varepsilon^2 \Sigma_X & 0 \\ 0 & \tau \Sigma_Z \end{bmatrix} . \quad \text{(A.8)}$$

So, we have

$$\left[\frac{1}{\sqrt{T}} X' \epsilon, \hat{\Sigma}_{XZ} \sqrt{T} (\hat{\hat{\Pi}} - \bar{\Pi}) a\right] \xrightarrow{L} N(0, \Omega^*) , \quad \text{(A.9)}$$

where

$$\Omega^* = \begin{bmatrix} I_K & 0 \\ 0 & \Sigma_{ZX}' \Sigma_Z^{-1} \end{bmatrix} \Omega \begin{bmatrix} I_K & 0 \\ 0 & Q_Z^{-1} Q_{ZX} \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2 \Sigma_X & 0 \\ 0 & \tau \Sigma_{ZX}' \Sigma_Z^{-1} \Sigma_{ZX} \end{bmatrix} . \quad \text{(A.10)}$$

We conclude that $\frac{1}{\sqrt{T}} X' \varepsilon$ and $\hat{\Sigma}_{XZ} \sqrt{T} (\hat{\hat{\Pi}} - \bar{\Pi}) a$ are asymptotically independent and that $\frac{1}{\sqrt{T}} X' \varepsilon \xrightarrow{L} N(0, \sigma_\varepsilon^2 \Sigma_X)$ and $\hat{\Sigma}_{XZ} \sqrt{T} (\hat{\Pi} - \Pi) a \xrightarrow{L} N(0, \tau \Sigma_{ZX}' \Sigma_Z^{-1} \Sigma_{ZX})$. So,

$\varepsilon_T = \frac{1}{\sqrt{T}}X'\varepsilon + \hat{\Sigma}_{XZ}\sqrt{T}(\hat{\Pi} - \Pi)a \xrightarrow{L} N(0, \sigma_\varepsilon^2\Sigma_X + \tau\Sigma'_{ZX}\Sigma_Z^{-1}\Sigma_{ZX})$. Thus,we have

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{L} N(0, \Sigma_\phi)$$

where Σ_ϕ is given in the theorem.

The asymptotic distribution of $\sqrt{T}(\hat{a} - a)$ follows from the identity

$$\sqrt{T}(\hat{a} - a) = C_2(\frac{1}{\sqrt{T}}X'\varepsilon^*)$$

and by the fact that $\varepsilon_T \overset{L}{\sim} N(0, \sigma_\varepsilon^2\Sigma_X + \tau\Sigma'_{ZX}\Sigma_Z^{-1}\Sigma_{ZX})$ and $\underset{T\to\infty}{\text{plim}}\,(C_2) = A_2$.

B. Froofs

PROOF OF THEOREM **3.1**

(A) Consistency : first, we can write

$$(\hat{\delta} - \delta) = \hat{\Sigma}_V(\hat{a} - a) + (\hat{\Sigma}_V - \Sigma_V)a\,. \tag{B.1}$$

Second, we know that $\hat{\Sigma}_V \xrightarrow{p} \Sigma_V$, $\hat{a} \xrightarrow{p} a$. Hence, $\hat{\delta} - \delta \xrightarrow{p} 0$.

(B) Normality : from (B.1), we have

$$\sqrt{T}(\hat{\delta} - \delta) = \sqrt{T}\hat{\Sigma}_V(\hat{a} - a) + \sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a\,. \tag{B.2}$$

From (3.1) - (3.3), we have

$$\begin{aligned} \sqrt{T}(\hat{\delta} - \delta) &= \hat{\Sigma}_V C_2\left[\frac{1}{\sqrt{T}}X'\varepsilon + \hat{\Sigma}_{XZ}\sqrt{T}(\hat{\Pi} - \Pi)a\right] + \sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a && \text{(B.3)} \\ &= \hat{\Sigma}_V C_2\varepsilon_T + \sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a\,, && \text{(B.4)} \end{aligned}$$

where ε_T is defined in (A.3). Since $\underset{T\to\infty}{\text{plim}}\,(\hat{\Sigma}_V C_2) = \Sigma_V A_2$ and $\varepsilon_T \xrightarrow{L} N(0, \sigma_\varepsilon^2\Sigma_X + \tau\Sigma'_{ZX}\Sigma_Z^{-1}\Sigma_{ZX})$, hence the first term of (B.4) is such that

$$\hat{\Sigma}_V C_2\varepsilon_T \quad \xrightarrow{L} \quad N[0, \Sigma_V A_2(\sigma_\varepsilon^2\Sigma_X + \tau\Sigma'_{ZX}\Sigma_Z^{-1}\Sigma_{ZX})A'_2\Sigma_V]$$

$$\equiv \quad N(0, \Sigma_V \Sigma_a \Sigma_V). \tag{B.5}$$

We have also showed that

$$\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a \quad \xrightarrow{L} \quad N(0, \Omega_a)\,, \tag{B.6}$$

where

$$\begin{aligned} \Omega_a &= (1+\kappa)(a'\Sigma_V a)\Sigma_V + (1+2\kappa)(\Sigma_V a)(\Sigma_V a)' \\ &= \tau\Sigma_V + \delta\delta' + \kappa(\tau\Sigma_V + 2\delta\delta')\,. \end{aligned} \tag{B.7}$$

We will now show that $\hat{\Sigma}_V C_2 \varepsilon_T$ and $\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a$ are asymptotically independent.

Since $\underset{T\to\infty}{\text{plim}}\,(\hat{\Sigma}_V C_2) = \Sigma_V A_2$ and $\underset{T\to\infty}{\text{plim}}\,\hat{\Sigma}_{XZ} = \Sigma_{XZ}$, this is equivalent to show that ε_T and $\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a$ are asymptotically independent. Moreover, we have $\varepsilon_T = \frac{1}{\sqrt{T}}X'\varepsilon + \hat{\Sigma}_{XZ}\sqrt{T}(\hat{\Pi} - \Pi)a$ and we have already showed that $\frac{1}{\sqrt{T}}X'\varepsilon$ and $\hat{\Sigma}_{XZ}\sqrt{T}(\hat{\Pi} - \Pi)a$ are independent. So, to establish the result, it is sufficient to show that $\sqrt{T}(\hat{\Pi} - \Pi)$ and $\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)$ are asymptotically independent. Obviously, $\sqrt{T}(\hat{\Pi} - \Pi)$ and $\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)$ are independent. To check this, let $\hat{\Pi}_k$ and $\hat{\omega}_l$ be respectively the kth and lth columns of $\hat{\Pi}$ and $\hat{\Sigma}_V$. Then, we have

$$E[(\hat{\Pi}_k - \Pi_k)(\hat{\omega}_l - \sigma_l^*)'] = E[(\hat{\Pi}_k - \Pi_k)\hat{\omega}_l'] - E[(\hat{\Pi}_k - \Pi_k)\sigma_l^{*\prime}] = 0,\ \forall\, k, l = 1,\, \ldots,\, G\,, \tag{B.8}$$

because $E[(\hat{\Pi}_k - \Pi_k)\hat{\omega}_l'] = 0$ and $E[(\hat{\Pi}_k - \Pi_k)\sigma_l^{*\prime}] = [E(\hat{\Pi}_k - \Pi_k)]\sigma_l^{*\prime} = 0$, σ_l^* is the lth column of Σ_V. Which means that $\sqrt{T}(\hat{\Pi} - \Pi)a$ and $\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a$ are asymptotically uncorrelated and normally distributed, *i.e.*, $\hat{\Sigma}_V C_2 \varepsilon_T$ and $\sqrt{T}(\hat{\Sigma}_V - \Sigma_V)a$ are asymptotically uncorrelated and normally distributed, thus asymptotically independent. Consequently,

$$\sqrt{T}(\hat{\delta} - \delta) \xrightarrow{L} N(0, \Sigma_\delta)\,, \tag{B.9}$$

where

$$\begin{aligned} \Sigma_\delta &= \Sigma_V \Sigma_a \Sigma_V + \Omega_a \\ &= \Sigma_V \Sigma_a \Sigma_V + \tau \Sigma_V + \delta\delta' + \kappa(\tau \Sigma_V + 2\kappa\delta\delta') \,. \end{aligned} \tag{B.10}$$

Q.E.D.

□

C. Bias and MSE of OLS, 2SLS and partial IV estimators

Table 3.6. DGP1: Bias of OLS, 2SLS, Partial IV and Pretest-estimators: strong IV for Y_2 ($\eta_2 = 1$)

$H_{\delta_1}: \delta_1 = 0$, strong IV for Y_2													
	$\delta_1 \rightarrow$	-2			0			1			5		
Estimators	$Bias \downarrow$ $\eta_1 \rightarrow$	0	.01	5	0	.01	5	0	.01	5	0	.01	5
OLS	$B(\hat{\beta}_1)$	-2.0000	-2.0074	.1575	-.0008	-.0002	.0000	1.0017	1.0030	-.0780	4.9988	5.0214	-.3904
	$B(\hat{\beta}_2)$	.0003	.0197	-.8077	.0004	.0000	.0002	.0000	-.0096	.3999	-.0003	-.0492	2.0028
2SLS	$B(\hat{\beta}_1)$	-2.0003	-2.0118	.1603	-.0022	.0030	-.0001	1.0042	1.0029	-.0810	5.0006	5.0208	-.3941
	$B(\hat{\beta}_2)$	.0003	.0203	-.8055	.0005	-.0002	.0010	-.0001	-.0098	.4069	-.0004	-.0503	1.9816
$\hat{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0000	-2.0074	.1575	-.0008	-.0002	.0000	1.0017	1.0030	-.0780	4.9988	5.0214	-.3904
	$B(\hat{\beta}_2)$	-.0072	-.0057	-.8700	-.0078	-.0250	-2.5584	-.0080	-.0355	-3.4214	-.0075	-.0757	-6.8326
$\tilde{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0000	-2.0074	.1575	-.0008	-.0002	.0000	1.0017	1.0030	-.0780	4.9988	5.0214	-.3904
	$B(\hat{\beta}_2)$	.0003	.0202	-2.9409	.0004	-.0001	.0030	.0000	-.0098	1.4644	-.0004	-.0503	7.2995
2 Stage estimators, Y_1 is not used as IV													
GW	$B(\hat{\beta}_1)$	-2.0000	-2.0075	.1603	-.0008	-.0001	.0000	1.0018	1.0030	-.0810	4.9989	5.0214	-.3941
	$B(\hat{\beta}_2)$	-.0070	-.0050	-.8056	-.0075	-.0243	-2.4706	-.0078	-.0348	.3774	-.0073	-.0750	1.9816
MGW	$B(\hat{\beta}_1)$	-2.0000	-2.0075	.1603	-.0008	-.0001	.0000	1.0018	1.0030	-.0810	4.9989	5.0214	-.3941
	$B(\hat{\beta}_2)$	-.0070	-.0050	-.8056	-.0075	-.0243	-2.4632	-.0078	-.0348	.3793	-.0073	-.0750	1.9816
2 Stage estimators, Y_1 used as IV													
GW	$B(\hat{\beta}_1)$	-2.0000	-2.0075	.1603	-.0008	-.0001	.0000	1.0018	1.0030	-.0810	4.9989	5.0214	-.3941
	$B(\hat{\beta}_2)$	.0003	.0202	-.8064	.0004	-.0001	.0030	.0000	-.0098	.4150	-.0004	-.0503	1.9816
MGW	$B(\hat{\beta}_1)$	-2.0000	-2.0075	.1603	-.0008	-.0001	.0000	1.0018	1.0030	-.0810	4.9989	5.0214	-.3941
	$B(\hat{\beta}_2)$	.0003	.0202	-.8064	.0004	-.0001	.0030	.0000	-.0098	.4145	-.0004	-.0503	1.9816

Table 3.7. DGP1: Bias of OLS, 2SLS Partial IV's and Pretest-estimators: weak IV for Y_2 ($\eta_2 = .01$)

$H_{\delta_1} : \delta_1 = 0$, strong IV for Y_2													
	$\delta_1 \rightarrow$	-2			0			1			5		
Estimators	$Bias \downarrow$ $\eta_1 \rightarrow$	0	.01	5	0	.01	5	0	.01	5	0	.01	5
OLS	$B(\hat{\beta}_1)$	-2.0004	-1.9990	-.0792	-.0018	.0007	-.0003	.9988	1.0018	.0393	5.0029	4.9989	.2010
	$B(\hat{\beta}_2)$	.0004	-.0006	-1.1530	.0013	.0002	-.0005	-.0016	.0000	.5774	-.0028	.0006	2.8807
2SLS	$B(\hat{\beta}_1)$	-1.9982	-1.9962	-.0151	-.0074	.0021	-.0003	1.0038	.9985	.0071	5.0013	5.0003	.0404
	$B(\hat{\beta}_2)$	-.0010	.0000	-1.1979	.0054	-.0014	-.0043	-.0069	-.0008	.6033	-.0037	-.0022	2.9821
$\hat{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0004	-1.9990	-.0792	-.0018	.0007	-.0003	.9988	1.0018	.0393	5.0029	4.9989	.2010
	$B(\hat{\beta}_2)$	-1.1290	-1.1217	-1.2049	-1.1230	-1.1321	-.0671	-1.1344	-1.1297	.5205	-1.1276	-1.1307	2.7966
$\tilde{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0004	-1.9990	-.0792	-.0018	.0007	-.0003	.9988	1.0018	.0393	5.0029	4.9989	.2010
	$B(\hat{\beta}_2)$	.0001	.0017	-2.8954	.0040	-.0012	-.0026	-.0070	-.0005	1.4519	-.0045	-.0027	7.2572
2 Stage estimators, Y_1 is not used as IV													
GW	$B(\hat{\beta}_1)$	-2.0004	-1.9989	-.0163	-.0020	.0007	-.0003	.9989	1.0017	.0106	5.0029	4.9989	.0415
	$B(\hat{\beta}_2)$	-1.1038	-1.0918	-1.1981	-1.0950	-1.1008	-.0651	-1.1084	-1.1030	.5943	-1.1015	-1.1015	2.9808
MGW	$B(\hat{\beta}_1)$	-2.0004	-1.9989	-.0162	-.0020	.0007	-.0003	.9989	1.0017	.0105	5.0029	4.9989	.0413
	$B(\hat{\beta}_2)$	-1.0999	-1.0891	-1.1980	-1.0930	-1.0998	-.0650	-1.1061	-1.1009	.5946	-1.0989	-1.0999	2.9810
2 Stage estimators, Y_1 used as IV													
GW	$B(\hat{\beta}_1)$	-2.0004	-1.9989	-.0163	-.0020	.0007	-.0003	.9989	1.0017	.0106	5.0029	4.9989	.0415
	$B(\hat{\beta}_2)$	.0001	.0017	-1.2298	.0041	-.0012	-.0026	-.0070	-.0006	.6954	-.0045	-.0027	3.0111
MGW	$B(\hat{\beta}_1)$	-2.0004	-1.9989	-.0162	-.0020	.0007	-.0003	.9989	1.0017	.0105	5.0029	4.9989	.0413
	$B(\hat{\beta}_2)$	.0001	.0017	-1.2263	.0041	-.0012	-.0026	-.0070	-.0006	.6922	-.0045	-.0027	3.0077

Table 3.8. DGP2: Bias of OLS, 2SLS, Partial IV and Pretest-estimators: strong IV for Y_2 ($\eta_2 = 1$)

$H_{\delta_1} : \delta_1 = 0$, strong IV for Y_2													
	$\delta_1 \rightarrow$	-2			0			1			5		
Estimators	$Bias \downarrow\ \eta_1 \rightarrow$	0	.01	5	0	.01	5	0	.01	5	0	.01	5
OLS	$B(\hat{\beta}_1)$	-2.0008	-2.0097	.1659	-.0025	.0002	.0000	.9986	1.0035	-.0825	5.0021	5.0201	-.4127
	$B(\hat{\beta}_2)$	.0001	.0189	-.8673	-.0002	.0002	-.0001	.0000	-.0089	.4309	-.0005	-.0472	2.1567
2SLS	$B(\hat{\beta}_1)$	-1.9996	-2.0077	.1735	-.0030	.0029	-.0006	.9991	1.0028	-.0854	5.0066	5.0195	-.4295
	$B(\hat{\beta}_2)$	.0000	.0197	-.8754	-.0001	.0001	.0029	.0000	-.0092	.4306	-.0006	-.0492	2.1659
$\hat{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0008	-2.0097	.1659	-.0025	.0002	.0000	.9986	1.0035	-.0825	5.0021	5.0201	-.4127
	$B(\hat{\beta}_2)$	-.0139	-.0135	-.8875	-.0161	-.0326	-2.6135	-.0156	-.0421	-3.4918	-.0157	-.0809	-6.9436
$\tilde{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0008	-2.0097	.1659	-.0025	.0002	.0000	.9986	1.0035	-.0825	5.0021	5.0201	-.4127
	$B(\hat{\beta}_2)$	.0001	.0197	-3.0053	-.0001	.0002	.0006	.0000	-.0093	1.4941	-.0006	-.0493	7.4445
2 Stage estimators, Y_1 is not used as IV													
GW	$B(\hat{\beta}_1)$	-2.0008	-2.0097	.1735	-.0025	.0002	.0000	.9986	1.0035	-.0854	5.0022	5.0201	-.4295
	$B(\hat{\beta}_2)$	-.0135	-.0125	-.8754	-.0156	-.0317	-2.5125	-.0152	-.0412	.3666	-.0153	-.0800	2.1558
MGW	$B(\hat{\beta}_1)$	-2.0008	-2.0097	.1735	-.0025	.0002	.0000	.9986	1.0035	-.0854	5.0022	5.0201	-.4295
	$B(\hat{\beta}_2)$	-.0137	-.0128	-.8755	-.0158	-.0321	-2.5465	-.0153	-.0415	.3215	-.0154	-.0803	2.1449
2 Stage estimators, Y_1 used as IV													
GW	$B(\hat{\beta}_1)$	-2.0008	-2.0097	.1735	-.0025	.0002	.0000	.9986	1.0035	-.0854	5.0022	5.0201	-.4295
	$B(\hat{\beta}_2)$	.0001	.0197	-.8773	-.0001	.0002	.0007	.0000	-.0093	.4479	-.0006	-.0493	2.1717
MGW	$B(\hat{\beta}_1)$	-2.0008	-2.0097	.1735	-.0025	.0002	.0000	.9986	1.0035	-.0854	5.0022	5.0201	-.4295
	$B(\hat{\beta}_2)$	.0001	.0197	-.8854	-.0001	.0002	.0007	.0000	-.0093	.4601	-.0006	-.0493	2.1780

Table 3.9. DGP2: Bias of OLS, 2SLS, Partial IV and Pretest-estimators: weak IV for Y_2 ($\eta_2 = .01$)

$H_{\delta_1} : \delta_1 = 0$, weak IV for Y_2													
	$\delta_1 \rightarrow$	-2			0			1			5		
Estimators	$Bias \downarrow \ \eta_1 \rightarrow$	0	.01	5	0	.01	5	0	.01	5	0	.01	5
OLS	$B(\hat{\beta}_1)$	-1.9995	-2.0004	-.1876	.0010	-.0012	.0001	.9981	1.0007	.0933	5.0001	5.0000	.4606
	$B(\hat{\beta}_2)$	.0014	.0001	.0037	-.0008	-.0001	-.0003	-.0019	.0010	-.0020	.0003	.0023	.0105
2SLS	$B(\hat{\beta}_1)$	-1.9974	-1.9980	-.0383	.0009	-.0062	.0000	.9994	.9995	.0188	4.9989	4.9993	.0881
	$B(\hat{\beta}_2)$	-.0003	-.0048	.0041	-.0028	-.0008	.0011	.0037	.0006	.0008	-.0004	.0060	.0021
$\hat{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-1.9995	-2.0004	-.1876	.0010	-.0012	.0001	.9981	1.0007	.0933	5.0001	5.0000	.4606
	$B(\hat{\beta}_2)$	-.0008	-.0052	-.0026	-.0017	-.0005	.0014	.0045	-.0014	.0032	-.0026	.0041	.0152
$\tilde{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-1.9995	-2.0004	-.1876	.0010	-.0012	.0001	.9981	1.0007	.0933	5.0001	5.0000	.4606
	$B(\hat{\beta}_2)$	-.0002	-.0037	.0295	-.0023	-.0002	.0021	.0031	.0011	-.0187	-.0018	.0050	.0029
2 Stage estimators, Y_1 is not used as IV													
GW	$B(\hat{\beta}_1)$	-1.9995	-2.0004	-.0383	.0010	-.0013	.0001	.9981	1.0006	.0188	5.0000	4.9653	.0881
	$B(\hat{\beta}_2)$	-.0008	-.0052	.0041	-.0017	-.0005	.0014	.0045	-.0014	.0008	-.0026	.0041	.0021
MGW	$B(\hat{\beta}_1)$	-1.9995	-2.0004	-.0383	.0010	-.0012	.0001	.9981	1.0006	.0188	5.0001	5.0000	.0881
	$B(\hat{\beta}_2)$	-.0008	-.0052	.0041	-.0017	-.0005	.0014	.0045	-.0014	.0008	-.0026	.0041	.0021
2 Stage estimators, Y_1 used as IV													
GW	$B(\hat{\beta}_1)$	-1.9995	-2.0004	-.0383	.0010	-.0013	.0001	.9981	1.0006	.0188	5.0000	4.9653	.0881
	$B(\hat{\beta}_2)$	-.0002	-.0037	.0041	-.0023	-.0002	.0021	.0031	.0011	.0008	-.0017	.0050	.0021
MGW	$B(\hat{\beta}_1)$	-1.9995	-2.0004	-.0383	.0010	-.0012	.0001	.9981	1.0006	.0188	5.0001	5.0000	.0881
	$B(\hat{\beta}_2)$	-.0002	-.0037	.0041	-.0023	-.0002	.0021	.0031	.0011	.0008	-.0017	.0050	.0021

Table 3.10. DGP1: MSE of OLS, 2SLS, Partial IV and Pretest-estimators)

	$H_{\delta_1} : \delta_1 = 0$, strong IV for Y_2											
$\delta_1 \rightarrow$	-2			0			1			5		
Estimators $\downarrow$ $\eta_1 \rightarrow$	0	.01	5	0	.01	5	0	.01	5	0	.01	5
OLS	4.0112	4.0410	1.7576	.0112	.0109	.0180	1.0143	1.0172	1.2003	24.9992	25.2278	6.5782
2SLS	4.0623	4.1078	1.1518	.0592	.0597	.1018	1.0665	1.0659	.3688	25.0663	25.2693	6.4758
$\hat{\beta}_{PIV}$	4.0131	4.0425	1.2952	.0131	.0134	7.3990	1.0162	1.0203	13.2980	25.0011	25.2329	54.8700
$\tilde{\beta}_{PIV}$	4.0112	4.0410	9.3253	.0112	.0110	.0672	1.0143	1.0173	2.3672	24.9992	25.2279	57.1610
2 Stage estimators, Y_1 is not used as IV												
GW	4.0131	4.0429	1.1517	.0130	.0133	6.9049	1.0162	1.0202	.3450	25.0015	25.2325	6.4758
MGW	4.0131	4.0429	1.1517	.0130	.0133	6.8639	1.0163	1.0202	.3465	25.0015	25.2325	6.4758
2 Stage estimators, Y_1 used as IV												
GW	4.0113	4.0415	1.1530	.0112	.0110	.0672	1.0144	1.0173	.3745	24.9997	25.2277	6.4758
MGW	4.0113	4.0416	1.1530	.0112	.0110	.0672	1.0144	1.0173	.3741	24.9998	25.2277	6.4758

	$H_{\delta_1} : \delta_1 = 0$, weak IV for Y_2											
$\delta_1 \rightarrow$	-2			0			1			5		
Estimators $\downarrow$ $\eta_1 \rightarrow$	0	.01	5	0	.01	5	0	.01	5	0	.01	5
OLS	4.0215	4.0249	0.0951	0.0233	0.0230	0.0132	1.0191	1.0250	0.0323	25.0240	25.0234	0.5238
2SLS	4.1188	4.1209	0.3251	0.1323	0.1280	0.0671	1.1326	1.1306	0.1324	25.1208	25.1250	1.7081
$\hat{\beta}_{PIV}$	4.0841	4.0870	0.4699	0.0863	0.0857	0.0728	1.0827	1.0893	0.1656	25.0886	25.0879	2.5021
$\tilde{\beta}_{PIV}$	4.0715	4.0738	1.2013	0.0740	0.0725	0.0634	1.0699	1.0751	0.3330	25.0740	25.0742	7.0935
2 Stage estimators, Y_1 is not used as IV												
GW	4.0832	4.0862	0.3251	0.0858	0.0852	0.0723	1.0822	1.0887	0.1324	25.0876	25.0871	1.7081
MGW	4.0835	4.0864	0.3251	0.0860	0.0854	0.0724	1.0823	1.0889	0.1324	25.0879	25.0874	1.7081
2 Stage estimators, Y_1 used as IV												
GW	4.0713	4.0736	0.3251	0.0740	0.0725	0.0634	1.0701	1.0751	0.1324	25.0737	25.0740	1.7081
MGW	4.0714	4.0736	0.3251	0.0740	0.0725	0.0634	1.0700	1.0751	0.1324	25.0738	25.0741	1.7081

Chapter 4

Identification-robust inference for error-regressors covariances and partial IV regression

1. Introduction

Testing or building confidence sets for subvectors of covariance is an important issue in econometrics. In models where more than one (supposedly) endogenous explanatory variable are involved, one often needs to determine which ones are exogenous, *i.e.* independent of the disturbances. This problem arise in many empirical applications. For example, in the New Keynesian Phillips Curve model, one should like to have a valid procedure for testing the partial exogeneity of interest rate or unemployment rate without imposing restriction about the exogeneity of inflation rate. In wage model, one often needs to assess whether mother's or father's education is a valid instruments (*i.e.*, exogenous) for schooling without assuming that ability and schooling are exogenous. Standard exogeneity tests proposed by Durbin-Wu-Hausman (DWH) [Durbin (1954), Wu (1973), Hausman (1978)] which are widely used in applied work cannot deal with such problems.

A solution to these problems is to use the Wald-type (GW) procedures proposed in Dufour (1987), Doko and Dufour (2009c). When the IV are strong, the Wald-type tests are typically valid and allow the construction of confidence sets on covariances or the test of linear restrictions on covariances. However, these procedures are in general size distorted when the available instruments are weak [see Doko and Dufour (2009c)]. So, the application of such tests in a model where the instruments are possibly weak as in–the returns to education [Dufour and Taamouti (2006), Angrist and Krueger (1991), Angrist and Krueger (1995), Angrist and al. (1999), Mankiw and al. (1992)], trade and growth [Irwin and Tervio (2002), Frankel and Romer (1999), Harrison (1996), Mankiw and al. (1992)]; the New Keynesian Phillips Curve [Dufour and Khalaf (2006)]; pregnancy and the demand for cigarettes [Bradford (2003)]; life cycle labor supply [MaCurdy (1981), Altonji (1986), Ham and Kevin (2002)]– is unreliable. This raises the following question: can one propose identification-robust procedure which can be used as partial exogeneity test even in presence of weak instruments? In other words, is it possible to build a valid partial identification-robust exogeneity tests which will have a good power even if the IV are weak?

This paper focuses on structural models and propose a finite- and large-sample procedure for building identification-robust confidence for: (1) the covariance between endogenous explanatory variables and errors; and (2) an auxiliary endogeneity parameter related to the error-regressors covariances. This later parameter has an interesting interpretation in many economic models. For example, in models with latent variables–Unanticipated money growth and unemployment in United States [Barro (1977)], Tobin's marginal q model of investment [Tobin (1969), Dufour and Jasiak (2001)], Students' academic achievements [Montmarquette and Mahseredjian (1989), Dufour and Jasiak (2001)]– it appears as the coefficient of omitted variable.

We provide analytic forms of the confidence sets and characterize necessary and sufficient conditions under which they are bounded. We also showed that the procedure is valid for subvectors of covariances. In particular, identification-robust confidence interval for the covariance between each endogenous regressor and the errors can be obtained. Then, the result can be applied to test the partial exogeneity of each regressor. Therefore, our procedure can be viewed as identification-robust partial exogeneity tests. Our asymptotic theory under weaker assumptions and allowing for heteroskedasticity and autocorrelation of errors confirms our finite-sample results. The Monte Carlo experiment shows that the procedure has a power even when IV are weak (contrary to GW-test and DWH-tests).

Moreover, we propose two new pretest-estimators where our procedure is used as pre-test in the first stage to assess whether a subset of explanatory variables is exogenous. Both pretest-estimators combine 2SLS and partial IV-estimators of the structural parameter. The first partial IV-estimator is obtained by treating a subset of variables as exogenous (dependently on the outcome of the pre-test) but not included in the set of instruments. The second partial IV-estimator is obtained by using this subset as additional instruments. We present Monte Carlo simulations indicating that: (a) like OLS [see Doko and Dufour (2009a)], partial IV-estimators outperform 2SLS when instruments are weak; (2) pretest-estimators have an excellent overall performance–bias and MSE– compared with

2SLS. This suggests that the projection-based procedure can be used as a variable selection method to decide which variables should be instrumented and which ones are valid instruments.

We illustrate our results through two empirical applications: the relation between trade and economic growth and the widely studied problem of returns to education. The results showed that the confidence sets for the covariance and the endogeneity parameter are unbounded. That is, the quality of the instruments in these models is poor, as questioned in the literature [Bound (1995), Frankel and Romer (1999)].

The paper is organized as follows. Section 2 formulates the model considered. Section 3 presents the finite-sample theory and Section 4, the asymptotic theory. Section 6 studies through a Monte Carlo experiment: (1) the properties (level and power) of our projection-based procedure; (2) the performance (bias and MSE) of different estimators including our new pretest-estimators. Section 7 illustrates our theoretical results through two empirical applications: the trade an growth model and the returns to education. We conclude in Section 9. Proofs are presented in the Appendix.

2. Framework

Consider the common simultaneous equation model described by the following assumptions:

$$y = Y\beta + X_1\gamma + u\,, \tag{2.1}$$

where y is a $T \times 1$ vector of observations on the dependent variable, Y is a $T \times G$ matrix of observations on the explanatory endogenous variables ($G \geq 1$), X_1 is a $T \times k_1$ matrix of observations on the included exogenous variables, $u = \begin{bmatrix} u_1 & \ldots & u_T \end{bmatrix}'$ is a vector of structural disturbances, β and γ are $G \times 1$ and $k_1 \times 1$ vectors of unknown coefficients.

Assume that

$$Y = [Y_1,\, Y_2]\,, \tag{2.2}$$

where Y_1 and Y_2 are respectively $T \times G_1$ and $T \times G_2$ matrices, $G_1 + G_2 = G$. If $G_1 = 0$ or $G_2 = 0$, we shall assume by convention that the corresponding variable Y_1 or Y_2 drops out in model (2.1). If we partition β according to the partition of Y in (2.2), then, (2.1) can becomes

$$y = Y_1\beta_1 + Y_2\beta_2 + X_1\gamma + u\,, \tag{2.3}$$

where $\beta = [\beta_1',\, \beta_2']'$ and $\beta_1,\ \beta_2$ are respectively $G_1 \times 1$ and $G_2 \times 1$ vectors.

Let

$$Y = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2 + V\,, \tag{2.4}$$

where X_2 is a $T \times k_2$ matrix of observations on the excluded exogenous variables, $\mathit{\Pi}_1$ and $\mathit{\Pi}_2$ are $k_1 \times G$ and $k_2 \times G$ matrices of unknown coefficients, $V = \begin{bmatrix} w_1 & \ldots & w_T \end{bmatrix}'$ is a $T \times G$ matrix of disturbances. From (2.2), (2.4) can be written as

$$Y_1 = X_1\mathit{\Pi}_{11} + X_2\mathit{\Pi}_{21} + V_1\,, \tag{2.5}$$

$$Y_2 = X_1\mathit{\Pi}_{12} + X_2\mathit{\Pi}_{22} + V_2\,, \tag{2.6}$$

where

$$\mathit{\Pi}_1 = [\mathit{\Pi}_{11},\, \mathit{\Pi}_{12}],\ \mathit{\Pi}_2 = [\mathit{\Pi}_{21},\, \mathit{\Pi}_{22}], \quad \text{and} \;\; V = [V_1,\, V_2]\,. \tag{2.7}$$

We assume that

$$X = [X_1 : X_2] \in \mathbb{R}^{T\times k} \text{ has full-column rank} \tag{2.8}$$

where $k = k_1 + k_2$. The usual necessary and sufficient condition for identification of this model is

$$\mathrm{rank}(\mathit{\Pi}_2) = G\,. \tag{2.9}$$

If $\mathrm{rank}(\mathit{\Pi}_2) < G$, then β is not identified and the instruments X_2 are weak. Nevertheless, remark that some linear combinations of β may be identified even if $\mathrm{rank}(\mathit{\Pi}_2) < G$ [see Dufour and Hsiao (2008)], except for $\mathit{\Pi}_2 = 0$. In particular, some components or subvectors of β may be identified.

The reduced form for $[y, Y]$ can be written as

$$y = X_1\pi_1 + X_2\pi_2 + v\,, \tag{2.10}$$

$$Y = X_1\Pi_1 + X_2\Pi_2 + V\,, \tag{2.11}$$

where $\pi_1 = \gamma + \Pi_1\beta,\ \pi_2 = \Pi_2\beta,$ and $v = u + V\beta = [v_1, \ldots, v_T]'$. Let

$$M = M_X = I - X(X'X)^{-1}X', \quad M_1 = M_{X_1} = I - X_1(X_1'X_1)^{-1}X_1'. \tag{2.12}$$

Then, we have

$$M_1 - M = M_1X_2(X_2'M_1X_2)^{-1}X_2'M_1. \tag{2.13}$$

Define

$$U = [u,\, V] = [U_1,\, \ldots,\, U_T]'\,, \tag{2.14}$$

$$W = [v,\, V] = [u + V\beta,\, V] = [W_1,\, W_2,\, \ldots,\, W_T]'\,. \tag{2.15}$$

We shall assume that the vectors $U_t = [u_t,\, V_t']'$, $t = 1, \ldots, T$, have the same nonsingular covariance matrix:

$$E[U_tU_t'] = \Sigma = \begin{bmatrix} \sigma_u^2 & \delta' \\ \delta & \Sigma_V \end{bmatrix} > 0\,, \quad t = 1, \ldots, T, \tag{2.16}$$

where Σ_V has dimension G. Then the covariance matrix of the reduced-form disturbances $W_t = [v_t, V_t']'$ errors in (2.10) - (2.11) also have the same covariance matrix, which takes the form:

$$\Omega = \begin{bmatrix} \sigma_u^2 + \beta'\Sigma_V\beta + 2\beta'\delta & \beta'\Sigma_V + \delta' \\ \Sigma_V\beta + \delta & \Sigma_V \end{bmatrix} \tag{2.17}$$

where Ω is positive definite. In the above framework, the standard exogeneity hypothesis can be expressed as

$$H_\delta : \delta = 0\,. \tag{2.18}$$

If we regress u on V, we get

$$u = Va + \varepsilon\,, \tag{2.19}$$

where a is a $G \times 1$ vector of unknown coefficients, ε is independent of V with mean zero and variance σ_{ε}^2.

From (2.19) and (2.15) - (4.36), we can easily see that

$$\Sigma_V a \;=\; \delta\,, \quad \sigma_{\varepsilon}^2 = \sigma_u^2 - \delta' \Sigma_V^{-1} \delta = \sigma_u^2 - a' \Sigma_V a\,. \tag{2.20}$$

So, $a = 0 \Leftrightarrow \delta = 0$, which means that the parameter a as well as δ characterize the exogeneity of Y in model (2.1) - (2.4). So, the exogeneity hypothesis of Y can be assessed here from

$$H_0 : a = 0 \tag{2.21}$$

by resorting to Durbin-Wu-Hausman (DWH)-type tests [Doko and Dufour (2009a)]. The difficulty with DWH-type arises when testing the hypotheses of the form

$$H_{a_0} : a = a_0 \quad \text{or} \quad H_{\delta_0} : \Gamma\delta = d_0\,, \tag{2.22}$$

where Γ is a given $r \times G$ matrix, $a_0 \neq 0$ and $d_0 \neq 0$. Durbin-Wu-Hausman (DWH)-type procedures are unusable for testing (2.22). The generalized Wald-type procedures proposed by Dufour (1987) and extended by Doko and Dufour (2009c) to no Gaussian errors typically deal with such hypotheses. However, the Wald-type procedures relies on the assumption that model parameters are identified by the available instruments. When IV are weak, Wald-type procedures are in general size distorted. Even these procedures are valid (level is controlled) when testing partial exogeneity [Doko and Dufour (2009c)], they have a low power when all IV are weak. So, the question now is can one propose identification-robust partial exogeneity which will exhibit power even in presence of weak IV?

Moreover, the endogenous parameter "a" itself may be of interest. In many economic models [see Barro (1977), Tobin (1969), Dufour and Jasiak (2001), Montmarquette and Mahseredjian (1989)], This parameter appears as the coefficient of omitted variables. So, we should like to assess whether $a = 0$ or not.

In this paper, we develop identification-robust procedure which overcomes the above difficulties.

We consider the problem of testing the joint hypothesis

$$H(\beta_0, a_0) : \beta = \beta_0,\, a = a_0\,, \tag{2.23}$$

where β_0 and a_0 are given $G \times 1$ vectors. First, we focus on building CS's for a and subvectors of a. Second, we deduce identification-robust CS's for δ and subvectors of δ from (2.20).

For our finite-sample distributional theory, we assume that

$$W_t = \Phi \bar{W}_t\,,\, t = 1, \ldots, T\,, \tag{2.24}$$

where the vector $W_{(T)} = vec(\bar{W}_1, \ldots, \bar{W}_T)$ has a known distribution $F_{\bar{W}}$ and $\Phi \in R^{(G+1)\times(G+1)}$ is an unknown upper triangular nonsingular matrix [for more details, see Doko and Dufour (2009a), Dufour and Khalaf (2002) and Dufour et al. (2008)]. Furthermore, the matrix Φ is the unique lower triangular matrix satisfying (2.24), *i.e.*

$$\Phi = \begin{bmatrix} \Phi_{11} & 0 \\ \Phi_{21} & \Phi_{22} \end{bmatrix} \tag{2.25}$$

where $\Phi_{11} \neq 0$ is a scalar and Φ_{22} is a nonsingular $G \times G$ matrix. In particular, these conditions are satisfied when

$$\bar{W}_t \overset{i.i.d.}{\sim} N[0,\, I_{G+1}]\,,\, t = 1, \ldots, T\,. \tag{2.26}$$

Following Doko and Dufour (2009a), we have

$$\bar{W} = U\Phi = [\bar{v},\, \bar{V}] = [\bar{W}_1, \ldots, \bar{W}_T]',\; \bar{W}_t = [\bar{v}_t,\, \bar{V}_t']', \tag{2.27}$$

$$\bar{v} = v\Phi_{11} + V\Phi_{21} = [\bar{v}_1, \ldots, \bar{v}_T]', \quad \bar{V} = V\Phi_{22} = \left[\bar{V}_1, \ldots, \bar{V}_T\right]', \tag{2.28}$$

so that

$$M(\bar{y} - \bar{Y}\beta) = M\bar{v}\,, \, M\bar{Y} = M\bar{V}\,, \tag{2.29}$$

$$M_1(\bar{y} - \bar{Y}\beta) = M_1[\bar{v} + \mu_2(\beta - \beta_0) + \bar{V}(\beta - \beta_0)]\,, \, M_1\bar{Y} = M_1(\mu_2 + \bar{V})\,. \tag{2.30}$$

where

$$\mu_2 = M_1 X_2 \Pi_2 P_{22}, \quad \left[\bar{y}, \bar{Y}\right] = [y, Y]\,\Phi. \tag{2.31}$$

We propose a five step procedure in order to build identification-robust CS's for a and covariances:

1. build identification-robust CS's with level $1 - \alpha_1$ for the structural parameter β;

2. build identification-robust CS's with level $1 - \alpha_2$ for the transformation $\theta = \beta + a$;

3. build identification-robust CS's with level $1 - \alpha$ for (β, θ) [where $\alpha \equiv \alpha(\alpha_1, \alpha_2)$];

4. use projection techniques [Dufour (1997), Abdelkhalek and Dufour (1998), Dufour and Jasiak (2001)] to get identification-robust CS's with $1 - \alpha$ for a (or subvectors of a);

5. deduce identification-robust CS's with level $1 - \alpha$ for δ (or subvectors of δ).

We distinguish two setups: finite-and -large sample theory.

Section 3 below presents the finite-sample theory.

3. Finite-sample theory

In this section, we derive finite-sample CS's for the full vector a as well as its subvectors under the assumption (2.24). The next subsection focuses on the construction of CS's for the full vector a.

3.1. Confidence sets for the full vector a

This subsection proposes identification-robust procedure for building CS's for a. We will proceed step by step as described in in Section 2.

3.1.1. Inference on structural parameter

We consider the hypothesis

$$H_{\beta_0} : \beta = \beta_0 \,, \tag{3.1}$$

where β_0 is a given $G \times 1$ vectors. The Anderson and Rubin (1949, AR) test for H_{β_0} involves considering the transformed equation

$$y - Y\beta_0 = X_1\pi_1^0 + X_2\pi_2^0 + v^0 \,, \tag{3.2}$$

where $\pi_1^0 = \gamma + \Pi_1(\beta - \beta_0)$, $\pi_2^0 = \Pi_2(\beta - \beta_0)$ and $v^0 = u + V(\beta - \beta_0)$. If any restriction is imposed on γ (which is typically the case), H_{β_0} can then be assessed by testing $H_{\pi_2} : \pi_2^0 = 0$ using the standard F-statistic [say $AR(\beta_0)$]. Under H_{π_2}, we have

$$AR(\beta_0) = \frac{(y - Y\beta_0)'(M_1 - M)(y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(y - Y\beta_0)/(T - k)} \,, \tag{3.3}$$

where the projection matrices M_1 and M are defined in (2.12). By a similar argument as those in Doko and Dufour (2009a, Lemma 4.5), we show in the following theorem that $AR(\beta_0)$ is pivotal irrespective of whether the instruments are weak or not.

Theorem 3.1 PIVOTALITY OF THE ANDERSON-RUBIN STATISTIC. *Under the assumptions of the model and if $\beta = \beta_0$ where β_0 is a $G \times 1$ constant vector, then*

$$AR(\beta_0) = \frac{\bar{v}'(M_1 - M)\bar{v}/k_2}{\bar{v}'M\bar{v}/(T - k)} \,. \tag{3.4}$$

If furthermore (2.26) holds and if u is independent of X, then

$$AR(\beta_0) \sim F(k_2, T-k)\,, \tag{3.5}$$

where $k = k_1 + k_2$.

The above theorem shows that given X, the distribution of $AR(\beta_0)$ only depends on $\bar{v}$. Hence, if the distribution of $\bar{v}$ given X can be simulated, one can get exact tests from $AR(\beta_0)$ by using Monte Carlo procedure [see Dufour (2006)] and these tests are robust to weak instruments even if u and X are correlated. So, a confidence set for β with level $1-\alpha_1$ is given by

$$C_\beta(\alpha_1) = \{\beta_0 : AR(\beta_0) \le c_{\alpha_1}\}\ , \tag{3.6}$$

where $c_{\alpha_1} \equiv c(\alpha_1;\, k_2, T-k)$ is the $1-\alpha_1$ quantile of the distribution of the statistic. In particular, under (2.26) and if u is independent of X, $c_{\alpha_1} = F_{\alpha_1}(k_2, T-k)$ is the $1-\alpha_1$ quantile of the F distribution with $(k_2, T-k)$ degrees of freedom. In general, for nonstandard errors, the critical value c_{α_1} can be obtained by Monte carlo procedure [see Dufour (2006)]. So, as in Dufour and Taamouti (2005), we have

$$C_\beta(\alpha_1) = \left\{\beta_0 : \beta_0' A \beta_0 + b'\beta_0 + c \le 0\right\}\ . \tag{3.7}$$

where

$$A = Y'HY\,, \quad b = -2Y'Hy\,, \quad c = y'Hy\,, \tag{3.8}$$

$$H = M_1 - [1 + \frac{k_2}{T-k} c_{\alpha_1}]M\,. \tag{3.9}$$

The confidence set defined by (3.7) is a quadric confidence set [see Dufour and Taamouti (2005)].

3.1.2. Inference on a transformation of structural and endogeneity parameters

We now focus on testing the hypothesis

$$H_{\theta_0} : \theta = \theta_0 \,, \tag{3.10}$$

where $\theta = \beta + a$ and θ_0 is a given $G \times 1$ vectors. If we Substitute (2.19) into (2.1), we get

$$y = Y\beta + X_1\gamma + Va + \varepsilon \,, \tag{3.11}$$

where ε is defined in (2.19). Equation (3.11) illustrates that the existence of correlation between Y and u may be viewed as a problem of omitted variables. If the matrix V were observed, we would test any set of linear restriction on the coefficients β, γ and a in (3.11) by standard F-test, and these tests would be exact in finite-sample. In particular, linear hypotheses regarding the parameter a could be tested. Furthermore, if $\varSigma_V$ were known, the transformation $\delta = \varSigma_V a$ would allow to test some linear restrictions on δ by standard F-test. The difficulty, of course, is that neither V nor $\varSigma_V$ is known so that F-type tests cannot be applied. An alternative consists on replacing V by $\hat{V}$ in (3.11) and considering the equation

$$y = Y\beta + X_1\gamma + \hat{V}a + \varepsilon^* \,, \tag{3.12}$$

where $\varepsilon^* = X(\hat{\varPi} - \varPi)a + \varepsilon$,

$$\hat{\varPi} = (X'X)^{-1}X'Y \tag{3.13}$$

is the OLS estimate of $\varPi = [\varPi_1' \,, \varPi_2']'$ in (2.4) and $\hat{V} = MY$, the corresponding residuals. If we estimate (3.12) by OLS, the estimators $\hat{\beta}$ and $\hat{\gamma}$ obtained are the 2SLS estimators of β and γ from (2.1) and (2.4) [see Dufour (1987)]. Furthermore, the OLS estimator of a in (3.12) is given by

$$\hat{a} = (\hat{V}'\bar{M}_1\hat{V})^{-1}\hat{V}'\bar{M}_1 y = a + (\hat{V}'\bar{M}_1\hat{V})^{-1}\hat{V}'\bar{M}_1\varepsilon^* = a + \left[\frac{\hat{V}'\bar{M}_1\hat{V}}{T}\right]^{-1}\frac{\hat{V}'\bar{M}_1\varepsilon^*}{T} \,, \tag{3.14}$$

where $\bar{M} = M_{\bar{X}_1} = I - \bar{X}_1(\bar{X}_1'\bar{X}_1)^{-1}\bar{X}_1'$ and $\bar{X}_1 = [Y,\, X_1]$. So, provided the identification condition $\text{rank}(\Pi_2) = G$ holds,

$$\frac{\hat{V}'\bar{M}_1\hat{V}}{T} \xrightarrow{p} \Sigma_V \left\{ \Sigma_V^{-1} - \left[\Sigma_V + \Pi_2'(\Sigma_{X_2} - \Sigma_{X_1X_2}'\Sigma_{X_1}^{-1}\Sigma_{X_1X_2})\Pi_2 \right]^{-1} \right\} \Sigma_V \,, \tag{3.15}$$

$$\frac{\hat{V}'\bar{M}_1\varepsilon^*}{T} \xrightarrow{p} 0\,, \tag{3.16}$$

so that

$$\underset{T\to\infty}{\text{plim}}\; \hat{a} = a\,. \tag{3.17}$$

However, (3.17) does not hold in general if $\text{rank}(\Pi_2) < G,$ *i.e.* if identification is deficient or weak (weak instruments).

Now, let subtract the LHS and RHS of (3.12) by $\hat{V}a_0 = MYa_0$. Then, we have

$$y - MYa_0 = Y\beta + X_1\gamma + \hat{V}a^* + \varepsilon^*\,, \tag{3.18}$$

where $a^* = a - a_0$. Hence, unless $a_0 = 0$, the OLS estimator of a^* in (3.18) is biased and inconsistent even when identification is strong [$\text{rank}(\Pi_2) = G$] since the error terms ε^* are correlated with the regressors Y if $a = a_0 \neq 0$. Thus, standard exogeneity tests of the type Wu-Durbin-Hausman [see Doko and Dufour (2009a)] cannot typically be applied to test $H_{a_0} : a = a_0 \neq 0$. This paper alleviates this difficulties by using the projection-based techniques.

Now, let replace V by

$$V = Y - X_1\Pi_1 - X_2\Pi_2$$

in (3.11). Then, we get

$$y = Y\theta + X_1(\gamma - \Pi_1 a) + X_2(-\Pi_2 a) + \varepsilon \tag{3.19}$$

or equivalently

$$y - Y\theta_0 = Y\psi + X_1\pi_1^* + X_2\pi_2^* + \varepsilon \tag{3.20}$$

where $\psi = \theta - \theta_0$, $\pi_1^* = \gamma - \Pi_1 a$, $\pi_2^* = -\Pi_2 a$ and ε is independent of all the regressors. Hence, the hypothesis H_{θ_0} can be assessed by testing whether $\psi = 0$ in (3.20). Remark that if any restriction is imposed on γ, a is identified if and only if rank$(\Pi_2) = G$, which is also the necessary and sufficient condition for identification of β. Therefore, the endogeneity parameter a is not identified when IV are weak. Since V is identified by assumption, this entails that δ is not identified if the IV are weak. However, $\theta = \beta + a$ is always identified. This result means that our procedure has power even when IV are weak. This is observed in the Monte Carlo experiment.

Define

$$Z = [Y \, , \, X_1 \, , \, X_2] \, , \qquad \bar{M} \equiv M_Z = I - Z(Z'Z)^{-1}Z' \, . \tag{3.21}$$

We can easily show that

$$\bar{M} = M - P_{MY} = M M_{MY} M \, , \qquad M_{MY} = I - P_{MY} \, , \tag{3.22}$$

$$P_{MY} = MY(Y'MY)^{-1}Y'M \, , \tag{3.23}$$

where M is defined in (2.12). The Anderson and Rubin (1949, AR) test statistic for H_{θ_0} in (3.20) is given by

$$AR(\theta_0) = \frac{\varepsilon(\theta_0)' P_{MY} \varepsilon(\theta_0)/G}{\varepsilon(\theta_0)' \bar{M} \varepsilon(\theta_0)/(T - G - k)} \, , \tag{3.24}$$

where $\varepsilon(\theta_0) = (y - Y\theta_0)$. As in Theorem **3.1**, we can characterize the distribution of $AR(\theta_0)$ and show that this statistic is pivotal irrespective of whether the instruments are strong or weak. Furthermore, if model have gaussian distribution,

$$AR(\theta_0) \sim F(G, T - k - G) \, . \tag{3.25}$$

So, tests based on $AR(\theta_0)$ are robust to weak instruments. A confidence set with level $1 - \alpha_2$ for $\theta = \beta + a$ can be obtained by inverting the statistic $AR(\theta_0)$ as

$$C_\theta(\alpha_2) = \{\theta_0 : AR(\theta_0) \leq c_{\alpha_2}\} \, . \tag{3.26}$$

With a simple algebraic calculus, we find

$$C_\theta(\alpha_2) = \left\{\theta_0 : \theta_0'\tilde{A}\theta_0 + \tilde{b}'\theta_0 + \tilde{c} \le 0\right\} . \tag{3.27}$$

where

$$\tilde{A} = Y'\tilde{H}Y \, , \tilde{b}' = -2Y'\tilde{H}y \, , \tilde{c} = y'\tilde{H}y \, , \tag{3.28}$$

$$\begin{aligned} \tilde{H} &= M - \left[1 + \frac{G}{T-G-k}c_{\alpha_2}\right]\bar{M} \\ &= P_{MY} - \frac{G}{T-G-k}c_{\alpha_2}\bar{M} \, . \end{aligned} \tag{3.29}$$

3.1.3. Joint inference on structural and endogeneity parameters

We focus in this subsection on the joint parameter (β, θ). By using Bonferroni inequality, we can show that

$$P\left[\beta \in C_\beta(\alpha_1)\, , \, \theta \in C_\theta(\alpha_2)\right] \ge 1 - \alpha_1 - \alpha_2 \, , \tag{3.30}$$

where the confidence sets $C_\beta(\alpha_1)$ and $C_\theta(\alpha_2)$ are defined above. If we choose $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, then (3.30) becomes

$$P\left[\beta \in C_\beta(\alpha_1)\, , \, \theta \in C_\theta(\alpha_2)\right] \ge 1 - \alpha \tag{3.31}$$

So, the set

$$C_{(\beta,\theta)}(\alpha) = \{(\beta_0', \theta_0')' : \beta_0 \in C_\beta(\alpha_1) \ , \ \theta_0 \in C_\theta(\alpha_2)\} \tag{3.32}$$

is the confidence set with level $1-\alpha$ for the joint parameter (β, θ). Note that $C_{(\beta,\theta)}(\alpha)$ defined in (3.32) is simultaneous in the sense of Scheffé. From (3.7) and (3.27), $C_{(\beta,\theta)}(\alpha)$ can be expressed:

$$C_{(\beta,\theta)}(\alpha) = \{(\beta_0', \theta_0')' : \beta_0' A\beta_0 + b'\beta_0 + c \le 0 \ , \ \theta_0'\tilde{A}\theta_0 + \tilde{b}'\theta_0 + \tilde{c} \le 0\} \, . \tag{3.33}$$

It is worthwhile to note that even though the simultaneous confidence sets $C_\beta(\alpha_1)$ and $C_\theta(\alpha_2)$ may be interpreted as a confidence sets based on inverting LR-type tests or as a profile likelihood confidence sets [see Meeker and al. (1998)], the confidence sets $C_{(\beta,\theta)}(\alpha)$ is not (strictly speaking)

LR-type confidence sets. $C_{(\beta,\theta)}(\alpha)$ is obtained by taking the intersection of two quadrics.

3.1.4. Confidence sets for the endogeneity parameter

We study in this section, the problem of building confidence sets for the endogeneity parameter a or any transformation $g(a)$ of this parameter. The methodology used is the projection-based techniques [see Dufour (1990), Dufour and Jan (1998), Dufour and Jasiak (2001)].

Let $\phi = (\beta', \theta')' \in \mathbb{R}^{2G}$ and consider the transformation $g(\phi) \in \mathbb{R}^G$. Since $\phi \in C_\phi(\alpha)$ entails $g(\phi) \in g[C_\phi(\alpha)]$, we have

$$P\left\{g(\phi) \in g[C_\phi(\alpha)]\right\} \geq P\left\{\phi \in C_\phi(\alpha)\right\} \geq 1 - \alpha\,, \tag{3.34}$$

where

$$g[C_\phi(\alpha)] = \left\{g(\phi) \in \mathbb{R}^G : \phi \in C_\phi(\alpha)\right\}. \tag{3.35}$$

We see that the confidence set $g[C_\phi(\alpha)]$ has level $1-\alpha$. Note that (3.34) holds for any transformation $g(.)$. Therefore, $g(.)$ may be discontinuous or non differentiable transformation. In particular when $g(\phi) = a$, $g[C_\phi(\alpha)]$ can be interpreted as the projection of $C_\phi(\alpha)$ on the space spanned by the columns of a. In this case, we have

$$g[C_\phi(\alpha)] = \left\{a \in \mathbb{R}^G : \phi \in C_\phi(\alpha)\right\} \equiv C_a(\alpha)\,. \tag{3.36}$$

To build CS's for the full vector a, we take the transformation $g(\phi) = a$. Define

$$\mathcal{Q}(\beta, a) = \theta'\tilde{A}\theta + \tilde{b}'\theta + \tilde{c} = a'\tilde{A}a + \tilde{b}'a + \tilde{c} + \beta'\tilde{A}\beta + (2\tilde{A}a + \tilde{b})'\beta\,, \tag{3.37}$$

$$f(\beta) = \beta' A\beta + b'\beta + c\,. \tag{3.38}$$

Then, we can see that the confidence set $C_a(\alpha)$ in (3.36) is obtained by minimizing (3.37) subject to (3.38), *i.e.*

$$C_a(\alpha) = \left\{a \in \mathbb{R}^G :\ \mathcal{Q}^*(a) \leq 0\right\}\,, \tag{3.39}$$

where

$$\mathcal{Q}^*(a) = \min_{\substack{\beta \\ s.c\ f(\beta)\leq 0}} \mathcal{Q}(\beta, a)\,. \tag{3.40}$$

For any given matrix Λ, we shall denote by Λ^-, its generalized inverse and $\mathcal{H}_\Lambda = \Lambda^-\Lambda$. Define

$$\mathcal{L}(\bar{\beta}^*, \lambda) \;=\; \phi_0\lambda^2 - \phi_1\lambda + \phi_2, \tag{3.41}$$

where

$$\begin{aligned} \phi_0 &= (A\bar{\beta}^* + \frac{b}{2})'\tilde{A}^-(A\bar{\beta}^* + \frac{b}{2}),\ \phi_1 = \bar{\beta}'^* A\bar{\beta}^* + \frac{b'\bar{\beta}^*}{2}\,, \\ \phi_2 &= \bar{\beta}'^*\tilde{A}\bar{\beta}^* + \tilde{c} + \frac{b'\bar{\beta}^*}{2}, \quad \bar{\beta}^* \text{ is given by } f(\bar{\beta}^*) = 0\,. \end{aligned} \tag{3.42}$$

Finally, for any matrix $\mathcal{K}$, $\mathcal{K} \geq 0$ means that $\mathcal{K}$ is positive semidefinite (p.s.d.). Then, the form of the set $C_a(\alpha)$ is given by the following theorem:

Theorem 3.2 CLOSE-FORM CS'S FOR a. *Assume that the assumptions* (2.1) - (2.8), (2.24) *hold. If furthermore* $\beta = \beta_0$ *and* $a = a_0$, *where* β_0 *and* a_0 *are* $G \times 1$ *constant vectors, then the sets* $C_a(\alpha)$ *take one of the following forms* :

(a) *if* $2A\beta + b = 0$, i.e. $\beta = \bar{\beta} = -\frac{1}{2}A^-b + (I - \mathcal{H}_A)\beta_{0*}$, *then*

$$C_a(\alpha) = \begin{cases} \left\{a : a'\tilde{A}a + \bar{b}'a + \bar{c} \leq 0\right\} & \textit{if} \quad c^* \leq 0 \\ \emptyset \quad \textit{otherwise} \end{cases} \tag{3.43}$$

where $\bar{b} = \tilde{b} + 2\tilde{A}\bar{\beta}$ *and* $\bar{c} = \tilde{c} + \bar{\beta}'\tilde{A}\bar{\beta} + \tilde{b}'\bar{\beta}$, $c^* = c - \frac{1}{4}b'A^-b + \frac{1}{2}b'(I - \mathcal{H}_A)\beta_{0*}$; β_{0*} *is any arbitrary vector in* $\mathbb{R}^G$.

(b) *if* $2A\beta + b \neq 0$, *then,*

$(b.1)$ *if* $\tilde{A}+\hat{\lambda}A\geq 0$ *and* $\hat{\lambda}>0$,

$$C_a(\alpha)=\Big\{a: a'\tilde{A}a+\tilde{b}'a+\tilde{c}_{\hat{\lambda}}\leq 0,\quad \bar{\beta}^* \text{ is such that } \quad f(\bar{\beta}^*)=0\Big\}, \tag{3.44}$$

where $\tilde{c}_{\hat{\lambda}}=\tilde{c}-\bar{\beta}^{'*}\tilde{A}\bar{\beta}^*-\hat{\lambda}(A\bar{\beta}^*+b)'\bar{\beta}^*$.

$(b.2)$ *if* $\tilde{A}+\hat{\lambda}A\geq 0$ *and* $\hat{\lambda}=0$,

$$C_a(\alpha)=\Big\{a:\tilde{b}_*'a+\tilde{c}_*\leq 0\Big\}\cap\{a'\Lambda a+b_*'a+c_*\leq 0\}, \tag{3.45}$$

where $\tilde{b}_*=(I-\mathcal{H}'_{\tilde{A}})\tilde{b},\ \tilde{c}_*=\tilde{c}-\frac{1}{4}\tilde{b}'\tilde{A}^-\tilde{b}+\frac{1}{2}\tilde{b}'(I-\mathcal{H}_{\tilde{A}})\beta_{30},\ \Lambda=\mathcal{H}'_{\tilde{A}}A\mathcal{H}_{\tilde{A}}$, $b_*=\mathcal{H}'_{\tilde{A}}[-b+A\tilde{A}^-\tilde{b}-2A(I-H_{\tilde{A}})\beta_{30}],\ c_*=c+\frac{1}{4}\tilde{b}'\tilde{A}^-A\tilde{A}^-\tilde{b}+b'(I-\mathcal{H}_{\tilde{A}})\beta_{30}-\frac{b'\tilde{A}^-\tilde{b}}{2}-\tilde{b}'\tilde{A}^-A(I-\mathcal{H}_{\tilde{A}})\beta_{30}+\beta_{30}'(I-\mathcal{H}'_{\tilde{A}})A(I-\mathcal{H}_{\tilde{A}})\beta_{30}$, *and* β_{30} *is any arbitrary vector in* $\mathbb{R}^G$;

$(b.3)$ *if* $\tilde{A}+\hat{\lambda}A$ *is not positive semidefinite,*

$$C_a(\alpha)=\mathbb{R}^G, \tag{3.46}$$

where $\hat{\lambda}=arg[\min_\lambda \mathcal{L}(\bar{\beta}^*,\lambda)]$ *and* $\mathcal{L}(\bar{\beta}^*,\lambda)$ *is defined by* (3.41).

We can now characterize the sufficient conditions under which the sets $C_a(\alpha)$ are bounded. Corollary **3.3** below gives these conditions.

Corollary 3.3 N.S. CONDITIONS FOR BOUNDED SETS. *Assume that the assumptions of Theorem 3.***2** *hold. Then, the necessary and sufficient condition under which the set* $C_a(\alpha)$ *is bounded is that* A *and* $\tilde{A}$ *are positive definite. If* A *or* $\tilde{A}$ *is not positive semidefinite,* $C_a(\alpha)$ *is bounded only when it is empty.*

Note first that Corollary **3.3** can be proved by using the spectral decomposition argument as in Dufour and Taamouti (2005). Second, observe that $C_a(\alpha)$ is a bounded non empty set only when A and $\tilde{A}$ are both positive definite. We now derive CS's for subvectors of a.

3.2. Confidence sets for subvectors of a

We consider the following partition of $a,\ \delta,$ and Σ_V according to the partition of Y in (2.2), *i.e.*

$$a = [a_1',\, a_2']',\ \delta = [\delta_1',\, \delta_2']'\,, \Sigma_V = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}. \tag{3.47}$$

From (3.47), we have

$$\delta_1 = \Sigma_1 a_1 + \Sigma_{12} a_2,\ \delta_2 = \Sigma_{21} a_1 + \Sigma_2 a_2 \tag{3.48}$$

or equivalently

$$a_1 = \Sigma^{11}\delta_1 + \Sigma^{12}\delta_2,\ a_2 = \Sigma^{21}\delta_1 + \Sigma^{22}\delta_2\,, \tag{3.49}$$

where

$$\begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} = \Sigma_V^{-1},\ \Sigma^{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}\,, \tag{3.50}$$

$$\Sigma^{22} = \Sigma_{22}^{-1}\Sigma_{21}\Sigma^{11}\Sigma_{12}\Sigma_{22}^{-1} + \Sigma_{22}^{-1},\ \Sigma^{12} = -\Sigma^{11}\Sigma_{12}\Sigma_{22}^{-1}\,, \tag{3.51}$$

$$\Sigma^{21} = \Sigma^{12\prime} = -\Sigma_{22}^{-1}\Sigma_{21}\Sigma^{11}\,. \tag{3.52}$$

If Σ_V were known and we have CS's for subvectors a_1 and a_2 (in large-or finite-sample), we should be able to get the corresponding CS's for δ_1 and δ_2. In other words, we should be able to test the partial exogeneity of Y_1 or Y_2. The difficulty here is that Σ_V is unknown. Nevertheless, Our final aim here is to get identification-robust CS's for δ_1 or δ_2 from those of a_1 and a_2. As it will be showed, this is done by replacing Σ_V by its consistent estimate, for example $\hat{\Sigma}_V = MY$ in (3.48).

Without loss of generality, we focus on the subvector a_1. We consider again the partition of $\tilde{A}$, $\tilde{b}$ and Λ, b_* according to those of a in (3.47), *i.e.*

$$\tilde{A} = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ \tilde{A}_{21} & \tilde{A}_{22} \end{bmatrix},\ \tilde{b} = \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix},\ b_* = \begin{bmatrix} b_{*1} \\ b_{*2} \end{bmatrix}, \tag{3.53}$$

where Λ and b_* are defined in Theorem **3.2**. Since $\tilde{A}$ and Λ are symmetric, we have $\tilde{A}_{12}' = \tilde{A}_{21}$ and

$\Lambda'_{12} = \Lambda_{21}$. Hence, we can rewrite

$$
\begin{aligned}
\mathcal{Q}(a,\beta) \equiv \mathcal{Q}(a_1,a_2,\beta) &= a_1'\tilde{A}_{11}a_1 + (\tilde{b}_1 + 2\tilde{A}_{11}\beta_1 + 2\tilde{A}_{21}\beta_2)'a_1 + \tilde{c} \\
&+ a_2'\tilde{A}_{22}a_2 + (\tilde{b}_2 + 2\tilde{A}_{21}a_1 + 2\tilde{A}_{21}\beta_1 + 2\tilde{A}_{22}\beta_2)'a_2 + \\
&+ \beta'\tilde{A}\beta + \tilde{b}'\beta\,, \qquad (3.54) \\
f_\beta(a_1,a_2) \equiv a'\Lambda a + b_*'a + c_* &= a_1'\Lambda_{11}a_1 + b_{*1}'a_1 + c_* + a_2'\Lambda_{22}a_2 \\
&+ (2\Lambda_{21}a_1 + b_{*2})'a_2\,, \qquad (3.55)
\end{aligned}
$$

where $\mathcal{Q}(a,\beta)$ is defined in (3.37) and c_* in Theorem **3.2**. So, a confidence set for a_1 with level $1-\alpha$ is obtained by minimizing $\mathcal{Q}(a_1,a_2,\beta)$ over β and a_2 subject to $f(\beta) \le 0,$ *i.e.*,

$$
C_{a_1}(\alpha) = \left\{a_1 \in \mathbb{R}^{G_1} : \mathcal{Q}^*(a_1) \le 0\right\}\,, \qquad (3.56)
$$

where

$$
\mathcal{Q}^*(a_1) = \min_{\substack{a_2,\,\beta \\ s.c\ f(\beta)\le 0}} \mathcal{Q}(a_1,a_2,\beta)\,, \qquad (3.57)
$$

and $f(\beta) = \beta'A\beta + b'\beta + c$. Let $r_2 = \text{rank}(\tilde{A}_{22})$ and $p_2 = \text{rank}(\Lambda_{22}),$ where $0 \le r_2, p_2 \le G_2$. Consider the following spectral decomposition [see Dufour and Taamouti (2006) for more details]:

$$
\begin{aligned}
\tilde{A}_{22} &= P_2D_2P_2',\ D_2 = \text{diag}(d_1,\,\ldots,\,d_{G_2})\,, \qquad (3.58) \\
\Lambda_{22} &= O_2\Delta_2O_2',\ \Delta_2 = \text{diag}(\mu_1,\,\ldots,\,\mu_{G_2})\,, \qquad (3.59)
\end{aligned}
$$

$d_1,\,\ldots,\,d_{G_2}$ and $\mu_1,\,\ldots,\,\mu_{G_2}$ are eigenvalues of $\tilde{A}_{22}$ and $\Lambda_{22},$ P_2 and O_2 are orthogonal matrices. We assume that

$$
P_2 = [P_{21},\, P_{22}] \quad \text{and} \quad O_2 = [O_{21},\, O_{22}]\,. \qquad (3.60)
$$

Then, the form of the set $C_{a_1}(\alpha)$ defined in (3.56) is given by the following theorem.

Theorem 3.4 CLOSE-FORM CS'S FOR a_1. *Assume that the assumptions of Theorem 3.***2***, the sets* $C_{a_1}(\alpha)$ *take one of the following forms* :

(a) *if* $2A\beta + b = 0,$

$(a.1)$ *if* $\tilde{A}_{22}$ *is positive semidefinite,* $\tilde{A}_{22} \neq 0$, *then*

$$C_{a_1}(\alpha) = \begin{cases} \left\{a_1 : a_1'\tilde{A}_{11*}a_1 + \bar{b}_{1*}'a_1 + \bar{c}_{1*} \leq 0\right\} \cup \bar{\mathcal{S}}_1 & \textit{if} \quad c^* \leq 0 \\ \emptyset \quad \textit{otherwise} \end{cases} \tag{3.61}$$

where $\tilde{A}_{11*} = \tilde{A}_{11} - \tilde{A}_{21}'\tilde{A}_{22}^{-}\tilde{A}_{21},\ \bar{b}_{1*} = \bar{b}_1 - \tilde{A}_{21}'\tilde{A}_{22}^{-}\bar{b}_2,\ \bar{c}_{1*} = \bar{c} - \frac{\bar{b}_2'\tilde{A}_{22}^{-}\bar{b}_2}{4},\ \bar{b}_1 = \tilde{b}_1 + 2\tilde{A}_{11}\bar{\beta}_1 + 2\tilde{A}_{21}\bar{\beta}_2,\ \bar{c} = \tilde{c} + \bar{\beta}'\tilde{A}\bar{\beta} + \tilde{b}'\bar{\beta},$ *and*

$$\bar{\mathcal{S}}_1 = \begin{cases} \emptyset \quad \textit{if} \quad \textit{rank}(\tilde{A}_{22}) = G_2 \\ \{a_1 : P_{22}'(2\tilde{A}_{21}a_1 + \bar{b}_2) \neq 0\} \quad \textit{if} \quad 1 \leq \textit{rank}(\tilde{A}_{22}) < G_2; \end{cases}$$

$(a.2)$ *if* $\tilde{A}_{22} = 0,$ *then*

$$C_{a_1}(\alpha) = \begin{cases} \left\{a_1 : a_1'\tilde{A}_{11}a_1 + \bar{b}_1'a_1 + \bar{c} \leq 0\right\} \cup \bar{\mathcal{S}}_2 & \textit{if} \quad c^* \leq 0 \\ \emptyset \quad \textit{otherwise,} \end{cases} \tag{3.62}$$

where

$$\bar{\mathcal{S}}_2 = \{a_1 : 2\tilde{A}_{21}a_1 + \bar{b}_2 \neq 0\}; \tag{3.63}$$

$(a.3)$ *if* $\tilde{A}_{22}$ *is not positive semidefinite, then*

$$C_{a_1}(\alpha) = \begin{cases} \mathbb{R}^{G_1} & \textit{if} \quad c^* \leq 0 \\ \emptyset \quad \textit{otherwise.} \end{cases} \tag{3.64}$$

(b) *if* $2A\beta + b \neq 0$, *then*

$(b.1)$ *if* $\tilde{A} + \hat{\lambda}A \geq 0$ *and* $\hat{\lambda} > 0,$

$(b.1.1)$ *if* $\tilde{A}_{22}$ *is positive semidefinite,* $\tilde{A}_{22} \neq 0$, *then*

$$C_{a_1}(\alpha) = \left\{a_1 : a_1'\tilde{A}_{11*}a_1 + \tilde{b}_{1*}'a_1 + \tilde{c}_{\hat{\lambda}*} \leq 0,\ f(\bar{\beta}^*) = 0\right\} \cup \mathcal{S}_1 \tag{3.65}$$

where $\tilde{b}_{1*} = \tilde{b}_1 - \tilde{A}'_{21}\tilde{A}^{-}_{22}\tilde{b}_2,\ \tilde{c}_{\hat{\lambda}*} = \tilde{c}_{\hat{\lambda}} - \frac{\tilde{b}'_2\tilde{A}^{-}_{22}\tilde{b}_2}{4},$ *and*

$$\mathcal{S}_1 = \begin{cases} \emptyset \quad \textit{if} \quad \textit{rank}(\tilde{A}_{22}) = G_2 \\ \{a_1 : P'_{22}(2\tilde{A}_{21}a_1 + \tilde{b}_2) \neq 0\} \quad \textit{if} \quad 1 \leq \textit{rank}(\tilde{A}_{22}) < G_2; \end{cases}$$

$(b.1.2)$ *if* $\tilde{A}_{22} = 0,$ *then*

$$C_{a_1}(\alpha) = \Big\{a_1 : a'_1\tilde{A}_{11}a_1 + \tilde{b}'_1a_1 + \tilde{c}_{\hat{\lambda}} \leq 0,\ f(\bar{\beta}^*) = 0\Big\} \cup \mathcal{S}_2 \tag{3.66}$$

where

$$\mathcal{S}_2 = \{a_1 : 2\tilde{A}_{21}a_1 + \tilde{b}_2 \neq 0\}; \tag{3.67}$$

$(b.1.3)$ *if* $\tilde{A}_{22}$ *is not positive semidefinite, then*

$$C_{a_1}(\alpha) = \begin{cases} \mathbb{R}^{G_1} \quad \textit{if} \quad f(\bar{\beta}^*) \leq 0 \\ \emptyset \quad \textit{otherwise.} \end{cases} \tag{3.68}$$

$(b.2)$ *if* $\tilde{A} + \hat{\lambda}A \geq 0$ *and* $\hat{\lambda} = 0,$

$(b.2.1)$ *if* $\varLambda_{22}$ *is positive semidefinite,* $\varLambda_{22} \neq 0,$ *then*

$$\begin{aligned} C_{a_1}(\alpha) \quad &= \quad \Big\{a_1 : \tilde{b}'_{*10}a_1 + \tilde{c}_{*10} \leq 0\Big\} \cap \\ &\quad \big[\{a_1 : a'_1\varLambda_{11*}a_1 + b'_{*11}a_1 + c_{*11} \leq 0\} \cup \mathcal{S}_3\big] \end{aligned} \tag{3.69}$$

where $\varLambda_{11*} = \varLambda_{11} - \varLambda'_{21}\varLambda^{-}_{22}\varLambda_{21},\ b_{*11} = b_{*1} - \varLambda'_{21}\varLambda^{-}_{22}b_{*2},\ c_{*11} = c_* - \frac{\bar{b}'_{*2}\varLambda^{-}_{22}\bar{b}_{*2}}{4},$
$\tilde{b}_{*10} = \tilde{b}_{*1} - \tilde{b}_{*2}\varLambda^{-}_{22}\varLambda_{21},\ \tilde{c}_{*10} = \tilde{c}_* - \frac{\tilde{b}'_{*2}\varLambda^{-}_{22}b_{*2}}{2},$

$$\begin{aligned} \tilde{b}_{*10} \quad &= \quad \tilde{b}_{*1} - \tilde{b}_{*2}\varLambda^{-}_{22}\varLambda_{21},\ \tilde{c}_{*10} = \tilde{c}_* - \frac{\tilde{b}'_{*2}\varLambda^{-}_{22}b_{*2}}{2} \quad \textit{if} \\ \textit{rank}(\varLambda_{22}) \quad &= \quad G_2 \quad \textit{or} \quad [1 \leq \textit{rank}(\varLambda_{22}) < G_2 \quad \textit{and} \quad O'_{22}(2\varLambda_{21}a_1 + b_{*2}) = 0] \end{aligned}$$

$$\begin{aligned} \tilde{b}_{*10} \quad &= \quad \tilde{b}_{*1}, \quad \tilde{c}_{*10} = \tilde{c}_* + \tilde{b}_{*2}[O^{-'}_{22}\tilde{a}_2 + (I - \mathcal{H}_{O'_{22}})\tilde{a}_{20}] \quad \textit{if} \\ &\quad 1 \leq \textit{rank}(\varLambda_{22}) < G_2 \quad \textit{and} \quad O'_{22}(2\varLambda_{21}a_1 + b_{*2}) \neq 0 \end{aligned}$$

$\tilde{a}_2 = O_{22}' a_2$ is chosen such that $f_\beta(a_1, a_2) < 0,$ $\tilde{a}_{20}$ is any arbitrary $G_2 \times 1$ vector,

$$\mathcal{S}_3 = \begin{cases} \emptyset \quad \textit{if} \quad \textit{rank}(\Lambda_{22}) = G_2 \\ \{a_1 : O_{22}'(2\Lambda_{21}a_1 + b_{*2}) \neq 0\} \quad \textit{if} \quad 1 \leq \textit{rank}(\Lambda_{22}) < G_2; \end{cases}$$

$(b.2.2)$ *if $\Lambda_{22} = 0,$ then*

$$C_{a_1}(\alpha) = \left\{a_1 : \tilde{b}_{*1}' a_1 + \tilde{c}_{*20} \leq 0\right\} \cap [\{a_1 : a_1' \Lambda_{11} a_1 + b_{*1}' a_1 + c_* \leq 0\} \cup \mathcal{S}_4] \tag{3.70}$$

where $\tilde{c}_{*20} = \tilde{c}_* + \tilde{b}_{*2} a_2,$ a_2 is chosen such that $f_\beta(a_1, a_2) < 0,$

$$\mathcal{S}_4 = \{a_1 : 2\Lambda_{21} a_1 + b_{*2} \neq 0\}; \tag{3.71}$$

$(b.2.3)$ *if Λ_{22} is not positive semidefinite, then*

$$C_{a_1}(\alpha) = \left\{a_1 : \tilde{b}_{*1}' a_1 + \tilde{c}_* \leq 0\right\} \quad \textit{if} \quad \tilde{b}_{*2} = 0 \tag{3.72}$$

$$= \mathbb{R}^{G_1} \quad \textit{if} \quad \tilde{b}_{*2} \neq 0 \tag{3.73}$$

$(b.3)$ *if $\tilde{A} + \hat{\lambda} A$ is not positive semidefinite,*

$$C_{a_1}(\alpha) = \mathbb{R}^{G_1}, \tag{3.74}$$

where $\hat{\lambda}$ is defined in Theorem **3.2**.

The results of this Theorem **3.4** suggest that unlike the CS's C_a for the full vector a, C_{a_1} can be bounded even when $\tilde{A}$ and A are not positive definite. In other words, even if a is not identified, some components or subvectors of a may be identified. Remark also that CS's for a_2 can be obtained by the same way. In particular, Theorem **3.4** holds for each component of a, say $a_j,\ j = 1, \ldots, G.$

The next section establishes the asymptotic validity of the above procedure and proposed asymptotic identification-robust CS's for covariances.

4. Asymptotic theory

Consider again the model described by (2.1) - (2.8). Let

$$u = Va + \varepsilon\,, \tag{4.1}$$

where ε is uncorrelated with V with mean zero and covariance matrix

$$\Sigma_\varepsilon = [\sigma_{ij}]_{\,1\leq i,\, j\leq T}\,, \Sigma_\varepsilon > 0\,. \tag{4.2}$$

If $\Sigma_\varepsilon = \sigma_\varepsilon^2 I_T$, then, $\varepsilon_t,\ t = 1, \ldots, T$ are homoskedastic. However, if $\Sigma_\varepsilon =$ diag$[\sigma_{11}, \sigma_{22}, \ldots, \sigma_{TT}]$, $\varepsilon_t,\ t = 1, \ldots, T$ are heteroskedastic. Finally if

$$\Sigma_\varepsilon = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \gamma_\varepsilon(1) & \ldots & \gamma_\varepsilon(T-1) \\ \gamma_\varepsilon(1) & \ddots & \ldots & \vdots \\ \vdots & \ldots & \ddots & \vdots \\ \gamma_\varepsilon(T-1) & \ldots & \ldots & 1 \end{bmatrix}, \tag{4.3}$$

$\varepsilon_t,\ t = 1, \ldots, T$ are autocorrelated, where $\gamma_\varepsilon(.)$ is the autocovariance function of ε. If the elements of the diagonal of Σ_ε in (4.3) are different, then, ε_t are both heteroskedastic and autocorrelated.

We can see from (4.1) that the covariance matrix of u is given by

$$\Sigma_u = [\sigma_{ij}^u]_{\,1\leq i,\, j\leq T} > 0, \quad \sigma_{ij}^u = a'\Sigma_V a + \sigma_{ij}\,, \tag{4.4}$$

where Σ_V is the covariance matrix of V (it is assumed implicitly that V is homoskedastic and that Σ_V is identified).

We consider the following generic assumptions on the asymptotic behaviour of model variables [where $\mathcal{B} > 0$ for any matrix $\mathcal{B}$ means that $\mathcal{B}$ is positive definite (p.d.), and $\to$ refers to limits as $T \to \infty$]:

$$\frac{1}{T}X'\Sigma_u^{-1}u \xrightarrow{p} 0\,,\ \frac{1}{T}X'\Sigma_u^{-1}X \xrightarrow{p} \Omega_X > 0\,, \tag{4.5}$$

$$\frac{1}{T}Z'\Sigma_{\varepsilon}^{-1}\varepsilon \xrightarrow{p} 0\,,\ \frac{1}{T}Z'\Sigma_{\varepsilon}^{-1}Z \xrightarrow{p} \Omega_Z > 0\,, \tag{4.6}$$

$$\frac{1}{T}X'\Sigma_u^{-1}\Sigma_{\varepsilon}\Sigma_u^{-1}X \xrightarrow{p} \Delta_{X\varepsilon} > 0\,,\ \frac{1}{T}Z'\Sigma_{\varepsilon}^{-1}\Sigma_u\Sigma_{\varepsilon}^{-1}Z \xrightarrow{p} \Delta_{Zu} > 0\,, \tag{4.7}$$

$$\frac{1}{\sqrt{T}}X'\Sigma_u^{-1}[u,\, V,\, \varepsilon] \xrightarrow{L} [S_u^x,\, S_V^x,\, S_\varepsilon^x]\,, \tag{4.8}$$

$$\frac{1}{\sqrt{T}}Z'\Sigma_\varepsilon^{-1}[u,\, V,\, \varepsilon] \xrightarrow{L} [S_u^z,\, S_V^z,\, S_\varepsilon^z]\,, \tag{4.9}$$

$$\text{vec}[S_u^x,\, S_V^x,\, S_\varepsilon^x] \sim N\left[0\,,\, \Sigma_{S^x}\right]\,, \text{vec}[S_u^z,\, S_V^z,\, S_\varepsilon^z] \sim N\left[0\,,\, \Sigma_{S^z}\right]\,, \tag{4.10}$$

$$S_\varepsilon^x \ \text{ is independent with } \ S_V^x, S_\varepsilon^z, \ \text{ and } \ S_V^z \tag{4.11}$$

$$S_u^x \sim N\left[0,\, \Omega_X\right]\,, \quad S_u^z \sim N\left[0,\, \Delta_{Zu}\right], \tag{4.12}$$

$$S_\varepsilon^x \sim N\left[0,\, \Delta_{X\varepsilon}\right]\,, \quad S_\varepsilon^x \sim N\left[0,\, \Omega_Z\right]\,, \tag{4.13}$$

where $X = [X_1\,,\, X_2]$ and $Z = [Y\,,\, X_1\,,\, X_2]$.

Two problems are studied here. First, we extend our finite-sample procedure for building CS's for a and its subvectors. Second, we build CS's for δ as well as its subvectors.

4.1. Asymptotic CS's for the full vector of endogeneity parameters

This subsection extends our finite-sample projection-based techniques developed in Subsection 3.1 to asymptotic setup. As before, we consider the problem of building CS's for the full vector a as well as its subvectors.

Let us consider again equations (3.2) and (3.20). Since $\Sigma_u > 0$ and $\Sigma_\varepsilon > 0$, we can multiply the LHS and RHS of (3.2) by $\Sigma_u^{-1/2}$ and those of (3.20) by $\Sigma_\varepsilon^{-1/2}$ so that (3.2) and (3.20) becomes respectively

$$\Sigma_u^{-1/2}(y - Y\beta_0) \;\; = \;\; \Sigma_u^{-1/2}X_1\pi_1^0 + \Sigma_u^{-1/2}X_2\pi_2^0 + \Sigma_u^{-1/2}\upsilon^0\,, \tag{4.14}$$

$$\textit{i.e.} \quad y_* - Y_*\beta_0 \;=\; X_{*1}\pi_1^0 + X_{*2}\pi_2^0 + \upsilon_1 \tag{4.15}$$

and

$$\Sigma_\varepsilon^{-1/2}(y - Y\theta_0) \;=\; \Sigma_\varepsilon^{-1/2}Y\psi + \Sigma_\varepsilon^{-1/2}X_1\pi_1^* + \Sigma_\varepsilon^{-1/2}X_2\pi_2^* + \Sigma_\varepsilon^{-1/2}\varepsilon\,, \tag{4.16}$$

$$\textit{i.e.} \quad y^* - Y_*\theta_0 \;=\; Y^*\psi + X_1^*\pi_1^* + X_2^*\pi_2^* + \varepsilon_1\,, \tag{4.17}$$

where $\psi = \theta - \theta_0$, $\pi_1^* = \gamma - \Pi_1 a$, $\pi_2^* = -\Pi_2 a$.

So, the Anderson and Rubin (1949, AR) test-statistics for testing the hypotheses $H_{\pi_2} : \pi_2^0 = 0$ and $H_\psi : \psi = 0$ respectively in (4.15) and (4.17) are given by

$$AR^{he}(\beta_0) = \frac{\upsilon^0(\beta_0)'\hat{\Sigma}_u^{-1/2}[P_X(\hat{\Sigma}_u) - P_{X_1}(\hat{\Sigma}_u)]\hat{\Sigma}_u^{-1/2}\upsilon^0(\beta_0)/k_2}{\upsilon^0(\beta_0)'\hat{\Sigma}_u^{-1/2}M_X(\hat{\Sigma}_u)\hat{\Sigma}_u^{-1/2}\upsilon^0(\beta_0)/(T - k_1 - k_2)}\,, \tag{4.18}$$

and

$$AR^{he}(\theta_0) = \frac{\varepsilon(\theta_0)'\hat{\Sigma}_\varepsilon^{-1/2}[P_Z(\hat{\Sigma}_\varepsilon) - P_X(\hat{\Sigma}_\varepsilon)]\hat{\Sigma}_\varepsilon^{-1/2}\varepsilon(\theta_0)/G}{\varepsilon(\theta_0)'\hat{\Sigma}_\varepsilon^{-1/2}M_Z(\hat{\Sigma}_\varepsilon)\hat{\Sigma}_\varepsilon^{-1/2}\varepsilon(\theta_0)/(T - k - G)}\,, \tag{4.19}$$

where $\upsilon^0(\beta_0) = y - Y\beta_0,\ \varepsilon(\theta_0) = y - Y\theta_0,$ and for any fixed matrix Λ and any random matrix $\mathcal{Z}$,

$$P_{\mathcal{Z}}(\Lambda) = \Lambda^{-1/2}\mathcal{Z}(\mathcal{Z}'\Lambda^{-1}\mathcal{Z})^{-1}\mathcal{Z}'\Lambda^{-1/2},\ M_{\mathcal{Z}}(\Lambda) = I - P_{\mathcal{Z}}(\Lambda)\,. \tag{4.20}$$

Note that $P_{\mathcal{Z}}(\Lambda)$ is a projection matrix on the space spanned by the columns of $\Lambda^{-1/2}\mathcal{Z}$. Σ_u and Σ_ε can be estimate consistently by using the HAC estimators that have been proposed in the literature, *e.g.*, Levine (1983), White (1984, pp. 147-161), White and Domowitz (1984), Gallant (1987, pp. 533, 551, 573), Newey and West (1987), Andrews (1991), Andrews (1992). In this paper, Σ_u and Σ_ε are estimated as in Andrews (1991) and Andrews (1992).

Define

$$J_{\upsilon,T} \;=\; \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}E(\upsilon_s^0\upsilon_t^{0\prime}) = \sum_{j=-T+1}^{T-1}\gamma_\upsilon(j)\,, \tag{4.21}$$

$$\gamma_{\upsilon}(j) \;=\; \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} E(\upsilon_t^0 \upsilon_{t-j}^{0\prime}) & \text{for } j \geq 0 \\ \\ \frac{1}{T}\sum_{t=-j+1}^{T} E(\upsilon_{t+j}^0 \upsilon_t^{0\prime}) & \text{for } j < 0\,, \end{cases} \tag{4.22}$$

$$J_{\varepsilon\,,T} \;=\; \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} E(\varepsilon_s \varepsilon_t') = \sum_{j=-T+1}^{T-1} \gamma_{\varepsilon}(j)\,, \tag{4.23}$$

$$\gamma_{\varepsilon}(j) \;=\; \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} E(\varepsilon_t \varepsilon_{t-j}') & \text{for } j \geq 0 \\ \\ \frac{1}{T}\sum_{t=-j+1}^{T} E(\varepsilon_{t+j} \varepsilon_t') & \text{for } j < 0\,. \end{cases} \tag{4.24}$$

We consider the following class of estimators of $J_{\upsilon\,,T}$ and $J_{\varepsilon\,,T}$

$$\hat{J}_{\upsilon\,,T} \;=\; \hat{J}_{\upsilon\,,T}(S_T) = \frac{T}{T-r}\sum_{j=-T+1}^{T-1} \kappa\left(\frac{j}{S_T}\right)\hat{\gamma}_{\upsilon}(j)\,, \tag{4.25}$$

$$\hat{\gamma}_{\upsilon}(j) \;=\; \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \hat{\upsilon}_t^0 \hat{\upsilon}_{t-j}^{0\prime} & \text{for } j \geq 0 \\ \\ \frac{1}{T}\sum_{t=-j+1}^{T} \hat{\upsilon}_{t+j}^0 \hat{\upsilon}_t^{0\prime} & \text{for } j < 0 \end{cases} \tag{4.26}$$

$$\hat{J}_{\varepsilon\,,T} \;=\; \hat{J}_{\varepsilon\,,T}(\tilde{S}_T) = \frac{T}{T-r}\sum_{j=-T+1}^{T-1} \tilde{\kappa}\left(\frac{j}{\tilde{S}_T}\right)\hat{\gamma}_{\varepsilon}(j)\,, \tag{4.27}$$

$$\hat{\gamma}_{\varepsilon}(j) \;=\; \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}' & \text{for } j \geq 0 \\ \\ \frac{1}{T}\sum_{t=-j+1}^{T} \hat{\varepsilon}_{t+j} \hat{\varepsilon}_t' & \text{for } j < 0 \end{cases} \tag{4.28}$$

where $\hat{\upsilon}_t^0 \equiv \hat{\upsilon}_t^0(\hat{\pi}_1^0, \hat{\pi}_2^0)$, $\hat{\varepsilon}_t = \hat{\varepsilon}_t(\hat{\psi},\, \hat{\pi}_1^*, \hat{\pi}_2^*)$, $\kappa(.)$ and $\tilde{\kappa}(.)$ are real-valued kernel and $S_T,\ \tilde{S}_T,$ are band-with parameters [see Andrews (1991) and Andrews (1992)]. So,

$$\hat{J}_{\upsilon\,,T} = \hat{\Sigma}_u \quad \text{and} \quad \hat{J}_{\varepsilon\,,T} = \hat{\Sigma}_{\varepsilon}\,. \tag{4.29}$$

are consistent estimators of Σ_u and Σ_{ε}, ***i.e.***,

$$\underset{T\to\infty}{\text{plim}}\,(\hat{J}_{\upsilon\,,T}) = \underset{T\to\infty}{\text{plim}}\,(\hat{\Sigma}_u) = \Sigma_u \quad \text{and} \quad \underset{T\to\infty}{\text{plim}}\,(\hat{J}_{\varepsilon\,,T}) = \underset{T\to\infty}{\text{plim}}\,(\hat{\Sigma}_{\varepsilon}) = \Sigma_{\varepsilon}\,. \tag{4.30}$$

Lemma 4.1 ASYMPTOTIC DISTRIBUTIONS OF AR STATISTICS. *Under the assumptions* (2.1) - (2.8), (4.1) - (4.13), *if* $\beta = \beta_0$ *and* $\theta = \theta_0$ *where* β_0 *and* θ_0 *are* $G \times 1$ *constant vectors, then*

$$AR^{he}(\beta_0) \xrightarrow{L} \frac{1}{k_2}\chi^2(k_2)\,, \tag{4.31}$$

$$AR^{he}(\theta_0) \xrightarrow{L} \frac{1}{G}\chi^2(G)\,, \tag{4.32}$$

irrespective of whether the instruments are weak or strong.

Consequently, the asymptotic confidence sets for β and θ with level $1-\alpha_1$ and $1-\alpha_2$ are given by

$$C_\beta^\infty(\alpha_1) = \left\{\beta_0 : AR^{he}(\beta_0) \leq \frac{1}{k_2}\chi^2_{\alpha_1}(k_2)\right\}\,, \tag{4.33}$$

$$C_\theta^\infty(\alpha_2) = \left\{\theta_0 : AR^{he}(\theta_0) \leq \frac{1}{G}\chi^2_{\alpha_2}(G)\right\}\,, \tag{4.34}$$

where $\chi^2_{\alpha_1}(k_2)$ and $\chi^2_{\alpha_2}(G)$ are respectively the $1-\alpha_1$ and $1-\alpha_2$ quantiles of the χ^2 distributions with k_2 and G degrees of freedom. As in our finite-sample setup, $C_\beta^\infty(\alpha_1)$ and $C_\theta^\infty(\alpha_2)$ can be written as

$$C_\beta^\infty(\alpha_1) = \{\beta_0 : \beta_0' A_h \beta_0 + b_h' \beta_0 + c_h \leq 0\}\,, \tag{4.35}$$

$$C_\theta^\infty(\alpha_2) = \{\theta_0 : \theta_0' \tilde{A}_h \theta_0 + \tilde{b}_h' \theta_0 + \tilde{c}_h \leq 0\}\,, \tag{4.36}$$

where

$$\begin{aligned}
A_h &= Y'\hat{\Sigma}_u^{-1} H_\infty \hat{\Sigma}_u^{-1} Y,\ \tilde{A}_h = Y'\hat{\Sigma}_\varepsilon^{-1} \tilde{H}_\infty \hat{\Sigma}_\varepsilon^{-1} Y,\ b_h' = -2Y'\hat{\Sigma}_u^{-1} H_\infty \hat{\Sigma}_u^{-1} y,\\
\tilde{b}_h' &= -2Y'\hat{\Sigma}_\varepsilon^{-1} \tilde{H}_\infty \hat{\Sigma}_\varepsilon^{-1} y,\ c_h = y'\hat{\Sigma}_u^{-1} H_\infty \hat{\Sigma}_u^{-1} y,\ \tilde{c}_h = y'\hat{\Sigma}_\varepsilon^{-1} \tilde{H}_\infty \hat{\Sigma}_\varepsilon^{-1} y,\\
H_\infty &= M_{X_1}(\hat{\Sigma}_u) - [1 + \frac{1}{T-k}\chi^2_{\alpha_1}(k_2)] M_X(\hat{\Sigma}_u),\\
\tilde{H}_\infty &= M_X(\hat{\Sigma}_\varepsilon) - [1 + \frac{1}{T-G-k}\chi^2_{\alpha_2}(G)] M_Z(\hat{\Sigma}_\varepsilon)\,.,
\end{aligned} \tag{4.37}$$

By applying Bonferroni inequality, the set

$$C^{\infty}_{(\beta,\theta)}(\alpha) = \{(\beta_0', a_0')' : \beta_0' A_h \beta_0 + b_h' \beta_0 + c_h \leq 0 \ , \ \theta_0' \tilde{A}_h \theta_0 + \tilde{b}_h' \theta_0 + \tilde{c}_h \leq 0\} \tag{4.38}$$

is an asymptotic confidence set with level $1-\alpha$ for the joint parameter(β, θ), where $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$. So,

$$C^{\infty}_{a}(\alpha) = \{a \in \mathbb{R}^G : (\beta', a')' \in C^{\infty}_{(\beta,\theta)}(\alpha)\} \, . \tag{4.39}$$

is also an asymptotic CS for a with level $1-\alpha$. Let us define

$$\mathcal{Q}_h(\beta, a) = \theta' \tilde{A}_h \theta + \tilde{b}_h' \theta + \tilde{c}_h = a' \tilde{A}_h a + \tilde{b}_h' a + \tilde{c}_h + \beta' \tilde{A}_h \beta + (2\tilde{A}_h a + \tilde{b}_h)' \beta \, , \tag{4.40}$$

$$f_h(\beta) = \beta' A_h \beta + b_h' \beta + c_h \, . \tag{4.41}$$

Then, we have

$$C_a(\alpha) = \{a : \ \mathcal{Q}^*_h(a) \leq 0\} \ , \tag{4.42}$$

where

$$\mathcal{Q}^*_h(a) = \min_{\substack{\beta \\ s.c \ f_h(\beta) \leq 0}} \mathcal{Q}_h(\beta, a) \, . \tag{4.43}$$

As in finite-sample setup, the forms of $C^{\infty}_{a}(\alpha)$ is given by Theorem **4.2** below.

Theorem 4.2 ASYMPTOTIC CS'S FOR ENDOGENEITY PARAMETER. *Assume that the assumptions* (2.1) - (2.8), (4.1) - (4.13) *hold. If furthermore* $\beta = \beta_0$ *and* $a = a_0$, *where* β_0 *and* a_0 *are* $G \times 1$ *constant vectors, then the sets* $C_a^{\infty}(\alpha)$ *take one of the following forms* :

(a) *if* $2A_h\beta + b_h = 0$, *then*

$$C_a^{\infty}(\alpha) = \begin{cases} \left\{a : a'\tilde{A}_h a + \bar{b}_h' a + \bar{c}_h \leq 0\right\} & \text{if} \quad c_h^* \leq 0 \\ \emptyset \quad \text{otherwise} \ ; \end{cases} \tag{4.44}$$

(b) *if* $2A_h\beta + b_h \neq 0$, *then,*

$(b.1)$ *if* $\tilde{A}_h + \hat{\lambda}A_h \geq 0$ *and* $\hat{\lambda} > 0$,

$$C_a^{\infty}(\alpha) = \left\{a : a'\tilde{A}_h a + \tilde{b}_h' a + \tilde{c}_{h\hat{\lambda}} \leq 0, \quad \bar{\beta}^* \text{ is such that} \quad f_h(\bar{\beta}^*) = 0\right\} \ ; \tag{4.45}$$

$(b.2)$ *if* $\tilde{A}_h + \hat{\lambda}A_h \geq 0$ *and* $\hat{\lambda} = 0$,

$$C_a^{\infty}(\alpha) = \left\{a : \tilde{b}_{h*}' a + \tilde{c}_{h*} \leq 0\right\} \cap \{a'\Lambda_h a + b_{h*}' a + c_{h*} \leq 0\} \, ; \tag{4.46}$$

$(b.3)$ *if* $\tilde{A}_h + \hat{\lambda}A_h$ *is not positive semidefinite,*

$$C_a^{\infty}(\alpha) = \mathbb{R}^G \, , \tag{4.47}$$

where $\bar{b}_h,\ \bar{c}_h,\ c_h^*,\ \hat{\lambda},\ \tilde{c}_{h\hat{\lambda}}\ \tilde{b}_{h*},\ \tilde{c}_{h*},\ \Lambda_h,\ b_{h*},$ *and* c_{h*} *are defined as in Theorem* $3.\mathbf{2}$ *by replacing* $A,\ \tilde{A},\ \tilde{b},\ b,\ \tilde{c},\ c,$ *by* $A_h,\ \tilde{A}_h,\ \tilde{b}_h,\ b_h,\ \tilde{c}_h,\ c_h$ *defined in* (4.37).

The above theorem shows the asymptotic validity of our procedure allowing the possibility of heteroskedasticity and/or autocorrelation of errors. Hence, the projection-based procedure is asymptotically robust to those problems, whether the instruments are strong or weak.

Furthermore, as in the finite-sample setup, $C_a^{\infty}(\alpha)$ is bounded if and only if $\tilde{A}_h$ and A_h are positive definite. Otherwise, $C_a^{\infty}(\alpha)$ is bounded only when it is empty.

We now focus on building asymptotic CS's for subvectors.

4.2. Asymptotic CS's for subvectors

As in Section 3.2, we consider again the problem of building CS's for the subvector a_1. We adopt the same notations as in Section 3.2 by just replacing $A,\ \tilde{A},\ \tilde{b},\ b,\ \tilde{c},\ c,$ by $A_h,\ \tilde{A}_h,\ \tilde{b}_h,\ b_h,\ \tilde{c}_h,\ c_h$ defined in (4.37).

Then, the form of the set $C^{\infty}_{a_1}(\alpha)$ is given by the following theorem [similar to Theorem **3.4**]:

Theorem 4.3 ASYMPTOTIC CS'S FOR SUBVECTOR OF ENDOGENEITY PARAMETER. *Assume that the assumptions* (2.1) - (2.8), (4.1) - (4.13) *hold. If furthermore* $\beta = \beta_0$ *and* $a = a_0$, *where* β_0 *and* a_0 *are* $G \times 1$ *constant vectors, then the sets* $C^{\infty}_{a_1}(\alpha)$ *take one of the following forms* :

(a) *if* $2A_h\beta + b_h = 0,$

$(a.1)$ *if* $\tilde{A}_{h22}$ *is positive semidefinite,* $\tilde{A}_{h22} \neq 0$, *then*

$$C^{\infty}_{a_1}(\alpha) = \begin{cases} \left\{a_1 : a_1'\tilde{A}_{h11*}a_1 + \bar{b}_{h1*}'a_1 + \bar{c}_{h1*} \leq 0\right\} \cup \bar{\mathcal{S}}_{h1} & \text{if} \quad c_h^* \leq 0 \\ \emptyset \quad \text{otherwise} & ; \end{cases} \tag{4.48}$$

$(a.2)$ *if* $\tilde{A}_{h22} = 0,$ *then*

$$C^{\infty}_{a_1}(\alpha) = \begin{cases} \left\{a_1 : a_1'\tilde{A}_{h11}a_1 + \bar{b}_{h1}'a_1 + \bar{c}_{h} \leq 0\right\} \cup \bar{\mathcal{S}}_{h2} & \text{if} \quad c_h^* \leq 0 \\ \emptyset \quad \text{otherwise,} & ; \end{cases} \tag{4.49}$$

$(a.3)$ *if* $\tilde{A}_{h22}$ *is not positive semidefinite, then*

$$C^{\infty}_{a_1}(\alpha) = \begin{cases} \mathbb{R}^{G_1} & \text{if} \quad c_h^* \leq 0 \\ \emptyset & \text{otherwise.} \end{cases} \tag{4.50}$$

(b) *if* $2A_h\beta + b_h \neq 0$, *then*

$(b.1)$ *if* $\tilde{A}_h + \hat{\lambda}A_h \geq 0$ *and* $\hat{\lambda} > 0,$

$(b.1.1)$ *if* $\tilde{A}_{h22}$ *is positive semidefinite,* $\tilde{A}_{h22} \neq 0$, *then*

$$C_{a_1}^{\infty}(\alpha) = \Big\{ a_1 : a_1' \tilde{A}_{h11*} a_1 + \tilde{b}_{h1*}' a_1 + \tilde{c}_{h\hat{\lambda}*} \leq 0,\ f_h(\bar{\beta}^*) = 0 \Big\} \cup \mathcal{S}_{h1} \quad (4.51)$$

$(b.1.2)$ *if* $\tilde{A}_{h22} = 0,$ *then*

$$C_{a_1}^{\infty}(\alpha) = \Big\{ a_1 : a_1' \tilde{A}_{h11} a_1 + \tilde{b}_{h1}' a_1 + \tilde{c}_{h\hat{\lambda}} \leq 0,\ f_h(\bar{\beta}^*) = 0 \Big\} \cup \mathcal{S}_{h2} \quad (4.52)$$

$(b.1.3)$ *if* $\tilde{A}_{h22}$ *is not positive semidefinite, then*

$$C_{a_1}^{\infty}(\alpha) = \begin{cases} \mathbb{R}^{G_1} & \text{if} \quad f_h(\bar{\beta}^*) \leq 0 \\ \emptyset & \text{otherwise.} \end{cases} \quad (4.53)$$

$(b.2)$ *if* $\tilde{A}_h + \hat{\lambda} A_h \geq 0$ *and* $\hat{\lambda} = 0,$

$(b.2.1)$ *if* $\mathit{\Lambda}_{h22}$ *is positive semidefinite,* $\mathit{\Lambda}_{h22} \neq 0$, *then*

$$\begin{aligned} C_{a_1}^{\infty}(\alpha) \quad = \quad & \Big\{ a_1 : \tilde{b}_{h*10}' a_1 + \tilde{c}_{h*10} \leq 0 \Big\} \cap \\ & \big[\{ a_1 : a_1' \mathit{\Lambda}_{h11*} a_1 + b_{h*11}' a_1 + c_{h*11} \leq 0 \} \cup \mathcal{S}_{h3} \big] \quad (4.54) \end{aligned}$$

$(b.2.2)$ *if* $\mathit{\Lambda}_{h22} = 0,$ *then*

$$\begin{aligned} C_{a_1}^{\infty}(\alpha) \quad = \quad & \Big\{ a_1 : \tilde{b}_{h*1}' a_1 + \tilde{c}_{h*20} \leq 0 \Big\} \cap \\ & [\{ a_1 : a_1' \mathit{\Lambda}_{h11} a_1 + b_{h*1}' a_1 + c_{h*} \leq 0 \} \cup \mathcal{S}_{h4}] \quad (4.55) \end{aligned}$$

$(b.2.3)$ *if* $\mathit{\Lambda}_{h22}$ *is not positive semidefinite, then*

$$\begin{aligned} C_{a_1}^{\infty}(\alpha) \quad &= \quad \Big\{ a_1 : \tilde{b}_{h*1}' a_1 + \tilde{c}_{h*} \leq 0 \Big\} \quad \text{if} \quad \tilde{b}_{h*2} = 0 \quad (4.56) \\ &= \quad \mathbb{R}^{G_1} \quad \text{if} \quad \tilde{b}_{h*2} \neq 0 \quad (4.57) \end{aligned}$$

$(b.3)$ *if* $\tilde{A}_h + \hat{\lambda} A_h$ *is not positive semidefinite,*

$$C_{a_1}^{\infty}(\alpha) = \mathbb{R}^{G_1}\, , \quad (4.58)$$

where $\tilde{A}_{h11*},\ \bar{b}_{h1*},\ \bar{c}_{h1*},\ \bar{b}_{h1},\ c_h^*,\ \bar{c}_h,\ \bar{\mathcal{S}}_{h1},\ \hat{\lambda},\ \tilde{b}_{h1*},\ \tilde{c}_{h\hat{\lambda}*},\ \mathcal{S}_{h1},\ \mathcal{S}_{h2},\ \Lambda_{h11*}, b_{h*11}, c_{h*11}, \tilde{b}_{h*10},$ $\tilde{c}_{h*10}, \tilde{b}_{h*10}, \tilde{c}_{h*10},\ \mathcal{S}_{h3},\ \tilde{c}_{*20},$ *and* $\mathcal{S}_{h4}$ *are defined as in Theorem 3.***4**.

Remark that $C_{a_1}^{\infty}(\alpha)$ in the above theorem may be bounded even if $\tilde{A}_h$ and A_h are not both positive definite. Furthermore, Theorem 4.**3** can be applied (as in our finite-sample setup) to get CS's for each component $a_j,\ j = 1,\ \ldots,\ G$.

We now focus on CS's construction for covariances in the next section.

5. Asymptotic confidence sets for covariances

This subsection considers the problem of building identification-robust CS's for covariances. This problem is important because as mentioned before, DWH-type tests are not usable for testing linear restrictions of the form

$$H_{d_0} : \Gamma\delta = d_0\,, \tag{5.59}$$

where $d_0 \neq 0$. The Wald-type procedures [Dufour (1987), Doko and Dufour (2009c)] which deal with such hypotheses assume that the available IV are strong. In general, the Wald-type procedures is size distorted when IV are weak. The goal of this subjection is to provide identification-robust CS's (*i.e.* valid CS's even if IV are weak) for covariances and subvectors of covariances.

We now focus on the full covariance.

5.1. Asymptotic CS's for the full covariance

As in the finite-sample section, from (4.1), we have

$$\delta = \Sigma_V a\,, \tag{5.60}$$

since V is uncorrelated with ε. Hence, $a = \Sigma_V^{-1}\delta$. Let $\tilde{\Sigma}_V$ be any consistent estimator of Σ_V, *i.e.*,

$$\operatorname*{plim}_{T\to\infty} \tilde{\Sigma}_V = \Sigma_V\,. \tag{5.61}$$

If we replace a by $\tilde{\Sigma}_V^{-1}\delta$ in Theorem **4.2**, we get an asymptotic confidence set with level $1-\alpha$ for δ. Furthermore, this confidence set is valid even in presence of weak instruments. In particular, because V is identified, we can choose

$$\tilde{\Sigma}_V = \hat{\Sigma}_V = Y'MY/T-k\,, \tag{5.62}$$

where M is defined in (2.12). Note that $\tilde{\Sigma}_V$ is the first stage OLS estimate of Σ_V. Theorem **5.1** below gives the form of the CS's $C_\delta^\infty(\alpha)$ for δ.

Theorem 5.1 ASYMPTOTIC CS'S FOR THE FULL COVARIANCE. *Assume that the assumptions* (2.1) - (2.8), (4.1) - (4.13) *hold. If furthermore* $\beta=\beta_0$ *and* $a=a_0$, *where* β_0 *and* a_0 *are* $G\times 1$ *constant vectors, then the sets* $C_\delta^\infty(\alpha)$ *take one of the following forms* :

(a) *if* $2A_h\beta+b_h=0$, *then*

$$C_\delta^\infty(\alpha)=\begin{cases}\left\{\delta : \delta'\hat{\Sigma}_V^{-1}\tilde{A}_h\hat{\Sigma}_V^{-1}\delta+\bar{b}_h'\hat{\Sigma}_V^{-1}\delta+\bar{c}_h\le 0\right\} & \textit{if} \quad c_h^*\le 0\\ \emptyset \quad \textit{otherwise}\,; & \end{cases} \tag{5.63}$$

(b) *if* $2A_h\beta+b_h\neq 0$, *then,*

$(b.1)$ *if* $\tilde{A}_h+\hat{\lambda}A_h\ge 0$ *and* $\hat{\lambda}>0$,

$$C_\delta^\infty(\alpha)=\left\{\delta : \delta'\hat{\Sigma}_V^{-1}\tilde{A}_h\hat{\Sigma}_V^{-1}\delta+\tilde{b}_h'\hat{\Sigma}_V^{-1}\delta+\tilde{c}_{h\hat{\lambda}}\le 0, \quad \bar{\beta}^*: \quad f_h(\bar{\beta}^*)=0\right\}\,; \tag{5.64}$$

$(b.2)$ *if* $\tilde{A}_h+\hat{\lambda}A_h\ge 0$ *and* $\hat{\lambda}=0$,

$$C_\delta^\infty(\alpha) \;=\; \left\{\delta : \tilde{b}_{h*}'\hat{\Sigma}_V^{-1}\delta+\tilde{c}_{h*}\le 0\right\}\cap$$

$$\{\delta : \delta' \hat{\Sigma}_V^{-1} \Lambda_h \hat{\Sigma}_V^{-1} \delta + b'_{h*} \hat{\Sigma}_V^{-1} \delta + c_{h*} \leq 0\}\,; \tag{5.65}$$

$(b.3)$ *if* $\tilde{A}_h + \hat{\lambda} A_h$ *is not positive semidefinite,*

$$C_\delta^\infty(\alpha) = \mathbb{R}^G\,, \tag{5.66}$$

where $\bar{b}_h,\ \bar{c}_h,\ c_h^*,\ \hat{\lambda},\ \tilde{c}_{h\hat{\lambda}}\ \tilde{b}_{h*},\ \tilde{c}_{h*},\ \Lambda_h,\ b_{h*},\ c_{h*}$ *are defined in Theorem 4.*$\mathbf{2}$ *and* $\hat{\Sigma}_V$ *in* (5.62).

We note that like in Theorem 4.$\mathbf{2}$ $C_\delta^\infty(\alpha)$ is bounded if and only if both $\tilde{A}_h$ and A_h are positive definite. Otherwise, $C_\delta^\infty(\alpha)$ is bounded only when it is empty. Furthermore, $C_\delta^\infty(\alpha)$ has level $1-\alpha$ even in presence of weak IV.

We now derive CS's for subvectors.

5.2. Asymptotic CS's for subvectors of covariance

We consider the partition as in Section 3.2 where we replace $A,\ \tilde{A},\ \tilde{b},\ b,\ \tilde{c},\ c,$ by $A_h,\ \tilde{A}_h,\ \tilde{b}_h,\ b_h,$ $\tilde{c}_h,$ and c_h. From (3.47) - (3.53), (3.54) - (3.55) becomes

$$\begin{aligned}
\mathcal{Q}_h(\delta_1, \delta_2, \beta) &= \mathcal{Q}(\delta_1, \delta_2, \beta) = \delta_1' \hat{\Sigma}^{11} \tilde{A}_{h11} \hat{\Sigma}^{11} \delta_1 + 2\delta_1' \hat{\Sigma}^{11} \tilde{A}_{h11} \hat{\Sigma}^{12} \delta_2 \\
&+ \delta_2' \hat{\Sigma}^{21} \tilde{A}_{h11} \hat{\Sigma}^{12} \delta_2 + (\tilde{b}_{h1} + 2\tilde{A}_{h11}\beta_1 + 2\tilde{A}_{h21}\beta_2)' \hat{\Sigma}^{11} \delta_1 \\
&+ (\tilde{b}_{h1} + 2\tilde{A}_{h11}\beta_1 + 2\tilde{A}_{h21}\beta_2)' \hat{\Sigma}^{12} \delta_2 + \tilde{c}_h \\
&+ \delta_1' \hat{\Sigma}^{12} \tilde{A}_{h22} \hat{\Sigma}^{21} \delta_1 + 2\delta_1' \hat{\Sigma}^{12} \tilde{A}_{h22} \hat{\Sigma}^{22} \delta_2 + \delta_2' \hat{\Sigma}^{22} \tilde{A}_{h22} \hat{\Sigma}^{22} \delta_2 \\
&+ (\tilde{b}_{h2} + 2\tilde{A}_{h21} \hat{\Sigma}^{11} \delta_1 + 2\tilde{A}_{h21}\beta_1 + 2\tilde{A}_{h22}\beta_2)' \hat{\Sigma}^{22} \delta_2 + \\
&+ 2\delta_1' \hat{\Sigma}^{11} \tilde{A}_{h12} \hat{\Sigma}^{21} \delta_1 + (\tilde{b}_{h2} + 2\tilde{A}_{h21}\beta_1 + 2\tilde{A}_{h22}\beta_2)' \hat{\Sigma}^{21} \delta_1 \\
&+ 2\delta_2' \hat{\Sigma}^{21} \tilde{A}_{h12} \hat{\Sigma}^{21} \delta_1 + 2\delta_2' \hat{\Sigma}^{21} \tilde{A}_{h12} \hat{\Sigma}^{22} \delta_2 \\
&+ \beta' \tilde{A}_h \beta + \tilde{b}_h' \beta \\
&= \delta_1' \tilde{A}_{h11}^* \delta_1 + \tilde{b}_{h1}^{*\prime} \delta_1 + \tilde{c}_h + \beta' \tilde{A}_h \beta + \tilde{b}_h' \beta
\end{aligned}$$

$$
\begin{aligned}
& + \quad \delta_2' \tilde{A}^*_{h22}\delta_2 + (\tilde{b}^*_{h2} + 2\tilde{A}^*_{h21}\delta_1)'\delta_2 && (5.67)\\
f_h(\delta_1,\delta_2) &= \delta_1' \Lambda^*_{h11}\delta_1 + \kappa_h^{*'}\delta_1 + c_{h*} + \delta_2'\Lambda^*_{h22}\delta_2 \\
& + \quad (2\Lambda^*_{h21}\delta_1 + \kappa^*_{h2})'\delta_2\,, && (5.68)
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{A}^*_{h11} &= \hat{\Sigma}^{11}\tilde{A}_{h11}\hat{\Sigma}^{11} + \hat{\Sigma}^{12}\tilde{A}_{h22}\hat{\Sigma}^{21} + 2\hat{\Sigma}^{11}\tilde{A}_{h12}\hat{\Sigma}^{21}\,,\\
\tilde{b}^*_{h1} &= \hat{\Sigma}^{11}(\tilde{b}_{h1} + 2\tilde{A}_{h11}\beta_1 + 2\tilde{A}_{h21}\beta_2) + \hat{\Sigma}^{12}(\tilde{b}_{h2} + 2\tilde{A}_{h21}\beta_1 + 2\tilde{A}_{h22}\beta_2)\,,\\
\tilde{A}^*_{h22} &= \hat{\Sigma}^{21}\tilde{A}_{h11}\hat{\Sigma}^{12} + \hat{\Sigma}^{22}\tilde{A}_{h22}\hat{\Sigma}^{22} + 2\hat{\Sigma}^{21}\tilde{A}_{h12}\hat{\Sigma}^{22}\,,\\
\tilde{b}^*_{h2} &= \hat{\Sigma}^{21}(\tilde{b}_{h1} + 2\tilde{A}_{h11}\beta_1 + 2\tilde{A}_{h21}\beta_2) + \hat{\Sigma}^{22}(\tilde{b}_{h2} + 2\tilde{A}_{h21}\beta_1 + 2\tilde{A}_{h22}\beta_2) +\\
\tilde{A}^*_{h21} &= \hat{\Sigma}^{21}\tilde{A}_{h11}\hat{\Sigma}^{11} + \hat{\Sigma}^{12}\tilde{A}_{h22}\hat{\Sigma}^{22} + \hat{\Sigma}^{22}\tilde{A}_{h21}\hat{\Sigma}^{11} + \hat{\Sigma}^{21}\tilde{A}_{h12}\hat{\Sigma}^{21}\,, && (5.69)
\end{aligned}
$$

$$
\begin{aligned}
\Lambda^*_{h11} &= \hat{\Sigma}^{11}\Lambda_{h11}\hat{\Sigma}^{11} + \hat{\Sigma}^{12}\Lambda_{h22}\hat{\Sigma}^{21} + 2\hat{\Sigma}^{11}\Lambda_{h12}\hat{\Sigma}^{21}\,,\\
\kappa^*_h &= \hat{\Sigma}^{11}b_{h*1}, \quad \Lambda^*_{h22} = \hat{\Sigma}^{21}\Lambda_{h11}\hat{\Sigma}^{12} + \hat{\Sigma}^{22}\Lambda_{h22}\hat{\Sigma}^{22} + 2\hat{\Sigma}^{21}\Lambda_{h12}\hat{\Sigma}^{22}\,,\\
\Lambda^*_{h21} &= \hat{\Sigma}^{21}\Lambda_{h11}\hat{\Sigma}^{11} + \hat{\Sigma}^{22}\Lambda_{h22}\hat{\Sigma}^{21} + 2\hat{\Sigma}^{22}\Lambda_{h21}\hat{\Sigma}^{11} + \hat{\Sigma}^{12}\Lambda_{h21}\hat{\Sigma}^{12}\,,\\
\kappa^*_{h2} &= \hat{\Sigma}^{21}b_{h*2} + \hat{\Sigma}^{22}b_{h*2}, \quad \hat{\Sigma}_V^{-1} = \begin{bmatrix} \hat{\Sigma}^{11} & \hat{\Sigma}^{12} \\ \hat{\Sigma}^{21} & \hat{\Sigma}^{22} \end{bmatrix}. && (5.70)
\end{aligned}
$$

So, the asymptotic projection-based confidence set with level $1-\alpha$ for δ_1 is defined by

$$C^{\infty}_{\delta_1}(\alpha) = \{\delta_1 :\ \mathcal{Q}_h(\delta_1) \le 0\}\ , \tag{5.71}$$

where

$$\mathcal{Q}_h(\delta_1) = \min_{\substack{\beta,\delta_2 \\ s.c\ f_h(\delta_1,\delta_2)\le 0}} \mathcal{Q}_h(\beta,\delta_1,\delta_2)\,. \tag{5.72}$$

More precisely, the form of $C^{\infty}_{\delta_1}(\alpha)$ is given by the following theorem.

Theorem 5.2 ASYMPTOTIC CS'S FOR SUBVECTORS OF COVARIANCE. *Assume that the as-*

sumptions (2.1) - (2.8), (4.1) - (4.13) *hold. If furthermore* $\beta = \beta_0$ *and* $a = a_0$*, where* β_0 *and* a_0 *are* $G \times 1$ *constant vectors, then the sets* $C^{\infty}_{\delta_1}(\alpha)$ *take one of the following forms* :

(a) *if* $2A_h\beta + b_h = 0,$

$(a.1)$ *if* $\tilde{A}^*_{h22}$ *is positive semidefinite,* $\tilde{A}^*_{h22} \neq 0$*, then*

$$C^{\infty}_{a_1}(\alpha) = \begin{cases} \left\{\delta_1 : \delta_1' \tilde{A}^*_{h11*}\delta_1 + \bar{b}^{*\prime}_{h1*}\delta_1 + \bar{c}^*_{h1*} \leq 0\right\} \cup \bar{\mathcal{S}}^*_{h1} & \textit{if} \quad c^*_h \leq 0 \\ \emptyset \quad \textit{otherwise} & ; \end{cases} \tag{5.73}$$

where $\tilde{A}^*_{11*} = \tilde{A}^*_{h11} - \tilde{A}^{*\prime}_{h21}\tilde{A}^{-*}_{h22}\tilde{A}^*_{h21}$, $\bar{b}^*_{h1*} = \bar{b}^*_{h1} - \tilde{A}^{*\prime}_{h21}\tilde{A}^{-*}_{h22}\bar{b}^*_{h2}$, $\bar{c}^*_{h1*} = \bar{c}_h - \frac{\bar{b}^{*\prime}_{h2}\tilde{A}^{-*}_{h22}\bar{b}^*_{h2}}{4}$, $\bar{c}^*_{h1*} = \tilde{c}_h + \bar{\beta}'\tilde{A}_h\bar{\beta} + \tilde{b}'_h\bar{\beta}$, *and*

$$\bar{\mathcal{S}}^*_{h1} = \begin{cases} \emptyset \quad \textit{if} \quad \textit{rank}(\tilde{A}^*_{h22}) = G_2 \\ \{\delta_1 : P^{*\prime}_{h22}(2\tilde{A}^*_{h21}\delta_1 + \bar{b}^*_{h2}) \neq 0\} \quad \textit{if} \quad 1 \leq \textit{rank}(\tilde{A}^*_{h22}) < G_2, \end{cases}$$

P^*_{h22} *is defined as in Theorems* 3.**4** - 4.**3**;

$(a.2)$ *if* $\tilde{A}^*_{h22} = 0,$ *then*

$$C^{\infty}_{\delta_1}(\alpha) = \begin{cases} \left\{\delta_1 : \delta_1' \tilde{A}^*_{h11}\delta_1 + \bar{b}^{*\prime}_{h1}\delta_1 + \bar{c}^*_{h} \leq 0\right\} \cup \bar{\mathcal{S}}^*_{h2} & \textit{if} \quad c^*_h \leq 0 \\ \emptyset \quad \textit{otherwise} & ; \end{cases} \tag{5.74}$$

where

$$\bar{\mathcal{S}}^*_{h2} = \left\{\{\delta_1 : 2\tilde{A}^*_{h21}\delta_1 + \bar{b}^*_{h2} \neq 0\}\right.; \tag{5.75}$$

$(a.3)$ *if* $\tilde{A}^*_{h22}$ *is not positive semidefinite, then*

$$C^{\infty}_{\delta_1}(\alpha) = \begin{cases} \mathbb{R}^{G_1} & \textit{if} \quad c^*_h \leq 0 \\ \emptyset & \textit{otherwise.} \end{cases} \tag{5.76}$$

(b) *if* $2A_h\beta + b_h \neq 0$*, then*

$(b.1)$ *if* $\tilde{A}_h + \hat{\lambda}A_h \geq 0$ *and* $\hat{\lambda} > 0,$

$(b.1.1)$ *if* $\tilde{A}^*_{h22}$ *is positive semidefinite,* $\tilde{A}^*_{h22} \neq 0$, *then*

$$C^\infty_{\delta_1}(\alpha) = \left\{\delta_1 : \delta_1' \tilde{A}^*_{h11*}\delta_1 + \tilde{b}^{*\prime}_{h1*}\delta_1 + \tilde{c}^*_{h\hat{\lambda}*} \leq 0,\ f_h(\bar{\beta}^*) = 0\right\} \cup \mathcal{S}^*_{h1}\,; \quad (5.77)$$

$(b.1.2)$ *if* $\tilde{A}^*_{h22} = 0,$ *then*

$$C^\infty_{\delta_1}(\alpha) = \left\{\delta_1 : \delta_1' \tilde{A}^*_{h11}\delta_1 + \tilde{b}^{*\prime}_{h1}\delta_1 + \tilde{c}^*_{h\hat{\lambda}} \leq 0,\ f_h(\bar{\beta}^*) = 0\right\} \cup \mathcal{S}^*_{h2}\,, \quad (5.78)$$

where $\tilde{c}^*_{h\hat{\lambda}*},\ \mathcal{S}^*_{h1},\ \tilde{c}^*_{h\hat{\lambda}},\ \mathcal{S}^*_{h2}$ *are defined as in Theorems* 3.**4** - 4.**3**;

$(b.1.3)$ *if* $\tilde{A}^*_{h22}$ *is not positive semidefinite, then*

$$C^\infty_{\delta_1}(\alpha) = \begin{cases} \mathbb{R}^{G_1} & \text{if} \quad f_h(\bar{\beta}^*) \leq 0 \\ \emptyset & \text{otherwise}\,; \end{cases} \quad (5.79)$$

$(b.2)$ *if* $\tilde{A}_h + \hat{\lambda}A_h \geq 0$ *and* $\hat{\lambda} = 0,$

$(b.2.1)$ *if* Λ^*_{h22} *is positive semidefinite,* $\Lambda^*_{h22} \neq 0$, *then*

$$\begin{aligned} C^\infty_{\delta_1}(\alpha) &= \left\{\delta_1 : \tilde{b}^{*\prime}_{h*10}\delta_1 + \tilde{c}^*_{h*10} \leq 0\right\} \cap \\ &\quad \left[\{\delta_1 : \delta_1' \Lambda^*_{h11*}\delta_1 + b^{*\prime}_{h*11}\delta_1 + c^*_{h*11} \leq 0\} \cup \mathcal{S}^*_{h3}\right] \end{aligned} \quad (5.80)$$

where $\tilde{b}^*_{h*10},\ \tilde{c}^*_{h*10},\ b^*_{h*11},\ c^*_{h*11},$ *and* $\mathcal{S}^*_{h3}$ *are defined as in Theorems* 3.**4** - 4.**3**;

$(b.2.2)$ *if* $\Lambda^*_{h22} = 0,$ *then*

$$\begin{aligned} C^\infty_{\delta_1}(\alpha) &= \left\{\delta_1 : \tilde{b}^{*\prime}_{h*1}\delta_1 + \tilde{c}^*_{h*20} \leq 0\right\} \cap \\ &\quad [\{\delta_1 : \delta_1' \Lambda^*_{h11}\delta_1 + b^{*\prime}_{h*1}\delta_1 + c^*_{h*} \leq 0\} \cup \mathcal{S}^*_{h4}] \end{aligned} \quad (5.81)$$

$(b.2.3)$ *if* Λ^*_{h22} *is not positive semidefinite, then*

$$\begin{aligned} C^\infty_{\delta_1}(\alpha) &= \left\{\delta_1 : \tilde{b}^{*\prime}_{h*1}\delta_1 + \tilde{c}^*_{h*} \leq 0\right\} \quad \text{if} \quad \tilde{b}^*_{h*2} = 0 && (5.82) \\ &= \mathbb{R}^{G_1} \quad \text{if} \quad \tilde{b}^*_{h*2} \neq 0\,, && (5.83) \end{aligned}$$

where $\tilde{b}^*_{h*1},\ \tilde{c}^*_{h*20},\ c^*_{h*},\ \mathcal{S}^*_{h4},$ *and* $\tilde{b}^*_{h*2}$ *are defined as in Theorems* 3.**4** - 4.**3**;

$(b.3)$ *if* $\tilde{A}_h + \hat{\lambda} A_h$ *is not positive semidefinite,*

$$C^{\infty}_{\delta_1}(\alpha) = \mathbb{R}^{G_1} . \tag{5.84}$$

The results of Theorem **5.2** are nearly similar to those of Theorems **4.3**. In particular, we observe that $C^{\infty}_{\delta_1}(\alpha)$ can be bounded even if $C^{\infty}_{\delta}(\alpha)$ is unbounded. Furthermore, Theorem **5.2** can be used as a partial identification-robust exogeneity tests to assess whether the subset of regressors Y_1 are endogenous or not. In particular, our procedure can be used as a pre-test for the partial exogeneity. Moreover, remark that our procedure is still valid for the general hypotheses of the form $H_0 : \delta_1 = \delta_{01} \neq 0$ even in presence of weak IV. This constitutes an important contribution compared with the GW-type tests [see Dufour (1987), Doko and Dufour (2009c)].

Section 6 below studies the properties (bias and MSE) of the pretest-estimators where our procedure is used as pre-test in the first stage.

6. Projection-based pretest-estimators

In this section, we first define the pretest-estimators which are considered. Second, we analyze through a Monte Carlo experiment: (1) the properties (level and power) of the projection-based pre-test; (2) the performance (bias and MSE) of the pretest-estimators where the projection-based procedure is used as pre-test in the first stage.

6.1. pretest-estimators

Suppose that we want to estimate β_1 and β_2 in model (2.2)-(2.4). If Y_1 were exogenous, estimating β_1 by OLS and use the available instruments only for Y_2 should improve the efficiency of the estimator. The challenge here is to propose a procedure which selects the subset of exogenous variables among a set of regressors in order to improve the estimation of model parameter.

To achieve this goal, we consider the following two pretest-estimators:

$$\hat{\beta} = \hat{\beta}_{PIV}\mathbb{1}_{[0\in C_{\delta_1}(\alpha)]} + \hat{\beta}_{2SLS}\mathbb{1}_{[0\notin C_{\delta_1}(\alpha)]}\,, \tag{6.1}$$

$$\tilde{\beta} = \tilde{\beta}_{PIV}\mathbb{1}_{[0\in C_{\delta_1}(\alpha)]} + \hat{\beta}_{2SLS}\mathbb{1}_{[0\notin C_{\delta_1}(\alpha)]}\,, \tag{6.2}$$

where α is the nominal size of the pre-test and $\mathbb{1}_{[.]}$ is the indicator function. The estimators $\hat{\beta}_{2SLS}$, $\hat{\beta}_{PIV}$ and $\tilde{\beta}_{PIV}$ are given by

$$\begin{aligned}\hat{\beta}_{2SLS} &= (\hat{Y}'M_{X_1}\hat{Y})^{-1}\hat{Y}'M_{X_1}y,\\ \hat{\beta}_{PIV} &= [\hat{\beta}'_{1OLS},\, \hat{\beta}'_{2IV}]',\\ \tilde{\beta}_{PIV} &= [\hat{\beta}'_{1OLS},\, \tilde{\beta}'_{2IV}]'\,,\end{aligned} \tag{6.3}$$

where

$$\begin{aligned}\hat{\beta}_{1OLS} &= (Y_1'M_{[X_1,\,Y_2]}Y_1)^{-1}Y_1'M_{[X_1,\,Y_2]}y,\\ \hat{\beta}_{2IV} &= (\hat{Y}_2'M_{[X_1,\,Y_1]}\hat{Y}_2)^{-1}\hat{Y}_2'M_{[X_1,\,Y_1]}y\,,\\ \tilde{\beta}_{2IV} &= (\tilde{Y}_2'M_{[X_1,\,Y_1]}\tilde{Y}_2)^{-1}\tilde{Y}_2'M_{[X_1,\,Y_1]}y,\end{aligned} \tag{6.4}$$

and

$$\begin{aligned}\hat{Y} &= [\hat{Y}_1,\, \hat{Y}_2] = P_X Y,\, \tilde{Y}_2 = P_W Y_2\,, W = [X,\, Y_1],\\ \hat{Y}^* &= [Y_1,\, \hat{Y}_2],\, \tilde{Y} = [Y_1,\, \tilde{Y}_2]\,.\end{aligned} \tag{6.5}$$

Note that for any matrix B, $P_B = B(B'B)^{-1}B'$ is the projection matrix in the space spanned by the columns of B and $M_B = 1 - P_B$. The main idea behind (6.1) - (6.2) is explained in Doko and Dufour (2009c).

The next subsection studies through a Monte Carlo experiment, the properties of the projection-based pre-test (level and power) and the performance of different pretest-estimators.

6.2. Simulation experiment

Consider the following two endogenous simultaneous equations model

$$\begin{aligned} y &= Y_1\beta_1 + Y_2\beta_2 + u, \\ (Y_1,\, Y_2) &= (X_2\mathit{\Pi}_{21},\, X_2\mathit{\Pi}_{22}) + (V_1,\, V_2), \\ u &= V_1 a_1 + V_2 a_2 + \varepsilon = V(\Sigma_V^{-1}\delta) + \varepsilon\,. \end{aligned} \tag{6.1}$$

where y: $T \times 1$, Y_1: $T \times G_1$, Y_2: $T \times G_2$, $\delta = (\delta_1,\, \delta_2)'$. δ_1 is the covariance between u and V_1, and δ_2 those between u and V_2.

The DGP is such that

$$(\varepsilon_t, V_t)' \overset{i.i.d}{\sim} N\left[0, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & .45 \\ 0 & .45 & 1 \end{pmatrix}\right], \quad \text{for all } t = 1,\, \ldots,\, T\ . \tag{6.2}$$

The instruments $Z_{2t} \overset{i.i.d}{\sim} N(0,\, I_{k_2}),\ t = 1,\, \ldots,\, T,$ are fixed over the experiment. $\mathit{\Pi}_{21}$ and $\mathit{\Pi}_{22}$ are vectors of dimension k_2 defined as

$$\mathit{\Pi}_{21} = \eta_1 \mathit{\Pi}_0, \quad \mathit{\Pi}_{22} = \eta_2 \mathit{\Pi}_1\,, \eta_j = \sqrt{\frac{\mu_j^2}{T\|Z_2 C\|}},\ j = 1,\, 2\,, \tag{6.3}$$

where the $k_2 \times 2$ fixed matrix $[\mathit{\Pi}_0,\, \mathit{\Pi}_1]$ is obtained by taking the first two columns of the identity matrix with dimension k_2, $\mu_1 \in \{0,\, 13,\, 200,\, 613,\, 2000\}$ and $\mu_2 = \{0,\, 13,\, 613,\, 2000\}$. Note that $\mu_j \leq 613,\ j = 1,\, 2,$ characterizes weak IV for the corresponding variable Y_j, whereas $\mu_j >$ 613 characterizes strong IV for Y_j [see Doko and Dufour (2009a) and Hansen et al. (2008)]. For example, If $\mu_j = 0$, then, the IV are irrelevant for Y_j and cannot identify β_j. However, for $\eta_j =$ 2000, the IV are strong and β_j is identified. The true β is $\beta = (\beta_1,\, \beta_2)' = (2,\, 5)'$ and the number of instruments k_2 belongs to $\{5,\, 20,\, 40\}$ in Subsection 6.2.1 and fixed at $k_2 = 20$ in Subsection 6.2.2. we generate $N = 10,000$ replications and the sample size is $T = 50$.

6.2.1. Level and Power of the projection-based procedure

This subsection studies numerically, the properties–level and power–of the projection-based procedure described above. We test the partial exogeneity of Y_1 in (2.2)-(2.4), *i.e.* the hypothesis

$$H_0 : \delta_1 = 0 . \tag{6.4}$$

In this section, the true δ_1 value of belongs to $\{-.9,\ 0,\ .8\}$. So, in the tables, the column $\delta_1 = 0$, corresponds to the level of the projection-based procedure. The other columns correspond to the power. The nominal level is 5 % [2.5 % for each AR-test of Section 3]. Table 4.1 below presents the results. The table is divided into four parts corresponding the four values of μ_2: $\{0,\ 13,\ 613,\ 2000\}$. Here, μ_2 is the quality of the IV for Y_2. The first column of the table contains the number of instruments k_2 whereas the other columns contain the empirical rejection frequencies of our procedure for each value of δ_1 and μ_1 (quality of the IV for Y_1). There are two main findings from these results: (1) Our procedure is valid (level is controlled in all cases) even when identification is deficient (weak instruments), as expected; (2) the procedure has a good power even if the IV are irrelevant (contrast to DWH-tests and GW-test). This later results is very interesting and is explained by the fact that even if β and a are not identified when the IV are weak, $\theta = \beta + a$ is always identified.

6.2.2. Performance of pretest-estimators: bias and MSE

We now examine the performance of the estimators defined in

(4.2) - (4.4). We test the exogeneity of Y_1 in the model (2.2)-(2.4), *i.e.* the hypothesis

$$H_0 : \delta_1 = 0 . \tag{6.5}$$

The value of δ_1 belongs to $\{-2,\ 0,\ 5\}$ and the number of instruments is fixed at $k_2 = 20$ over the experiment. Tables 4.4- 4.8 in Appendix B present the results. The first column of the tables contains the estimators studied and the other columns present the bias (for each estimator of β_1 and

Table 4.1. Level and Power of the projection-based procedure with nominal level of 5%, $T = 50$, irrelevant IV for Y_2

$H_0 : \delta_1 = 0$, irrelevant IV for $Y_2 : \mu_2 = 0$															
$\delta_1 \rightarrow$	-.9					0					.8				
$k_2 \downarrow$ $\mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
5	98,16	98,22	98,25	98,28	98,15	4,88	5,06	4,63	4,65	5,05	95,15	95,52	95,30	94,74	94,85
20	88,76	89,22	88,91	89,17	88,68	4,94	5,00	4,94	4,98	5,00	80,26	81,04	80,68	81,17	80,94
40	30,46	30,54	30,73	30,73	30,16	5,15	5,17	4,99	4,43	4,83	24,60	25,06	24,28	24,04	24,74
$H_0 : \delta_1 = 0$, weak IV for $Y_2 : \mu_2 = 13$															
5	98,31	98,36	98,21	98,30	98,39	4,91	4,86	4,99	4,98	4,95	95,26	95,09	95,74	94,89	95,19
20	89,28	89,05	88,89	88,62	89,47	5,21	4,78	4,92	4,76	5,13	81,07	80,82	80,99	81,36	80,99
40	31,10	30,14	30,49	30,54	30,71	4,48	4,96	5,00	4,98	4,90	24,99	25,16	25,45	25,31	24,43
$H_0 : \delta_1 = 0$, weak IV for $Y_2 : \mu_2 = 613$															
5	98,18	98,27	98,33	98,24	98,38	4,66	5,16	4,59	4,84	4,94	95,20	95,21	95,01	95,18	95,21
20	88,55	88,75	89,08	88,92	89,31	4,97	4,87	5,05	5,07	5,04	81,31	80,33	80,41	80,84	80,34
40	29,93	30,56	30,75	31,26	30,85	5,28	5,13	4,99	4,87	4,88	25,30	25,06	24,34	25,03	24,65
$H_0 : \delta_1 = 0$, strong IV for $Y_2 : \mu_2 = 2000$															
5	98,10	98,40	98,43	98,32	98,26	4,80	4,97	5,25	4,91	5,18	95,63	95,21	95,01	95,35	95,20
20	89,22	89,56	89,05	89,11	89,16	5,19	5,09	4,85	4,80	5,36	81,51	80,76	80,80	81,26	81,60
40	30,94	30,29	30,62	30,50	30,29	5,13	5,08	4,74	4,92	4,84	24,30	25,07	24,02	24,73	24,86

β_2) or the mean square errors (MSE).

Our results indicate that OLS and partial IV-estimators outperform 2SLS when the IV are weak [similar to Doko and Dufour (2009a)]. Further, the pretest-estimators have a good performance (bias and MSE) compared to 2SLS when the IV are weak. Overall, the results of this section suggest that the practice of pre-testing has a nice feature and may not be abandoned as recommended by Guggenberger (2008).

7. Empirical applications

To illustrate our theoretical results, we consider two empirical applications. The first one is the relation between trade and growth [Doko and Dufour (2009a), Dufour and Taamouti (2006), Irwin and Tervio (2002), Frankel and Romer (1999), Harrison (1996), Mankiw and al. (1992)]. In this application, it is known that the quality of the instruments is not too poor. The second application is the problem of returns to education [Doko and Dufour (2009a), Dufour and Taamouti (2006), Angrist and Krueger (1991), Angrist and Krueger (1995), Angrist and al. (1999), Mankiw and al. (1992)] where the quality of the instruments is poor.

7.1. Trade and growth model

This model studies the relationship between standards of living and openness. Frankel and Romer (1999) argue that trade share (ratio of imports or exports to GDP) which is the commonly used indicator of openness should be viewed as endogenous variable, and similarly for the other indicators such as trade policies. The authors suggest that instrumental variables method should be applied for estimating the income-trade relationship. The equation studied is specified as

$$ln(\text{Income}_i) = \beta_0 + \beta_1 \text{Trade}_i + \gamma_1 ln(\text{Pop}_i) + \gamma_2 ln(\text{Area}_i) + u_i, \tag{7.1}$$

where Income_i is the income per capita in country i, Trade_i is the trade share (measured as a ratio of imports and exports to GDP), Pop_i the logarithm of population, and Area_i the logarithm of country area. The instrument suggested by Frankel and Romer (1999) is constructed on the basis of geographic characteristics. The first stage equation is given by

$$\text{Trade}_i = b_0 + b_1 Z_i + c_1 \text{Pop}_i + c_2 \text{Area}_i + V_i, \tag{7.2}$$

where Z_i is a constructed instrument from geographic characteristics. We use the sample of 150 countries and the data include for each country the trade share in 1985, the area and population (1985), per capita income (1985), and the fitted trade share (instrument). For this sample, it is not clear how"weak "the instruments are[1].

In this application, we compare the projection-based pre-test procedure to earlier procedures: Durbin-Wu-Hausman (DWH) and generalized Wald (GW) tests. The hypothesis to test is the exogeneity of trade share in (7.1).

Our results are as follow. When applying DWH-tests to assess whether trade share is exogenous, the p-value are .1222, .1193, .04969, .0495, .1243, .0499 for respectively $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$, $\mathcal{T}_2$, $\mathcal{T}_3$ and $\mathcal{T}_4$ [see Doko and Dufour (2009a)]. Hence, except for $\mathcal{H}_3$, $\mathcal{T}_2$, and $\mathcal{T}_4$, the exogeneity cannot be rejected by the other statistics. This suggests that the IV is not possibly too strong. This is supported by the Wald-type test which has a p-value of .9624 .

Now, if we estimate the endogeneity "a" by OLS in the transformed equation

$$ln(\text{Income}_i) = \beta_0 + \beta_1 \text{Trade}_i + \gamma_1 ln(\text{Pop}_i) + \gamma_2 ln(\text{Area}_i) + \hat{V}_i a + e_i, \tag{7.3}$$

we get $\hat{a} = -1.817$. And using the relation $\hat{\delta} = \hat{\Sigma}_V \hat{a}$, we find $\hat{\delta} = -.3805$.

The Table 4.2 below contains the confidence sets of β_1, $\theta = \beta_1 + a$, a and δ for different nominal levels α.

[1] The F-statistic of the first stage is about 13 as indicated in Frankel and Romer (1999, Table 2, p.385)

Table 4.2. Projection-based confidence sets for different parameters in growth model

AR-type CS's	99.5 %	97.5 %	95 %
$C_{\beta_1}(\alpha)$	$\{\beta_1 : -.735\beta_1^2 - 4.509\beta_1 - 2.923 \leq 0\}$ $=]-\infty, -5.401] \cup [-.737, +\infty[$	$\{\beta_1 : .4612\beta_1^2 - 4.757\beta_1 + .0426 \leq 0\}$ $= [.284, 4.652]$	$\{\beta_1 : .963\beta_1^2 - 4.754\beta_1 + 1.274 \leq 0\}$ $= [.009, 10.307]$
$C_{\theta}(\alpha)$	$\{\theta : .611\theta^2 - .127\theta - .068 \leq 0\}$ $= [-.245, .453]$	$\{\theta : .611\theta^2 - .127\theta - .0385 \leq 0\}$ $= [-.1678, .3755]$	$\{\theta : .611\theta^2 - .127\theta - .026 \leq 0\}$ $= [-.128, .337]$
Scheffé-type CS's	99 %	95 %	90 %
$C_a(\alpha)$	$]-\infty, 1.19] \cup [5.156, +\infty[$	$[-10.4746, .3666]$	$[-4.780, .052]$
$C_\delta(\alpha)$	$]-\infty, .249] \cup [1.08, +\infty[$	$[-2.1932, .0768]$	$[-1.001, .011]$

7.2. Education and earnings

We now consider the problem of estimating returns to education. The model studies a relationship between log weekly earning and the number of years of education and several other covariates (age, age squared, year of birth, ...). Education is probably an endogenous variable in the model and several authors including Angrist and Krueger (1991) have proposed to use the birth quarter as an instrument. The idea is that individuals born in the first quarter of the year start school at an older age, and can therefore drop out after completing less schooling than individuals born near the end of the year. Therefore, individuals born at the beginning of the year are likely to earn less than those born during the rest of the year. Moreover, it is well known that the instruments used by Angrist and Krueger (1991) are very weak , *e.g.* Bound (1995). So, standard DWH-tests for the exogeneity of education conclude that education is exogenous [see Doko and Dufour (2009a)]. This is surprising since DWH test have no power when all instruments are weak. Applying the generalized Wald (GW)-test also leads to the same conclusion. The GW-test p-value for the test of the exogeneity of education is too large, about .997, suggesting that education can be treated as exogenous. As showed in Doko and Dufour (2009c), the relevance of the GW-tests is questionable when the instruments are weak, as it is the case in this model. Clearly, these results are coherent with the theory. The only plausible is the projection-based procedure which still has a power even if when all the IV are irrelevant.

The goal of this application is to compare the results when applying the projection-based procedure to earlier procedures (DWH and GW tests) for assessing the exogeneity of education. The model is:

$$y = \beta_0 + \beta_1 E + \sum_{i=1}^{k_1} \gamma_i X_i + u \,, \tag{7.4}$$

$$E = \pi_0 + \sum_{i=1}^{k_2} \pi_i Z_i + \sum_{i=1}^{k_1} \phi_i X_i + V\,, \tag{7.5}$$

where y is log-weekly earnings, E is the number of years of education (possibly endogenous), X contains the exogenous covariates (age, age squared, 10 dummies for birth of year). Z contains 40 dummies obtained by interacting the quarter of birth with the year of birth. In this model, β_1 measures the return to education. The data set consists of the 5% public-use sample of the 1980 US census for men born between 1930 and 1939. The sample size is 329 509 observations.

Our results with the projection-based procedure are presented in Table 4.3 below.

8. Conclusion

This paper focuses on structural models and develops identification-robust inference for covariances between errors and regressors.

First, we propose a finite-and large-sample identification-robust confidence sets for covariances and an auxiliary endogeneity parameter. We derive analytic forms of these confidence sets and characterize necessary and sufficient conditions under which they are bounded. Then, the results are applied to get identification-robust partial exogeneity tests. Our asymptotic theory under weaker assumptions and allowing for heteroskedasticity and autocorrelation of errors confirms the finite-sample results. The Monte Carlo experiment shows that our procedure has power even when identification is weak (contrary to GW-test and DWH-tests).

Second, we propose pretest-estimators of the structural parameter which are more efficient than usual IV estimator in presence of weak instruments. When the instruments are strong, our new estimators behave like usual IV estimator. Therefore, this suggests that the projection-based techniques can be used as a variable selection method to decide which variables should be instrumented and which ones are valid instruments.

We illustrate our results through two empirical applications: the relation between trade and

Table 4.3. Projection-based confidence sets for different parameters in earning equation

AR-type CS's	99.5 %	97.5 %	95 %
$C_{\beta_1}(\alpha)$	$\{\beta_1 : -2.7\beta_1^2 + .377\beta_1 - .121 \leq 0\} = \mathbb{R}$	$\{\beta_1 : -2.382\beta_1^2 + .332\beta_1 - .107 \leq 0\} = \mathbb{R}$	$\{\beta_1 : -2.229\beta_1^2 + .31\beta_1 - .1 \leq 0\} = \mathbb{R}$
$C_\theta(\alpha)$	$\{\theta : 4.398\theta^2 - .5\theta - .576 \leq 0\}$ $= [-.31,\ .423]$	$\{\theta : 4.301\theta^2 - .5\theta - .575 \leq 0\}$ $= [-.312,\ .428]$	$\{\theta : 4.254\theta^2 - .5\theta - .575 \leq 0\}$ $= [-.314,\ .431]$
Scheffé-type CS's	**99 %**	**95 %**	**90 %**
$C_a(\alpha)$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$
$C_\delta(\alpha)$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$

economic growth and the widely studied problem of returns to education. The results showed that the confidence sets for the covariance and endogeneity parameter are unbounded. That is, the quality of the instruments in these models is poor, as questioned in the literature [Bound (1995), Frankel and Romer (1999)].

APPENDIX

A. Proofs

PROOF OF THEOREM **3.1** Consider the AR-statistic defined by (3.3):

$$AR(\beta_0) = \frac{(y - Y\beta_0)'(M_1 - M)(y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(y - Y\beta_0)/(T - k)}\,. \tag{A.1}$$

As in Doko and Dufour (2009a), we can show that if we replace y and Y by $\bar{y}$ and $\bar{Y}$ defined in (**??**), then the expression of $AR(\beta_0)$ does not change. So, we can write

$$AR(\beta_0) = \frac{(\bar{y} - \bar{Y}\beta_0)'(M_1 - M)(\bar{y} - \bar{Y}\beta_0)/k_2}{(\bar{y} - \bar{Y}\beta_0)'M(\bar{y} - \bar{Y}\beta_0)/(T - k)}\,. \tag{A.2}$$

However, from (4.56), we have

$$\begin{aligned} M(\bar{y} - \bar{Y}\beta) &= M\bar{v},\ M_1\bar{Y} = M_1(\mu_2 + \bar{V}), \\ M_1(\bar{y} - \bar{Y}\beta) &= M_1[\bar{v} + \mu_2(\beta - \beta_0) + \bar{V}(\beta - \beta_0)]\,. \end{aligned} \tag{A.3}$$

If $\beta = \beta_0$, we have $M_1(\bar{y} - \bar{Y}\beta) = M_1\bar{v}$ so that

$$AR(\beta_0) = \frac{\bar{v}'(M_1 - M)\bar{v}/k_2}{\bar{v}'M\bar{v}/(T - k)}\,. \tag{A.4}$$

If further assumption (4.43) holds and if u is independent of X, since $(M_1 - M)^2 = M_1 - M$ and $M^2 = M$, we have

$$\bar{v}'M\bar{v} \sim \chi^2(T - k),\ \bar{v}'(M_1 - M)\bar{v} \sim \chi^2(k_2)\,. \tag{A.5}$$

Furthermore, we have $M(M_1 - M) = 0$, thus $AR(\beta_0) \sim F(k_2, T - k)$.

□

PROOF OF THEOREM **3.2**

(a) Suppose that $2A\beta + b = 0$ *i.e* $\beta = \bar{\beta} = -\frac{1}{2}A^{-}b + (I - \mathcal{H}_A)\beta_{0*}$, where β_{0*} is any arbitrary $G \times 1$ vector and $\mathcal{H}_N = N^{-}N$, where N^{-} is any generalized inverse of N. Then, we have

$$Q(a, \bar{\beta}) \quad = \quad a'\tilde{A}a + \bar{b}'a + \bar{c}\,, \tag{A.6}$$

where $\bar{b} = \tilde{b} + 2\tilde{A}\bar{\beta},$ and $\bar{c} = \tilde{c} + \bar{\beta}'\tilde{A}\bar{\beta} + \tilde{b}'\bar{\beta}$. Moreover, we also have

$$f(\bar{\beta}) \quad = \quad c - \frac{1}{4}b'A^{-}b + \frac{1}{2}b'(I - \mathcal{H}_A)\beta_{0*} \equiv c^{*}\,. \tag{A.7}$$

So, we see immediately from (3.40) that

$$C_a(\alpha) = \begin{cases} \left\{ a \in \mathbb{R}^G : a'\tilde{A}a + \bar{b}'a + \bar{c} \leq 0 \right\} & \text{if} \quad c^{*} \leq 0 \\ \emptyset \quad \text{otherwise} \end{cases} \tag{A.8}$$

(b) Suppose now that $2A\beta + b \neq 0$. Define the Lagrangian of the problem (3.40) as

$$\mathcal{L}(\beta, a, \lambda) = \mathcal{Q}(\beta, a) + \lambda f(\beta)\,. \tag{A.9}$$

The F.O.C are

$$\frac{\partial \mathcal{Q}(\beta, a)}{\partial \beta} + \lambda \frac{\partial f(\beta)}{\partial \beta} \quad = \quad 0 \tag{A.10}$$

$$\lambda f(\beta) \quad = \quad 0 \tag{A.11}$$

$$\lambda \geq 0, \quad f(\beta) \quad \leq \quad 0 \tag{A.12}$$

i.e.

$$2\tilde{A}\beta + 2\tilde{A}a + \tilde{b} + \lambda(2A\beta + b) \quad = \quad 0 \tag{A.13}$$

$$\lambda f(\beta) \quad = \quad 0 \tag{A.14}$$

$$\lambda \geq 0, \quad f(\beta) = \beta'A\beta + b'\beta + c \quad \leq \quad 0\,. \tag{A.15}$$

The bordered hessian for binding constraint is given by

$$\bar{H}_{br} = \begin{bmatrix} 0 & 2\beta' A + b' \\ 2A\beta + b & 2\tilde{A} + 2\lambda A \end{bmatrix} . \tag{A.16}$$

(b.1) Assume that $\tilde{A} + \lambda A \geq 0$. Then, the bordered hessian based approach in Magnus and Neudecker (1998, Theorem 12) is satisfied. Suppose that $\lambda > 0$ then, the minimum of $\mathcal{L}(\beta, a, \lambda)$ is obtained for $\bar{\beta}^*$ such that $f(\bar{\beta}^*) = 0$ and from (A.10), we have

$$\begin{aligned} 2\tilde{A}a &= -2\tilde{A}\bar{\beta}^* - \tilde{b} - \lambda(2A\bar{\beta}^* + b)\,, \quad \text{i.e} \\ a &= -\tilde{A}^-\tilde{A}\bar{\beta}^* - \frac{\tilde{A}^-\tilde{b}}{2} - \lambda(\tilde{A}^- A\bar{\beta}^* + \frac{\tilde{A}^- b}{2})\,. \end{aligned} \tag{A.17}$$

So, with a little algebraic, $\mathcal{L}(\beta, a, \lambda)$ becomes

$$\mathcal{L}(\bar{\beta}^*, \lambda) = \phi_0\lambda^2 - \phi_1\lambda + \phi_2\,, \tag{A.18}$$

where $\phi_0 = (A\bar{\beta}^* + \frac{b}{2})'\tilde{A}^-(A\bar{\beta}^* + \frac{b}{2})$, $\phi_1 = \bar{\beta}^{'*}A\bar{\beta}^* + \frac{b'\bar{\beta}^*}{2}$, $\phi_2 = \bar{\beta}^{'*}\tilde{A}\bar{\beta}^* + \tilde{c} + \frac{b'\bar{\beta}^*}{2}$, and where we also use the identity $\tilde{A}^-\tilde{A}\bar{\beta}^* = \bar{\beta}^*$. Hence, $\hat{\lambda} \in \mathbb{R}_+ \setminus \{0\}$ which minimizes (A.18) is given by

$$\hat{\lambda} = \begin{cases} \frac{\phi_1}{2\phi_0} & \text{if} \quad \phi_0 > 0 \\ \mathbb{R}_+ & \text{if} \quad \phi_0 = 0 \quad \text{and} \quad \phi_1 \geq 0. \end{cases} \tag{A.19}$$

So, the set C_a takes the form

$$C_a(\alpha) = \left\{ a \in \mathbb{R}^G : a'\tilde{A}a + \tilde{b}'a + \tilde{c}_{\hat{\lambda}} \leq 0 \right\}, \tag{A.20}$$

where $\tilde{c}_{\hat{\lambda}} = \tilde{c} - \bar{\beta}^{'*}\tilde{A}\bar{\beta}^* - \hat{\lambda}(A\bar{\beta}^* + b)'\bar{\beta}^*$.

Suppose now that $\lambda = 0$. Then, we have to minimize

$$\mathcal{L}(\beta, a, 0) \equiv \mathcal{L}(\beta, a) = \mathcal{Q}(\beta, a) \tag{A.21}$$

over β. So, from (A.13),

$$\beta \;\; = \;\; -\tilde{A}^-\tilde{A}a - \frac{\tilde{A}^-\tilde{b}}{2} + (I - \mathcal{H}_{\tilde{A}})\beta_{30} = -\mathcal{H}_{\tilde{A}}a - \frac{\tilde{A}^-\tilde{b}}{2} + (I - \mathcal{H}_{\tilde{A}})\beta_{30}, \tag{A.22}$$

is the minimum, where β_{30} is an arbitrary vector. By substituting (A.22) in the objective $\mathcal{Q}(a, \beta)$,

$$\mathcal{Q}_0^*(a) = \tilde{b}_*' a + \tilde{c}_* \,, \tag{A.23}$$

where

$$\tilde{b}_* \;=\; (I - \mathcal{H}_{\tilde{A}}')\tilde{b},\ \tilde{c}_* = \tilde{c} - \frac{1}{4}\tilde{b}'\tilde{A}^-\tilde{b} + \frac{1}{2}\tilde{b}'(I - \mathcal{H}_{\tilde{A}})\beta_{30}\,. \tag{A.24}$$

We note that if $\tilde{A} > 0,\ \tilde{b}_* = 0$ and $\mathcal{Q}_0^*(a) = \tilde{c}_* = \tilde{c} - \frac{1}{4}\tilde{b}'\tilde{A}^-\tilde{b}$ for any $a \in \mathbb{R}^G.$ On the other hand, if $\tilde{A} = 0,\ \tilde{b}_* = \tilde{b}$ and $\mathcal{Q}_0^*(a) = \tilde{b}'a + \tilde{c}$. Furthermore, from (A.22), we get

$$f(\beta) = a'\varLambda a + b_*' a + c_* \,, \tag{A.25}$$

where

$$\begin{aligned}
\varLambda \;&=\; \mathcal{H}_{\tilde{A}}' A \mathcal{H}_{\tilde{A}},\ b_* = \mathcal{H}_{\tilde{A}}'[-b + A\tilde{A}^-\tilde{b} - 2A(I - H_{\tilde{A}})\beta_{30}]\,,\\
c_* \;&=\; c + \frac{1}{4}\tilde{b}'\tilde{A}^- A\tilde{A}^-\tilde{b} + b'(I - \mathcal{H}_{\tilde{A}})\beta_{30} - \frac{b'\tilde{A}^-\tilde{b}}{2} - \tilde{b}'\tilde{A}^- A(I - \mathcal{H}_{\tilde{A}})\beta_{30}\\
&\quad\ \beta_{30}'(I - \mathcal{H}_{\tilde{A}}')A(I - \mathcal{H}_{\tilde{A}})\beta_{30}\,.
\end{aligned} \tag{A.26}$$

So, we have

$$C_a(\alpha) = \left\{a \in \mathbb{R}^G : \tilde{b}_*' a + \tilde{c}_* \le 0\right\} \cap \{a : a'\varLambda a + b_*' a + c_* \le 0\}. \tag{A.27}$$

(b.2) If $\tilde{A} + \lambda A < 0$, this entail that $\tilde{A} + \lambda A \neq 0$, and we can find a vector β_0 such that $\beta_0'(\tilde{A} + \lambda A)\beta_0 \equiv q_0 < 0$. So, for any scalar δ_0, we have

$$\mathcal{L}(a, \varDelta_0\beta_0, \lambda) = a'\tilde{A}a + \tilde{b}'a + \tilde{c} + \varDelta_0^2 q_0 + \varDelta_0(2\tilde{A}a + \tilde{b} + \lambda b)'\beta_0\,. \tag{A.28}$$

Since $q_0 < 0$, we can choose $\varDelta_0$ sufficiently large to have $\mathcal{L}(a, \varDelta_0\beta_0, \lambda) < 0$, irrespective of the values of a and λ. Hence, all values of a belong to C_a, thus $C_a = \mathbb{R}^G$. $\square$

PROOF OF COROLLARY **3.3**

(A) Suppose that $2A\beta + b = 0$. From Theorem **3.2-**(a), we have

$$C_a(\alpha) = \begin{cases} \left\{ a \in \mathbb{R}^G : a'\tilde{A}a + \bar{b}'a + \bar{c} \leq 0 \right\} & \text{if} \quad c^* \leq 0 \\ \emptyset \quad \text{otherwise} \end{cases} \tag{A.29}$$

where $f(\bar{\beta}) = c - \frac{1}{4}b'A^{-}b + \frac{1}{2}b'(I - \mathcal{H}_A)\beta_{0*} = c^*$. If $c^* > 0$, $C_a(\alpha) = \emptyset$. If $c^* \leq 0$ [this corresponds to the case where $f(\bar{\beta}) = 0$ defines an ellipsoid, hence, $A > 0$), we have $C_a(\alpha) = \left\{ a \in \mathbb{R}^G : a'\tilde{A}a + \bar{b}'a + \bar{c} \leq 0 \right\}$. From Dufour and Taamouti (2005, Econmetrica), $C_a(\alpha)$ is an unbounded non empty set if only if $\tilde{A} > 0$.

(B) Suppose now that $2A\beta + b \neq 0$.

First, assume that $\tilde{A} + \hat{\lambda}A \geq 0,$ where $\hat{\lambda}$ is defined in Theorem **3.2**. If $\hat{\lambda} > 0$, then, $C_a(\alpha) = \left\{ a \in \mathbb{R}^G : a'\tilde{A}a + \tilde{b}'a + \tilde{c}_{\hat{\lambda}} \leq 0 \right\}$ where $\tilde{c}_{\hat{\lambda}} = \tilde{c} - \bar{\beta}'^{*}\tilde{A}\bar{\beta}^* - \hat{\lambda}(A\bar{\beta}^* + b)'\bar{\beta}^*$ and $\bar{\beta}^*$ is such that $f(\bar{\beta}^*) = 0.$ If A is not positive definite, (1) we can find $\bar{\beta}^*$ as large as possible satisfying $f(\bar{\beta}^*) = 0$ [see Dufour and Taamouti (2005) for more details] so that $a'\tilde{A}a + \tilde{b}'a + \tilde{c}_{\hat{\lambda}} \leq 0$. Consequently, $C_a(\alpha)$ is unbounded; or (2) there is no $\bar{\beta}^*$ satisfying $f(\bar{\beta}^*) = 0$ and $C_a(\alpha) = \emptyset$. If $\tilde{A} > 0$ and $A > 0,$ then, $a'\tilde{A}a + \tilde{b}'a + \tilde{c}_{\hat{\lambda}} \leq 0$ defines an ellipsoid and $C_a(\alpha)$ is bounded or empty. If $\hat{\lambda} = 0$, the above arguments hold.

Second, assume that $\tilde{A} + \hat{\lambda}A < 0,$ then, from Theorem **3.2**, $C_a = \mathbb{R}^G$ is unbounded.

It is worthwhile to note that the spectral decomposition argument in Dufour and Taamouti (2005) can also be applied here.

□

PROOF OF THEOREM **3.4**

(a)- Assume that $2A\beta + b = 0$. From, we have

$$C_a(\alpha) = \Big\{a \in \mathbb{R}^G : a'\tilde{A}a + \bar{b}'a + \bar{c} \leq 0\Big\} \quad \text{if} \quad c^* \leq 0 \tag{A.30}$$

$$= \emptyset \quad \text{otherwise} \quad , \tag{A.31}$$

where

$$c^* = c - \frac{1}{4}b'A^{-}b + \frac{1}{2}b'(I - \mathcal{H}_A)\beta_{0*}\,. \tag{A.32}$$

So, if $c^* > 0$, we have $C_{a-1}(\alpha) = \emptyset$. If $c^* \leq 0$, we

$$Q(a, \bar{\beta}) = Q(a_1, a_2) = a'\tilde{A}a + \bar{b}'a + \bar{c}\,, \tag{A.33}$$

where $\bar{b} = \tilde{b}+2\tilde{A}\bar{\beta},$ and $\bar{c} = \tilde{c}+\bar{\beta}'\tilde{A}\bar{\beta}+\tilde{b}'\bar{\beta}$, and a CS for a_1 is obtained by solving the unconstrained problem

$$\min_{a_2} \mathcal{Q}(a_1, a_2)\,. \tag{A.34}$$

By following the same steps as in Dufour and Taamouti (2006), we get the results of Theorem **3.4**-(a).

(b)- Assume now that $2A\beta + b \neq 0$.

□

PROOF OF LEMMA **4.1** Note first that the AR-type statistics defined in (4.18) - (4.19) can be written as

$$AR^{he}(\beta_0) = \frac{\upsilon_1(\beta_0)'[P_{X^*} - P_{X_1^*}]\upsilon_1(\beta_0)/k_2}{\upsilon_1(\beta_0)'M_{X^*}\upsilon_1(\beta_0)/(T-k_1-k_2)}\,, \tag{A.35}$$

and

$$AR^{he}(\theta_0) = \frac{\varepsilon_1(\theta_0)'[P_{Z^*} - P_{X^{**}}]\varepsilon_1(\theta_0)/G}{\varepsilon_1(\theta_0)'M_{Z^*}\varepsilon_1(\theta_0)/(T-k-G)}\,, \tag{A.36}$$

where $\upsilon_1(\beta_0) = \hat{\Sigma}_u^{-1/2}\upsilon^0(\beta_0) \equiv \upsilon_1$, $\varepsilon_1(\theta_0) = \hat{\Sigma}_\varepsilon^{-1/2}\varepsilon(\theta_0) \equiv \varepsilon_1$, $X^* = [X_1^*,\, X_2^*] = \hat{\Sigma}_u^{-1/2}X$,

$X^{**} = [X_1^{**},\, X_2^{**}] = \hat{\Sigma}_\varepsilon^{-1/2} X,\; Z^* = [Y^{**},\, X^{**}] = \hat{\Sigma}_\varepsilon^{-1/2} Y,\; P_{\mathcal{Z}} = \mathcal{Z}(\mathcal{Z}'\mathcal{Z})^{-1}\mathcal{Z}'$ and $M_{\mathcal{Z}} = I - P_{\mathcal{Z}},$ for any matrix $\mathcal{Z}$. Now, consider the denominators of (A.35), we have

$$\frac{\upsilon_1' M_{X^*} \upsilon_1}{T-k} = \frac{\upsilon_1'\upsilon_1}{T-k} - \frac{T}{T-k}\frac{\upsilon_1' X^*}{T}\left(\frac{X^{*\prime}X^*}{T}\right)^{-1}\frac{X^{*\prime}\upsilon_1}{T}. \tag{A.37}$$

Under H_{β_0}, we have $\upsilon^0 = u$, hence $\frac{\upsilon_1'\upsilon_1}{T-k} = \frac{u'\hat{\Sigma}_u^{-1}u}{T-k} = \frac{u'(\hat{\Sigma}_u^{-1}-\Sigma_u^{-1})u}{T-k} + \frac{u'\Sigma_u^{-1}u}{T-k} \xrightarrow{p} 1,\; \frac{X^{*\prime}\upsilon_1}{T} = \frac{X'\hat{\Sigma}_u^{-1}u}{T} = \frac{X'(\hat{\Sigma}_u^{-1}-\Sigma_u^{-1})u}{T} + \frac{X'\Sigma_u^{-1}u}{T} \xrightarrow{p} 0,\; \frac{X^{*\prime}X^*}{T} = \frac{X'\hat{\Sigma}_u^{-1}X}{T} = \frac{X'(\hat{\Sigma}_u^{-1}-\Sigma_u^{-1})X}{T} + \frac{X'\Sigma_u^{-1}X}{T} \xrightarrow{p} \Omega_X > 0$ because $\underset{T\to\infty}{\text{plim}}\,(\hat{\Sigma}_u) = \Sigma_u$. So, we have $\frac{\upsilon_1' M_{X^*}\upsilon_1}{T-k} \xrightarrow{p} 1.$ The numerator of (A.35) can be written as

$$\begin{aligned}
&\upsilon_1'[P_{X^*} - P_{X_1^*}]\upsilon_1/k_2 \\
&= \left[\begin{array}{cc} \frac{u'\Sigma_u^{-1}X}{\sqrt{T}} & \frac{u'\Sigma_u^{-1}X_1}{\sqrt{T}} \end{array}\right] \left[\begin{array}{cc} \left(\frac{X'\hat{\Sigma}_u^{-1}X}{T}\right)^{-1} & 0 \\ 0 & -\left(\frac{X_1'\hat{\Sigma}_u^{-1}X_1}{T}\right)^{-1} \end{array}\right] \left[\begin{array}{c} \frac{X'\Sigma_u^{-1}u}{\sqrt{T}} \\ \frac{X_1'\Sigma_u^{-1}u}{\sqrt{T}} \end{array}\right] &\text{(A.38)} \\
&\xrightarrow{L} S_u^{x\prime}\Omega_X^{-1}S_u^x - S_u^{x_1'}\Omega_{X_1}^{-1}S_u^{x_1} &\text{(A.39)}
\end{aligned}$$

because $\left[\begin{array}{cc} \left(\frac{X'\hat{\Sigma}_u^{-1}X}{T}\right)^{-1} & 0 \\ 0 & \left(\frac{X_1'\hat{\Sigma}_u^{-1}X_1}{T}\right)^{-1} \end{array}\right] \xrightarrow{p} \left[\begin{array}{cc} \Omega_X^{-1} & 0 \\ 0 & \Omega_{X_1}^{-1} \end{array}\right]$, where $S_u^x = \left[\begin{array}{c} S_u^{x_1} \\ S_u^{x_2} \end{array}\right]$ is partitioned according to $[X_1,\, X_2\,]$. We easily see that (A.39) can be written as

$$S_u^{x\prime}\Omega_X^{-1}S_u^x - S_u^{x_1'}\Omega_{X_1}^{-1}S_u^{x_1} = S_u^{x\prime}\Omega_X^{-1/2}\Omega\Omega_X^{-1/2}S_u^x\,, \tag{A.40}$$

where $\Omega = I_k - \Omega_X^{1/2}\left[\begin{array}{cc} \Omega_{X_1}^{-1} & 0 \\ 0 & 0 \end{array}\right]\Omega_X^{1/2}$. Moreover, $\Omega_X^{-1/2}S_u^x \sim N[0\,, I_k]$ and by noting that $\Omega^2 = \Omega$, hence Ω is idempotent, we have

$$S_u^{x\prime}\Omega_X^{-1/2}\Omega\Omega_X^{-1/2}S_u^x \sim \chi^2(m)\,,$$

where $m = rank(\Omega)$. Furthermore, we have

$$Trace(\Omega) = k - Trace\left(\Omega_X^{1/2}\left[\begin{array}{cc} \Omega_{X_1}^{-1} & 0 \\ 0 & 0 \end{array}\right]\Omega_X^{1/2}\right)$$

and

$$Trace\left(\Omega_X^{1/2}\begin{bmatrix}\Omega_{X_1}^{-1} & 0\\ 0 & 0\end{bmatrix}\Omega_X^{1/2}\right) = Trace\left(\Omega_X\begin{bmatrix}\Omega_{X_1}^{-1} & 0\\ 0 & 0\end{bmatrix}\right) = Trace\left(\begin{bmatrix}I_{k_1} & 0\\ \Omega_{X_2X_1} & 0\end{bmatrix}\right) = k_1\,,$$

where $\Omega_X = \begin{bmatrix}\Omega_{X_1} & \Omega'_{X_2X_1}\\ \Omega_{X_2X_1} & \Omega_{X_2}\end{bmatrix}$. So, we have $Trace(\Omega) = k_2$. Since Ω is idempotent, its eigenvalues λ_j are 1 or 0 [see Magnus and Neudecker (1998, Theorem 7, p.14)]. Because, $Trace(\Omega) = \sum_{j=1}^{k} = k_2$ then Ω has k_2 eigenvalues equal 1. So, $m = rank(\Omega) = k_2$ and $S_u^{x\prime}\Omega_X^{-1/2}\Omega\Omega_X^{-1/2}S_u^x \sim \chi^2(k_2)$, *i.e.*, $AR^{he}(\beta_0) \xrightarrow{L} \frac{1}{k_2}\chi^2(k_2)$. By following the same steps as above, we can show that $AR^{he}(\theta_0) \xrightarrow{L} \frac{1}{G}\chi^2(G)$.

□

PROOF OF THEOREM **4.2** Same proof as in Theorem **3.2**. □

B. Performance of OLS, 2SLS, Partial IV and Projection-based pretest-estimators

Table 4.4. Bias of OLS, 2SLS, Partial IV and Pretest-estimators: weak IV for Y_2 ($\mu_{22} = 0$)

$H_0 : \delta_1 = 0$, irrelevant IV for $Y_2 : \mu_{22} = 0$																
	$\delta_1 \rightarrow$	-2					0					5				
	$k_2 \downarrow$ $\mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
OLS	$B(\hat{\beta}_1)$	-2.0012	-1.0579	-1.9923	-1.9756	-1.9218	0.0000	-0.0011	-0.0006	-0.0002	0.0005	5.0010	4.9975	4.9784	4.9377	4.8042
	$B(\hat{\beta}_2)$	-0.0006	0.0001	-0.0041	-0.0116	-0.0337	0.0003	0.0002	0.0012	0.0003	-0.0008	-0.0004	0.0011	0.0087	0.0279	0.0872
2SLS	$B(\hat{\beta}_1)$	-1.9974	-1.9886	-1.8821	-1.6757	-1.2191	0.0043	0.0010	-0.0007	0.0030	0.0001	5.0017	4.9812	4.7001	4.1864	3.0499
	$B(\hat{\beta}_2)$	-0.0040	-0.0048	-0.0553	-0.1443	-0.3427	0.0036	-0.0028	0.0027	-0.0008	-0.0030	-0.0019	0.0085	0.1248	0.3681	0.8775
$\hat{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0012	-1.0579	-1.9923	-1.9756	-1.9218	0.0000	-0.0011	-0.0006	-0.0002	0.0005	5.0010	4.9975	4.9784	4.9377	4.8042
	$B(\hat{\beta}_2)$	-1.2633	-1.2606	-1.2579	-1.2511	-1.2523	-1.2546	-1.2571	-1.2494	-1.2348	-1.2241	-1.2614	-1.2601	-1.2507	-1.2219	-1.1310
$\tilde{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0012	-1.0579	-1.9923	-1.9756	-1.9218	0.0000	-0.0011	-0.0006	-0.0002	0.0005	5.0010	4.9975	4.9784	4.9377	4.8042
	$B(\hat{\beta}_2)$	-0.0039	-0.0056	-0.0568	-0.1601	-0.4555	0.0028	-0.0034	0.0020	-0.0003	-0.0015	-0.0025	0.0083	0.1297	0.4076	1.1625
2 Stage estimators, Y_1 is not used as IV																
Pre-test	$B(\hat{\beta}_1)$	-1.9974	-1.9886	-1.8821	-1.6757	-1.2191	0.0001	-0.0010	-0.0006	-0.0002	0.0005	5.0017	4.9812	4.7001	4.1864	3.0499
	$B(\hat{\beta}_2)$	-0.0040	-0.0048	-0.0553	-0.1443	-0.3427	-1.2273	-1.2254	-1.2232	-1.2091	-1.1967	-0.0019	0.0085	0.1248	0.3681	0.8775
2 Stage estimators, Y_1 is used as IV																
Pre-test	$B(\hat{\beta}_1)$	-1.9974	-1.9886	-1.8821	-1.6757	-1.2191	0.0001	-0.0010	-0.0006	-0.0002	0.0005	5.0017	4.9812	4.7001	4.1864	3.0499
	$B(\hat{\beta}_2)$	-0.0040	-0.0048	-0.0553	-0.1443	-0.3427	0.0028	-0.0034	0.0020	-0.0003	-0.0016	-0.0019	0.0085	0.1248	0.3681	0.8775

Table 4.5. Bias of OLS, 2SLS, Partial IV and Pretest-estimators: weak IV for Y_2 ($\mu_{22} = 13$)

$H_0 : \delta_1 = 0$, weak IV for $Y_2 : \mu_{22} = 13$																
	$\delta_1 \rightarrow$	-2					0					5				
	$k_2 \downarrow$ $\mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
OLS	$B(\hat{\beta}_1)$	-1.9997	-1.9997	-1.9909	-1.9738	-1.9211	0.0006	0.0003	-0.0002	-0.0005	-0.0014	4.0903	4.9981	4.9800	4.9377	4.8016
	$B(\hat{\beta}_2)$	0.0016	0.0001	-0.0042	-0.0115	-0.0355	0.0001	0.0005	0.0009	0.0006	0.0000	0.0008	-0.0001	0.0094	0.0267	0.0875
2SLS	$B(\hat{\beta}_1)$	-2.0038	-1.9895	-1.8764	-1.6758	-1.2224	0.0003	-0.0072	-0.0001	-0.0028	-0.0032	4.9949	4.9796	4.7030	4.1969	3.0564
	$B(\hat{\beta}_2)$	0.0026	-0.0052	-0.0547	-0.1451	-0.3436	-0.0046	0.0017	-0.0024	0.0047	-0.0018	0.0087	0.0096	0.1262	0.3447	0.8576
$\hat{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-1.9997	-1.9997	-1.9909	-1.9738	-1.9211	0.0006	0.0003	-0.0002	-0.0005	-0.0014	4.0903	4.9981	4.9800	4.9377	4.8016
	$B(\hat{\beta}_2)$	-1.2601	-1.2568	-1.2617	-1.2563	-1.2529	-1.2642	-1.2534	-1.2690	-1.2509	-1.2174	-1.2468	-1.2517	-1.2494	-1.2338	-1.1746
$\tilde{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-1.9997	-1.9997	-1.9909	-1.9738	-1.9211	0.0006	0.0003	-0.0002	-0.0005	-0.0014	4.0903	4.9981	4.9800	4.9377	4.8016
	$B(\hat{\beta}_2)$	0.0024	-0.0051	-0.0561	-0.1581	-0.4572	-0.0040	0.0034	-0.0020	0.0044	-0.0004	0.0072	0.0102	0.1321	0.3827	1.1450
2 Stage estimators, Y_1 is not used as IV																
Pre-test	$B(\hat{\beta}_1)$	-2.0038	-1.9895	-1.8764	-1.6758	-1.2224	0.0006	0.0001	-0.0002	-0.0006	-0.0014	4.9949	4.9796	4.7030	4.1969	3.0564
	$B(\hat{\beta}_2)$	0.0026	-0.0052	-0.0547	-0.1451	-0.3436	-1.2360	-1.2204	-1.2406	-1.2218	-1.1913	0.0087	0.0096	0.1262	0.3447	0.8576
2 Stage estimators, Y_1 is used as IV																
Pre-test	$B(\hat{\beta}_1)$	-2.0038	-1.9895	-1.8764	-1.6758	-1.2224	0.0006	0.0001	-0.0002	-0.0006	-0.0014	4.9949	4.9796	4.7030	4.1969	3.0564
	$B(\hat{\beta}_2)$	0.0026	-0.0052	-0.0547	-0.1451	-0.3436	-0.0040	0.0034	-0.0020	0.0044	-0.0004	0.0087	0.0096	0.1262	0.3447	0.8576

Table 4.6. Bias of OLS, 2SLS, Partial IV and Pretest-estimators: moderate IV for Y_2 ($\mu_{22} = 613$)

$H_0 : \delta_1 = 0$, moderate IV for $Y_2 : \mu_{22} = 613$																
	$\delta_1 \rightarrow$	-2					0					5				
	$k_2 \downarrow$ $\mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
OLS	$B(\hat{\beta}_1)$	-1.9993	-1.9988	-1.9927	-1.9761	-1.9211	-0.0007	0.0007	0.0011	0.0003	-0.0008	4.9994	4.9992	4.9802	4.9406	4.8050
	$B(\hat{\beta}_2)$	0.0003	0.0011	-0.0019	-0.0098	-0.0331	0.0004	-0.0002	0.0005	-0.0001	0.0001	-0.0004	0.0000	0.0067	0.0246	0.0823
2SLS	$B(\hat{\beta}_1)$	-2.0037	-1.9939	-1.8897	-1.6903	-1.2341	-0.0022	-0.0005	0.0041	0.0028	0.0036	5.0038	4.9840	4.7242	4.2268	3.0917
	$B(\hat{\beta}_2)$	-0.0003	0.0038	-0.0402	-0.1136	-0.2949	-0.0024	0.0024	-0.0028	-0.0030	-0.0056	-0.0033	0.0025	0.0974	0.2877	0.7326
$\hat{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-1.9993	-1.9988	-1.9927	-1.9761	-1.9211	-0.0007	0.0007	0.0011	0.0003	-0.0008	4.9994	4.9992	4.9802	4.9406	4.8050
	$B(\hat{\beta}_2)$	-1.1472	-1.1482	-1.1447	-1.1493	-1.1391	-1.1527	-1.1386	-1.1507	-1.1549	-1.1399	-1.1372	-1.1477	-1.1555	-1.1541	-1.1275
$\tilde{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-1.9993	-1.9988	-1.9927	-1.9761	-1.9211	-0.0007	0.0007	0.0011	0.0003	-0.0008	4.9994	4.9992	4.9802	4.9406	4.8050
	$B(\hat{\beta}_2)$	0.0003	0.0041	-0.0411	-0.1267	-0.3976	-0.0026	0.0018	-0.0027	-0.0038	-0.0050	-0.0027	0.0028	0.1005	0.3228	0.9918
2 Stage estimators, Y_1 is not used as IV																
Pre-test	$B(\hat{\beta}_1)$	-2.0037	-1.9939	-1.8897	-1.6903	-1.2341	-0.0007	0.0007	0.0011	0.0003	-0.0007	5.0038	4.9840	4.7242	4.2268	3.0917
	$B(\hat{\beta}_2)$	-0.0003	0.0038	-0.0402	-0.1136	-0.2949	-1.1251	-1.1122	-1.1281	-1.1314	-1.1134	-0.0033	0.0025	0.0974	0.2877	0.7326
2 Stage estimators, Y_1 is used as IV																
Pre-test	$B(\hat{\beta}_1)$	-2.0037	-1.9939	-1.8897	-1.6903	-1.2341	-0.0007	0.0007	0.0011	0.0003	-0.0007	5.0038	4.9840	4.7242	4.2268	3.0917
	$B(\hat{\beta}_2)$	-0.0003	0.0038	-0.0402	-0.1136	-0.2949	-0.0026	0.0018	-0.0027	-0.0038	-0.0050	-0.0033	0.0025	0.0974	0.2877	0.7326

Table 4.7. Bias of OLS, 2SLS, Partial IV and Pretest-estimators: strong IV for Y_2 ($\mu_{22} = 2000$)

$H_0 : \delta_1 = 0$, strong IV for $Y_2 : \mu_{22} = 2000$																
	$\delta_1 \rightarrow$	-2					0					5				
	$k_2 \downarrow$ $\mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
OLS	$B(\hat{\beta}_1)$	-2.0004	-2.0003	-1.9913	-1.9753	-1.9200	-0.0006	-0.0011	-0.0002	0.0000	0.0004	5.0009	4.9977	4.9766	4.9349	4.8013
	$B(\hat{\beta}_2)$	-0.0011	0.0004	-0.0052	-0.0126	-0.0369	0.0013	0.0000	0.0006	0.0007	-0.0010	-0.0001	0.0018	0.0115	0.0314	0.0915
2SLS	$B(\hat{\beta}_1)$	-2.0016	-1.9957	-1.8858	-1.6913	-1.2484	0.0028	-0.0001	0.0009	-0.0039	0.0016	4.9978	4.9722	4.7084	4.2261	3.1281
	$B(\hat{\beta}_2)$	0.0003	-0.0034	-0.0483	-0.1144	-0.2679	0.0066	0.0013	-0.0001	0.0019	0.0000	-0.0016	0.0223	0.1197	0.2879	0.6578
$\hat{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0004	-2.0003	-1.9913	-1.9753	-1.9200	-0.0006	-0.0011	-0.0002	0.0000	0.0004	5.0009	4.9977	4.9766	4.9349	4.8013
	$B(\hat{\beta}_2)$	-0.9312	-0.9343	-0.9294	-0.9208	-0.9191	-0.9305	-0.9281	-0.9123	-0.8978	-0.8637	-0.9438	-0.9267	-0.8857	-0.8491	-0.7421
$\tilde{\beta}_{PIV}$	$B(\hat{\beta}_1)$	-2.0004	-2.0003	-1.9913	-1.9753	-1.9200	-0.0006	-0.0011	-0.0002	0.0000	0.0004	5.0009	4.9977	4.9766	4.9349	4.8013
	$B(\hat{\beta}_2)$	0.0000	-0.0027	-0.0505	-0.1275	-0.3625	0.0054	0.0015	-0.0004	0.0026	-0.0006	-0.0012	0.0212	0.1236	0.3245	0.8940
2 Stage estimators, Y_1 is not used as IV																
Pre-test	$B(\hat{\beta}_1)$	-2.0016	-1.9957	-1.8858	-1.6913	-1.2484	-0.0005	-0.0011	-0.0002	-0.0001	0.0004	4.9978	4.9722	4.7084	4.2261	3.1281
	$B(\hat{\beta}_2)$	0.0003	-0.0034	-0.0483	-0.1144	-0.2679	-0.9088	-0.9074	-0.8938	-0.8764	-0.8441	-0.0016	0.0223	0.1197	0.2879	0.6578
2 Stage estimators, Y_1 is used as IV																
Pre-test	$B(\hat{\beta}_1)$	-2.0016	-1.9957	-1.8858	-1.6913	-1.2484	-0.0005	-0.0011	-0.0002	-0.0001	0.0004	4.9978	4.9722	4.7084	4.2261	3.1281
	$B(\hat{\beta}_2)$	0.0003	-0.0034	-0.0483	-0.1144	-0.2679	0.0054	0.0015	-0.0004	0.0026	-0.0006	-0.0016	0.0223	0.1197	0.2879	0.6578

Table 4.8. MSE of OLS, 2SLS, Partial IV and Pretest-estimators)

$H_0: \delta_1 = 0$, irrelevant IV for $Y_2: \mu_{22} = 0$															
$\delta_1 \rightarrow$	-2					0					5				
$k_2 \downarrow \ \mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
OLS	4.0158	4.0104	3.9802	3.9141	3.7060	0.0108	0.0110	0.0107	0.0108	0.0106	25.0205	24.9862	24.7962	24.3947	23.1036
2SLS	4.1838	4.1394	3.7472	3.0455	1.8295	0.1934	0.1925	0.1769	0.1669	0.1401	25.2098	25.0122	22.4269	18.1402	10.7795
$\hat{\beta}_{PIV}$	5.9270	5.9184	5.8780	5.7943	5.5952	1.8996	1.9029	1.9009	1.8981	1.9546	26.9317	26.8942	26.7879	26.5446	25.7733
$\tilde{\beta}_{PIV}$	4.0883	4.0792	4.0618	4.0314	4.0310	0.0822	0.0830	0.0780	0.0778	0.0717	25.0935	25.0613	24.9415	24.7847	24.8728
2 Stage estimators, Y_1 is not used as IV															
Pre-test	4.1838	4.1394	3.7472	3.0455	1.8295	1.8205	1.8113	1.8243	1.8227	1.8711	25.2098	25.0122	22.4269	18.1402	10.7795
2 Stage estimators, Y_1 used as IV															
Pre-test	4.1838	4.1394	3.7472	3.0455	1.8295	0.0823	0.0831	0.0780	0.0778	0.0717	25.2098	25.0122	22.4269	18.1402	10.7795
$H_0: \delta_1 = 0$, weak IV for $Y_2: \mu_{22} = 13$															
$\delta_1 \rightarrow$	-2					0					5				
$k_2 \downarrow \ \mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
OLS	4.0095	4.0095	3.9749	3.9072	3.7032	0.0107	0.0110	0.0108	0.0108	0.0106	25.0099	24.9916	24.8117	24.3945	23.0794
2SLS	4.2076	4.1475	3.7253	3.0451	1.8359	0.1869	0.1885	0.1823	0.1688	0.1391	25.1317	24.9929	22.4401	18.2182	10.7699
$\hat{\beta}_{PIV}$	5.8942	5.9044	5.8868	5.7980	5.6020	1.9238	1.9030	1.9436	1.9219	1.9482	26.8757	26.8701	26.8039	26.5555	25.8166
$\tilde{\beta}_{PIV}$	4.0799	4.0806	4.0551	4.0244	4.0268	0.0804	0.0816	0.0802	0.0801	0.0714	25.0775	25.0667	24.9537	24.7674	24.8193
2 Stage estimators, Y_1 is not used as IV															
Pre-test	4.2076	4.1475	3.7253	3.0451	1.8359	1.8418	1.8075	1.8605	1.8365	1.8682	25.1317	24.9929	22.4401	18.2182	10.7699
2 Stage estimators, Y_1 used as IV															
Pre-test	4.2076	4.1475	3.7253	3.0451	1.8359	0.0804	0.0816	0.0803	0.0801	0.0715	25.1317	24.9929	22.4401	18.2182	10.7699

Table 4.8. MSE of OLS, 2SLS, Partial IV and Pretest-estimators)

$H_0: \delta_1 = 0$, moderate IV for $Y_2: \mu_{22} = 613$															
$\delta_1 \rightarrow$	-2					0					2				
$k_2 \downarrow$ $\mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
OLS	4.0076	4.0058	3.9818	3.9161	3.7031	0.0108	0.0106	0.0106	0.0105	0.0106	25.0050	25.0024	24.8140	24.4228	23.1107
2SLS	4.1880	4.1519	3.7601	3.0700	1.8272	0.1762	0.1761	0.1700	0.1541	0.1271	25.2137	25.0218	22.6110	18.3870	10.7360
$\hat{\beta}_{PIV}$	5.6071	5.6193	5.5822	5.5281	5.3035	1.6384	1.6019	1.6388	1.6641	1.7111	26.5858	26.6183	26.5250	26.2964	25.5173
$\tilde{\beta}_{PIV}$	4.0687	4.0690	4.0514	4.0132	3.9700	0.0726	0.0720	0.0717	0.0679	0.0641	25.0662	25.0652	24.9312	24.7165	24.4657
2 Stage estimators, Y_1 is not used as IV															
Pre-test	4.1880	4.1519	3.7601	3.0700	1.8272	1.5637	1.5309	1.5773	1.5994	1.6349	25.2137	25.0218	22.6110	18.3870	10.7360
2 Stage estimators, Y_1 used as IV															
Pre-test	4.1880	4.1519	3.7601	3.0700	1.8272	0.0727	0.0721	0.0718	0.0680	0.0641	25.2137	25.0218	22.6110	18.3870	10.7360
$H_{\delta_1}: \delta_1 = 0$, strong IV for $Y_2: \mu_{22} = 2000$															
$\delta_1 \rightarrow$	-2					0					5				
$k_2 \downarrow$ $\mu_1 \rightarrow$	0	13	200	613	2000	0	13	200	613	2000	0	13	200	613	2000
OLS	4.0125	4.0119	3.9761	3.9126	3.6990	0.0106	0.0107	0.0107	0.0106	0.0104	25.0198	24.9876	24.7781	24.3669	23.0764
2SLS	4.1631	4.1400	3.7179	3.0437	1.8094	0.1531	0.1533	0.1483	0.1407	0.1118	25.1320	24.8839	22.4256	18.3016	10.7419
$\hat{\beta}_{PIV}$	5.1144	5.1233	5.0830	5.0029	4.7951	1.1179	1.1140	1.0902	1.0722	1.0660	26.1523	26.0964	25.8652	25.5162	24.4490
$\tilde{\beta}_{PIV}$	4.0604	4.0600	4.0280	3.9864	3.9030	0.0583	0.0569	0.0563	0.0556	0.0505	25.0663	25.0376	24.8675	24.5981	24.1063
2 Stage estimators, Y_1 is not used as IV															
Pre-test	4.1631	4.1400	3.7179	3.0437	1.8094	1.0688	1.0671	1.0486	1.0240	1.0203	25.1320	24.8839	22.4256	18.3016	10.7419
2 Stage estimators, Y_1 used as IV															
Pre-test	4.1631	4.1400	3.7179	3.0437	1.8094	0.0583	0.0570	0.0564	0.0556	0.0505	25.1320	24.8839	22.4256	18.3016	10.7419

References

Abdelkhalek, T., Dufour, J.-M., 1998. Statistical inference for computable general equilibrium models with applications to a model of the moroccan economy. Review of Economics and Statistics, pp. 520–534.

Altonji, J., 1986. Intertemporal substitution in labour supply: Evidence from mico data. Journal of Political Economy 94, S176–215.

Anderson, T. W., 2003. An Introduction to Multivariate Statistical Analysis third edn. John Wiley & Sons, New York.

Anderson, T. W., Kadane, J. B., 1977. A comment on the tes of overidentifying restrictions. Econometrica 45, 1027–1031.

Anderson, T. W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46–63.

Andrews, D. W. K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D. W. K., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60(4), 953–966.

Angrist, J. D., al., 1999. Jackknife instrumental variables estimates. Journal of Applied Econometrics 14, 57–67.

Angrist, J. D., Krueger, A. B., 1991. Does compulsory school attendance affect schooling and earning?. Quarterly Journal of Economics CVI, 979–1014.

Angrist, J. D., Krueger, A. B., 1995. Split-sample instrumental variables estimates of the return to schooling. Journal of Business and Economic Statistics 13, 225–235.

Ashley, R. , 2006. Assessing the credibilility of instrumental variables inference with imperfect instruments via sensitivity analysis. Technical report, Economics Department, Virginia Polytechnic Institute Blacksburg, Virginia.

Barro, R., 1977. Unanticipated money growth and unemployment in united states. American Economic Review 67, 101–115.

Bekker, P., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62, 657–681.

Bound, J. e. a., 1995. Problem with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. Journal of the American Statistical Association 90, 443–450.

Bradford, W. D. , 2003. Pregnancy and the demand for cigarettes. American Economic Review 93, 1752–1763.

Doko, F., Dufour, J.-M., 2009a. Exogeneity tests, non gaussian distributions and weak identification : finite-sample and asymptotic distributional theory. Technical report, C.R.D.E., Université de Montréal.

Doko, F., Dufour, J.-M., 2009b. Identification-robust inference for error-regressors covariance and partial iv regression. Technical report, C.R.D.E., Université de Montréal.

Doko, F., Dufour, J.-M., 2009c. Wald-type tests for error-regressors covariances, partial exogeneity tests and partial iv estimation. Technical report, C.R.D.E., Université de Montréal.

Donald, S. G., Newey, W. K., 2001. Choosing the number of instruments. Econometrica 69, 1161–1191.

Dufour, J.-M., 1987. Linear Wald methods for inference on covariances and weak exogeneity tests in structural equations. In: I. B. MacNeill, G. J. Umphrey, eds, Advances in the Statistical Sciences: Festschrift in Honour of Professor V.M. Joshi's 70th Birthday. Volume III, Time Series and Econometric Modelling. D. Reidel, Dordrecht, The Netherlands, pp. 317–338.

Dufour, J.-M., 1990. Exact tests and confidence sets in linear regressions with autocorrelated errors. Econometrica 58, 475–494.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics, with applications to structural and dynamic models. Econometrica 65, 1365–1389.

Dufour, J.-M., 2003. Identification, weak instruments and statistical inference in econometrics. Canadian Journal of Economics 36(4), 767–808.

Dufour, J.-M., 2006. Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. Journal of Econometrics 138, 2649–2661.

Dufour, J.-M., Hsiao, C., 2008. Identification. In:. L. E. Blume, S. N. Durlauf, eds, The New Palgrave Dictionary of Economics 2nd edn, Palgrave Macmillan, Basingstoke, Hampshire, England. forthcoming.

Dufour, J.-M., Jan, K., 1998. Exact inference methods for first-order autoregressive distributed lag models. Econometrica 66, 9–104.

Dufour, J.-M., Jasiak, J., 2001. Finite sample limited information inference methods for structural equations and models with generated regressors. International Economic Review 42, 815–843.

Dufour, J.-M., Khalaf, L., 2002. Simulation based finite and large sample tests in multivariate regressions. Journal of Econometrics 111(2), 303–322.

Dufour, J.-M., Khalaf, L., 2006. Inflation dynamics and the new keynesian phillips curve: an identification robust econometric analysis. JEDC 30(9).

Dufour, J.-M., Khalaf, L., Beaulieu, M.-C., 2008. Multivariate residual-based finite-sample tests for serial dependence and GARCH with applications to asset pricing models. Journal of Applied Econometrics forthcoming.

Dufour, J.-M., Taamouti, M., 2003. Point-optimal instruments and generalized Anderson-Rubin procedures for nonlinear models. Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M., Taamouti, M., 2005. Projection-based statistical inference in linear structural models with possibly weak instruments. Econometrica 73(4), 1351–1365.

Dufour, J.-M., Taamouti, M., 2006. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. Journal of Econometrics 139(1), 133–153.

Durbin, J., 1954. Errors in variables. Review of the International Statistical Institute 22, 23–32.

Frankel, J. A., Romer, D., 1999. Does trade cause growth?. American Economic Review 89(3), 379–399.

Gallant, A. R., ed., 1987. Nonlinear Statistical Model. New York: Wiley, New York, New York.

Guggenberger, P., 2008. The impact of a Hausman pretest on the size of the hypothesis tests. forthcoming. Technical report, Department of Economics, UCLA Los Angeles, California.

Hahn, J., Ham, J., Moon, H. R., 2008. The Hausman test and weak instruments. Technical report, Department of Economics, UCLA Los Angeles, California.

Hall, A. R., Peixe, F. P. M., 2003. A consistent method for the selection of relevant instruments. Econometric Reviews 2(3), 269–287.

Hall, A. R., Rudebusch, G. D., Wilcox, D. W., 1996. Judging instrument relevance in instrumental variables estimation. International Economic Review 37, 283–298.

Ham, J. C., Kevin, T. R., 2002. Testing intertemporal substitution, implicit contracts, and hours restriction models of labour market using micro data. American Economic Review 92, 905–927.

Hansen, C., Hausman, J., Newey, W., 2008. Estimation with many instrumental variables. Journal of Business and Economic Statistics 26, 398–422.

Harrison, A., 1996. Oponness and growth: a time-series, cross-country analysis for developing countries. 48, 419–447.

Harville, D. A., 1997. Matrix Algebra from a Statistician's Perspective. Springer-Verlag, New York.

Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–1272.

Hausman, J., Hahn, J., 2002. A new specification test for the validity of the instrumental variables. Econometrica 70, 163–189.

Irwin, A.-D., Tervio, M., 2002. Does trade raise income? evidence from twentieth century. Journal of International Economics 58, 1–18.

Kiviet, J. F., Niemczyk, J., 2006. On the limiting and empirical distribution of IV estimators when some of the instruments are invalid. Technical report, Department of Quantitative Economics, University of Amsterdam Amsterdam, The Netherlands.

Kiviet, J. F., Niemczyk, J., 2007. The asymptotic and finite-sample distributions of ols and simple iv in simultaneous equations. Computational Statistics and Data Analysis 51, 3296–3318.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70(5), 1781–1803.

Levine, D., 1983. A remark on serial correlation in maximum likelihood. JOE 23, 337–342.

MaCurdy, T., 1981. An empirical model of labour supply in a life-cycle setting. JPE 89, 1059–1085.

Magnus, J. R., Neudecker, H., 1998. Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised Edition. John Wiley & Sons, New York.

Mankiw, G. N., al., 1992. A contribution to the empirics of economic growth. The Quarterly Journal of Economics 107(2), 407–437.

Meeker, W. Q., al., 1998. Accelaration degradation tests: Modeling and analysis. Technometrics 40(2), 89–99.

Montmarquette, C., Mahseredjian, S., 1989. Could teacher grading practices account for unexplained variation in school achievements?. European Economic Review 8, 335–343.

Moreira, M. J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71(4), 1027–1048.

Nerlove, ed., 1963. Returns to Scale in Electricity Supply. Stanford: Stanford University Press, New York, New York.

Newey, W., West, K., 1987. A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Phillips, P. C. B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.

Rao, T., Mitra, C., eds, 1971. Generalized Inverse of Matrices and its Applications. John Wiley & Sons, Inc. Hoboken, New Jersey.

Revankar, N. S., Hartley, M. J., 1973. An independence test and conditional unbiased predictions in the context of simultaneous equation systems. International Economic Review 14, 625–631.

Rodrik, D., 1995. Trade and industrial policy reform. Review of Economic Studies 17, 1–20.

Sargan, J., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26(3), 393–415.

Small, D. S., 2007. Sensitivity analisis for instrumental variables regression with overidentifying restrictions. Annals of Mathematical Statistics 102, 1049–1058.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica 65(3), 557–586.

Stock, J. H., Wright, J. H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20(4), 518–529.

Swanson, N. R., Chao, J. C., 2005. Notes and comments: Consistent estimation with a large number of weak instruments. Econometrica 73, 1673–1692.

Tobin, J., 1969. A general equilibrium approach to monetary theory. Journal of Money, Credit and Banking 1, 15–29.

Wang, J., Zivot, E., 1998. Inference on structural parameters in instrumental variables regression with weak instruments. Econometrica 66(6), 1389–1404.

White, H., Domowitz, I., 1984. Nonlinear regression with dependent observations. Econometrica 52, 143–161.

White, H., ed., 1984. Asymptotic Theory for Econometricians. New York: Academic Press, New York, New York.

Wu, D.-M., 1973. Alternative tests of independence between stochastic regressors and disturbances. Econometrica 41, 733–750.

Conclusion générale

Cette thèse étudie les effets de l'endogénéité et de la faiblesse des variables instrumentales sur les statistiques de tests robustes à l'identification et les tests d'exogénété et propose les procedures de sélection des instruments.

Dans le premier essai, nous analysons les effets de l'endogénéité des instruments sur la statistique d'Anderson et Rubin (AR) et celle de la statistique de Kleibergen (K), avec ou sans instruments faibles. Nous montrons que toutes ces procédures sont en général consistantes contre la présence d'instruments invalides (c'est-à-dire détectent la présence d'instruments invalides) indépendamment de leur qualité lorsque le paramètre qui contrôle l'endogénéité des instruments est fixe. Nous décrivons aussi des cas où cette consistance peut ne pas tenir, mais la distribution asymptotique est modifiée d'une manière qui pourrait conduire aux distorsions de niveau en grands échantillons. Ensuite, lorsque les instruments sont localement exogènes (c'est-à-dire le paramètre d'endogénéité converge vers zéro lorsque la taille de l'échantillon augmente), nous montrons que ces tests convergent vers des distributions de chi carré non centré, que les instruments soient forts ou faibles. Nous caractérisons aussi les situations où le paramètre de non centralité est nul et la distribution asymptotique des statistiques demeure la même que dans le cas des instruments valides (malgré la présence des instruments invalides).

Dans le deuxième essai, nous étudions l'impact des instruments faibles sur les tests d'exogénéité du type Durbin-Wu-Hausman (DWH) ainsi que le test de Revankar et Hartley (1973). Nous proposons une analyse en petit-et grand-échantillon de la distribution de ces tests sous l'hypothèse nulle et l'alternative, incluant les cas où les instruments faibles. Notre analyse en petit-échantillon founit plusieurs nouvelles perspectives et extensions des précédentes procédures. La caractérisation de la distribution de ces statistiques permet la construction des tests de Monte Carlo exacts pour l'exogénéité même avec les erreurs non Gaussiens. Nous montrons que ces tests sont typiquement robustes aux intruments faibles. De plus, nous fournissons une caractérisation de la puissance des tests, qui exhibe clairement les facteurs qui déterminent la puissance. Nous montrons que les tests n'ont pas de puissance lorsque tous les instruments sont faibles [similaire à Guggenbergen (2008)]. Cependant, la puissance existe tant qu'au moins un seul instruments est fort. Notre théorie asymptotique sous les hypothèses affaiblies confirme la théorie en échantillon fini.

Par ailleurs, nous présentons une analyse de Monte Carlo indiquant que: (1) L'estimateur des moindres carrés ordinaires est plus efficace que celui des doubles moindres carrés lorsque les instruments sont faibles et l'endogenéité modérée [conclusion similaire à celle de Kiviet and Niemczyk (2007)]; (2) les estimateurs pré-test basés sur les tests d'exogenété ont une excellente performance comparés aux doubles moindres carrés. Ce qui suggère que la méthode des variables instrumentales ne devrait être appliquée que si l'on a la certitude d'avoir les instruments fort. Nous illustrons nos résultats théoriques à travers deux applications empiriques: la relation entre le taux d'ouverture et la croissance économique et le problème bien connu de rendements à l'éducation.

Dans le troisième essai, nous étendons le test d'exogénéité de Wald généralisé proposé par Dufour (1987) aux cas où les erreurs de la régression ont une distribution non-normale. Nous proposons une nouvelle version du précédent test qui est valide même en présence d'erreurs non-Gaussiens. Contrairement aux tests de Durbin-Wu-Hausman, ce test permet de tester l'exogénéité partielle d'un sous ensemble de variables. Par ailleurs, nous proposons deux nouveaux estimateurs pré-test basés sur ce test. Nos analyses de Monte Carlo permettent de montrer que ces estimateurs sont plus efficaces que l'estimateur IV usuel lorsque les variables instrumentales sont faibles et l'endogénéité modérée. Nous montrons également que ce test peut servir de procédure de sélection de variables instrumentales. Nous illustrons nos résultats théoriques par deux applications empiriques: le modèle bien connu d'équation du salaire [Angist et Krueger (1991, 1999)] et les rendements d'échelle [Nerlove (1963)]. Nos résultats suggèrent que l'éducation de la mère expliquerait le décrochage de son fils et que l'output est une variable endogène dans l'estimation du coût de la firme et le prix du fuel en est un instrument valide pour l'output.

Dans le quatrième essai, nous apportons une solution à deux problèmes très importants en économétrie. D'abord, bien que le test de Wald initial ou étendu permette de construire les régions de confiance et de tester les restrictions linéaires sur les covariances, il suppose que les paramètres du modèle sont identifiés. Lorsque l'identification est faible, ce test n'est en général plus valide. Cet essai développe une procédure d'inférence robuste à l'identification (instruments faibles) pour construire des régions de confiance pour la matrices de covariances entre les erreurs de la régression et les variables explicatives (possiblement endogènes). Nous fournissons les expressions analytiques des régions de confiance et caractérisons les conditions nécessaires et suffisantes sous lesquelles ils sont bornés. La procédure proposée demeure valide même pour les petits échantillons et est aussi asymptotiquement robuste à l'hétéroscédasticité et l'autocorrélation des erreurs.

Ensuite, les résultats sont utilisés pour développer les tests d'exogénéité partielle robustes à l'identification. Les simulations Monte Carlo indiquent que ces tests contrôlent le niveau et ont de la puissance même si les instruments sont faibles. Ce qui nous permet de proposer une procédure valide de sélection de variables instrumentales même s'il y a un problème d'identification. La procédure de sélection des instruments est basée sur deux nouveaux estimateurs pré-test qui combinent l'estimateur IV usuel et les estimateurs IV partiels. Nos simulations montrent que: (1) les estimateurs IV partiels sont plus efficaces que l'estimateur IV usuel lorsque les instruments sont faibles et l'endogénéité modérée; (2) les estimateurs pré-test ont globalement une excellente performance comparés à l'estimateur IV usuel. Nous illustrons nos résultats théoriques par deux apllications empiriques: la relation entre le taux d'ouverture et la croissance économique et le modèle de rendements à l'éducation. Dans la première application, les études antérieures ont conclu que les instruments n'étaient pas trop faibles [Dufour et Taamouti (2007)] alors qu'ils le sont fortement dans la seconde [Bound (1995), Doko and Dufour (2009)]. Nos résultats montrent les régions de confiance non bornées pour la covariance dans le cas les instruments faibles.