Université de Montréal

Techniques d'inférence exacte dans les modèles structurels avec applications macroéconomiques

par

Mohamed Taamouti

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

Septembre 2001

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée:

Techniques d'inférence exacte dans les modèles structurels avec applications macroéconomiques

présentée par

Mohamed Taamouti

a été évaluée par un jury composé des personnes suivantes:

René Garcia : président-rapporteur

Jean-Marie Dufour : directeur de recherche

Benoit Perron : membre du jury

Eric Zivot : examinateur externe (University of Washington)

Roch Roy : représentant du doyen de la FES

Thèse acceptée le: 29 novembre 2001

Sommaire

Cette thèse a pour objectif de développer des méthodes d'inférence exacte dans les modèles structurels. Elle est motivée par les résultats reportés récemment dans la littérature économétrique concernant les problèmes des méthodes d'inférence usuelles en présence d'instruments faibles.

Dans le premier essai, nous étudions le problème de construction de régions de confiance pour des transformations du vecteur des paramètres inconnus dans un modèle à équations simultanées linéaires. Plusieurs tests proposés pour l'inférence en présence d'instruments faibles présentent l'inconvénient de ne tester que des hypothèses spécifiant le vecteur entier des paramètres. C'est le cas notamment de la statistique d'Anderson-Rubin (1949, Ann. Math. Stat.) et des deux statistiques proposées récemment par Wang et Zivot (1998, Econometrica). En principe, ce problème peut être résolu en utilisant la technique de projection [Dufour (1997, Econometrica), Dufour and Jasiak (2001, International Economic Review)]. Cependant, cette dernière exige, de façon générale, l'utilisation des méthodes numériques. Dans cet essai, nous fournissons une solution analytique complète au problème de construction d'intervalles de confiance par projection à partir de régions de confiance obtenues par inversion de statistiques de type Anderson-Rubin. Cette solution se base sur la théorie des quadriques et peut être perçue comme une extension des intervalles et des ellipsoïdes de confiance usuels. Les calculs requis sont semblables à ceux des moindres carrés ordinaires. Nous étudions également par des simulations Monte Carlo le degré de conservatisme des régions de confiance par projection. Enfin, nous illustrons les méthodes proposées par deux applications macroéconomiques: La relation entre le commerce international et la croissance économique et le problème des rendements d'échelle dans l'industrie américaine.

Dans le deuxième essai, nous proposons des méthodes d'inférence exacte dans un modèle structurel non linéaire. Nous développons une approche simple pour construire des tests exacts. Cette approche généralise celle de Hartley (1964, Biometrika) pour le test d'hypothèses et la construction de régions de confiance dans les régressions

non linéaires et la procédure d'Anderson-Rubin spécifique aux modèles structurels à équations linéaires. Nous abordons par la suite le problème d'optimalité des instruments utilisés dans le test. La notion d'optimalité réfère à la maximisation de la puissance du test proposé. Ceci contraste avec la notion d'optimalité typiquement considérée dans la littérature économétrique et qui réfère à la minimisation de la variance asymptotique d'un estimateur [Amemiya (1977, Econometrica)]. Les instruments optimaux que nous dérivons dépendent de l'hypothèse alternative, et donc nous qualifions ces derniers d'instruments "point-optimaux" [King (1988, Econometric Reviews)]. La matrice des instruments optimaux est inconnue de façon générale, nous proposons une méthode pour son estimation basée sur la technique du split-sample.

Dans le troisième essai, nous examinons le problème de sélection d'instruments dans le cas d'un modèle structurel linéaire où la méthode 2SLS pour l'estimation et/ou le test d'Anderson-Rubin sont utilisés. Nous nous concentrons sur le cas d'une seule variable explicative endogène. Nous commençons par analyser les déterminants de la performance de ces méthodes et comment la matrice des instruments affecte la qualité de l'estimation et de l'inférence. Nous proposons par la suite une méthode de sélection d'instruments sur la base de ces résultats. Cette méthode est basée sur la maximisation séquentielle du paramètre de concentration. Nous étudions si les méthodes de diagnostic et de sélection d'instruments sont utiles en pratique ou comme conclu par Hall, Rudebusch and Wilcox (1996, International Economic Review) sont inutilisables. Nous comparons également, par des simulations Monte Carlo, les principales méthodes de sélection d'instruments proposées dans la littérature.

Mots clés: Inférence exact; modèle structurel; endogeniété; variable instrumentale; pertinence; instrument faible; équations simultanées; région de confiance; projection; inférence simultanée; modèle non linéaire; test d'Anderson-Rubin; régression artificielle; instrument point-optimal ; technique du split-sample; estimateur des 2SLS; paramètre de concentration; sélection d'instruments; convergence.

Summary

The objective of this thesis is to develop exact inference methods in structural models. It is motivated by recent results on the problems of standard inference methods in the presence of weak instruments.

In the first essay, we consider the problem of building confidence sets for transformations of the vector of unknown parameters in linear simultaneous equations models. Several standard tests present the drawback of requiring the specification of the entire vector of unknown parameters. This is particularly the case of the Anderson-Rubin (1949, Ann. Math. Stat.) test and two statistics proposed recently by Wang et Zivot (1998, Econometrica). This problem may in principle be overcome by using projection techniques [Dufour (1997, Econometrica), Dufour and Jasiak (2001, International Economic Review)]. The latter, however, apparently can be implemented only by using costly numerical techniques. In this paper, we provide a complete close-form solution to the problem of building projection-based confidence sets from Anderson-Rubin-type confidence sets. The latter involves the geometric properties of "quadrics" and can be viewed as an extension of usual confidence intervals and ellipsoids. Only least squares techniques are required for building the confidence intervals. We also study by simulation how "conservative" projection-based confidence sets are. Finally, we illustrate the methods proposed by applying them in two different examples: the relationship between trade and growth in a cross-section of countries, and a study of production functions in the U.S. economy.

In the second essay, we consider exact statistical inference in nonlinear structural models. We propose an exact and simple test which may be inverted to obtain confidence sets for the vector of unknown parameters. This test generalizes the approach of Hartley (1964, Biometrika) for hypothesis testing and confidence sets construction in nonlinear regressions, and the Anderson-Rubin (1949, Ann. Math. Stat.) procedure in linear structural models. We also study the problem of building optimal instruments for testing purposes. The optimal instruments are those that maximize the power of the proposed test. This contrasts with the optimality of the instruments

which is well known in the econometric literature [Amemiya (1977, Econometrica)] and which refers to the minimization of the asymptotic variance of an estimator. Our optimal instruments depend on the alternative hypothesis, yielding **point-optimal instruments**. The matrix of optimal instruments is unknown in general, to estimate it, we use the split-sample technique [Dufour-Jasiak (1993)]. The exact distribution of the test is derived under the assumptions of independence and normality of the disturbances, but we show that the test remains valid asymptotically under weaker assumptions similar to those usually assumed to derive asymptotic distributions.

In the third essay, we examine the problem of selecting instruments in the case of a linear structural equation for estimating or testing. We concentrate on the case of 2SLS method for estimation and Anderson-Rubin (1949, Ann. Math. Stat.) test for inference. We first analyze what determines the performance of these methods and how the matrix of instruments affects the quality of estimation and inference. We then consider the problem of choosing a subset from a large set of potential instruments on the basis of these results in the case of one endogenous explanatory variable. The method we propose is based on the sequential maximization of the concentration parameter. It is simple to apply and is reliable even in small samples as is shown by Monte Carlo simulations. In large samples, the method is shown to be consistent in the sense that it selects with probability 1 only instruments with non zero coefficient in the true DGP of the endogenous explanatory variable.

Key words: Exact inference; structural model; endogeneity; instrumental variable; relevance; weak instrument; simultaneous equations; confidence set; projection; simultaneous inference; nonlinear model; Anderson-Rubin test; artificial regression; point-optimal instrument; split-sample technique; 2SLS estimation; concentration parameter; selection of instruments; consistency.

Contents

List of Tables

List of Figures

Remerciements

Cette thèse n'aurait pu aboutir sans l'aide de plusieurs personnes et institutions. En premier lieu, je tiens à remercier mon directeur de recherche Jean-Marie Dufour pour tout ce qu'il m'a appris en termes de savoir et de rigueur scientifique.[a] Je m'estime chanceux d'avoir côtoyé un scientifique de son calibre. J'ai également apprécié en lui sa générosité et ses qualités humaines extraordinaires.

Je voudrais remercier aussi le directeur de l'I.N.S.E.A. pour les facilités qu'il m'a accordées pour mon détachement. Le financement de ce projet de recherche a été assuré par le programme canadien des bourses de la francophonie (P.C.B.F.), le centre de recherche et développement en économique (C.R.D.E.), et le centre interuniversitaire de recherche en analyse des organisations (C.I.R.A.N.O.).

Je voudrais également remercier Craig Burnside pour m'avoir fourni les données qui ont servi aux applications empiriques dans cette thèse. Le papier a aussi bénéficié des commentaires des participants aux ateliers du C.R.D.E. (mai, 2000, 2001), aux congrès de la Société Canadienne de Sciences Economiques (2000, 2001), au congrès du Canadian Economics Association (2001), et au North American Summer Meeting of the Econometric Society (2001). En particulier, je remercie Lynda Khalaf pour ses commentaires sur les essais 1 et 2 de cette thèse.

J'ai également bénéficié de conseils et commentaires de plusieurs autres personnes, notamment, Touhami Abdelkhalek, Abdeljelil Farhat, Denis Pelletier et Gabriel Rodriguez.

Enfin, un gros merci à Fatima-Zohra pour le soutien moral.

[a] Les essais de cette thèse ont été rédigés conjointement avec Jean-Marie Dufour.

Introduction

Un des problèmes classiques en économétrie consiste à faire l'inférence statistique sur les paramètres d'un modèle structurel. Ce type de modèle fait intervenir de façon générale des variables explicatives qui sont endogènes, c'est à dire liées avec le terme d'erreur dans le modèle. Cette endogeniété nécessite l'utilisation de variables instrumentales (instruments) pour assurer l'identification des paramètres du modèle et la faisabilité de l'estimation et de l'inférence statistique. L'approche qui consiste à "instrumenter les variables endogènes" est communément connue sous le nom de la méthode des variables instrumentales (IV), c'est l'une des méthodes les plus utilisées dans la recherche empirique en économie. Un instrument doit vérifier deux principales conditions. La première c'est **l'exogeniété**: il doit être non corrélé avec le terme d'erreur du modèle. La deuxième est **la pertinence**: il doit être corrélé avec les variables explicatives endogènes du modèle.

Récemment, le problème de la modélisation structurelle a connu un important regain d'intérêt. Cet intérêt est dû au fait que les instruments utilisés dans la recherche empirique sont assez souvent faiblement corrélés avec les variables explicatives endogènes du modèle et donc violent ou presque la deuxième condition: nous parlons alors d'**instruments faibles**. La conséquence de ceci est que les paramètres sont presque non identifiés. La littérature sur le problème des instruments faibles est considérable, voir par exemple Nelson-Startz (1990,1990b), Buse (1992), Maddala-Jeong (1992), Bound-Jaeger-Baker (1993,1995), Angrist-Krueger (1995), Hall-Rudebusch-Wilcox (1996), Dufour (1997), Shea (1997), Staiger-Stock (1997), Wang-Zivot (1998), Zivot-Startz-Nelson (1998), Startz-Nelson-Zivot (1999), Perron (1999), Stock-Wright (2000), Dufour-Jasiak (2001), et Kleibergen (2001a, 2001b).

Dans ce contexte, plusieurs articles ont documenté, par des simulations et par des approximations asymptotiques, la faible performance des méthodes et approximations asymptotiques usuelles [Nelson-Startz (1990,1990b), Buse (1992), Bound-Jaeger-Baker (1993,1995), Hall-Rudebusch-Wilcox (1996), Staiger-Stock (1997), Zivot-Startz-Nelson (1998), Dufour-Jasiak (2001)]. La principale difficulté dans ce cas est

la distribution des statistiques usuelles dans les échantillons finis. En particulier, les tests usuels deviennent très sensibles aux paramètres de nuisance. D'un autre coté, les distributions limites sont non standards quand les conditions d'identification ne sont pas vérifiées, et les approximations basées sur les grandes tailles d'échantillon ne convergent pas de façon uniforme et peuvent être arbitrairement imprécises dans les petits échantillons même si l'identification des paramètres est vérifiée et les conditions de régularité usuelles sont vérifiées. Le fait que la théorie asymptotique standard peut être arbitrairement imprécise dans les échantillons finis (de n'importe quelle taille) est démontré de façon rigoureuse dans Dufour (1997) où il a été observé qu'une région de confiance valide doit être non bornée avec une probabilité positive et que les statistiques de type Wald ont des distributions qui peuvent dévier de façon arbitraire de leur distribution asymptotique (même quand l'identification est vérifiée). Le fait que les distributions à distance finie et asymptotiques soient caractérisées par une forte dépendance des paramètres de nuisance a été aussi démontré en théorie distributionnelle exacte [voir Choi-Phillips (1992)] et en théorie asymptotique sous l'hypothèse de non identification locale [voir Staiger-Stock (1997) et Wang-Zivot (1998)].

Comme résultat, il apparaît important en présence de ce type de problèmes aussi bien de construire des tests et des régions de confiance proprement pivotales ou bornés par des statistiques pivotales, que d'étudier les procédures d'inférence dans une perspective de distance finie. L'exigence que les tests doivent être basés sur des statistiques pivotales ou dont les distributions sont bornables et que les régions de confiance soient construites à partir de statistiques pivotales est une condition de base de la théorie statistique [voir Lehmann (1986)].

Les recherches sur le problème des instruments faibles ont tout d'abord privilégié le développement de méthodes de diagnostic et de tests de la pertinence des instruments, ceci dans l'objectif de juger de la validité des instruments et de l'opportunité d'utiliser les méthodes d'inférence et d'estimation usuelles. Dans les modèles à équations simultanées linéaires, Bound-Jaeger-Baker (1995), en se basant sur les résultats de Buse (1992), ont montré que si un ensemble d'instruments est faiblement corrélé avec les variables explicatives endogènes, même une faible corrélation entre les instru-

ments et le terme d'erreur peut conduire à des biais importants dans les estimations par IV par rapport à la méthode des moindres carrés ordinaires et ceci même pour de grandes tailles d'échantillon. Ils suggèrent d'utiliser le R^2 partiel et la statistique de Fisher pour les instruments exclus de l'équation structurelle comme indicateur de la qualité de l'estimation par IV. Dans le même contexte, Hall-Rudebusch-Wilcox (1996) proposent la corrélation canonique entre les instruments et les variables explicatives endogènes comme mesure de la pertinence des instruments. Cependant leurs simulations montrent qu'une telle mesure a peu de mérite en pratique puisque, lorsqu'elle est utilisée comme test préliminaire elle conduit à une détérioration des propriétés de l'estimation par IV. Shea (1997) propose une autre mesure de la pertinence, il s'agit de la corrélation partielle. Cette mesure est calculée pour chaque variable endogène et se base sur le fait que pour qu'un ensemble d'instruments soit pertinent il faut pour chaque variable explicative endogène au moins un instrument qui contient un pouvoir explicatif de cette variable indépendant des composantes pertinentes pour les autres variables endogènes. Cette mesure coïncide avec le R^2 usuel dans le cas d'une seule variable explicative endogène mais dans le cas de plusieurs variables explicatives endogènes, elle permet de faire le diagnostic de façon séparée pour chacune d'elles. Dans le cadre des modèles non linéaires estimés par la méthode des moments généralisés (GMM), Wright (2000) propose un test pour l'identification du modèle en utilisant les tests de rang comme dans Cragg-Donald (1993). Ce test coïncide dans le cas d'un modèle structurel à équations linéaires avec le test proposé par Cragg-Donald (1993). Dans le cas des modèles non linéaires, la statistique du test dépend de paramètres de nuisance et l'auteur utilise l'inégalité de Bonferroni pour contourner ce problème et obtient donc un test conservateur.

Le développement de nouvelles méthodes d'inférence valides dans le cas de problèmes d'instruments faibles constitue un des principaux sujets de la recherche actuelle en économétrie. Les méthodes proposées jusqu'à date sont principalement spécifiques aux modèles structurels à équations linéaires. L'une de ces méthodes consiste à utiliser [Dufour (1997), Staiger-Stock (1997)] une vieille statistique proposée par Anderson-Rubin (1949). Cette statistique suit une distribution de Fisher sous

l'hypothèse d'indépendance et de normalité des perturbations et reste asymptotiquement valide sous des hypothèses standard plus faibles. L'inversion de cette statistique permet d'obtenir des régions de confiance valides pour le vecteur des paramètres du modèle. Wang-Zivot (1998) propose sous l'hypothèse de non identification locale deux statistiques alternatives. La première est de type LR et la deuxième est de type LM. Ces deux statistiques sont asymptotiquement distribuées selon une loi khi-deux lorsque le modèle est exactement identifié (nombre d'instruments égal au nombre de variables explicatives endogènes) et sont bornées par une distribution khi-deux dans le cas de sur-identification (nombre d'instruments supérieur au nombre de variables explicatives endogènes) et conduisent donc à des tests et des régions de confiance conservateurs. Startz-Nelson-Zivot (1999) propose une procédure d'inférence "améliorée", le "S-test" qui peut être inversé pour construire des régions de confiance basées sur le test conjoint de l'hypothèse structurelle et de la condition d'identification. Kleibergen (2001a) propose une modification de la statistique AR pour obtenir une statistique pivotale dont la distribution asymptotique est une khi-deux avec un nombre de degrés de liberté égal au nombre de variables explicatives endogènes. Ceci permet d'améliorer la puissance du test AR qui diminue lorsque le nombre d'instruments non pertinents augmente. Enfin dans les modèles non linéaires estimés par GMM, Stock-Wright (2000), propose une statistique de test robuste dans le cas d'instruments faibles. Cette statistique est basée sur le critère de la méthode GMM et son inversion permet d'obtenir des régions de confiance (S-sets) asymptotiquement valides.

Cette thèse considère à nouveau le problème des instruments faibles et particulièrement l'invalidité des méthodes asymptotiques usuelles dans ce cas. Notre premier objectif est de développer des méthodes d'inférence exacte valides même en cas de problèmes d'identification. Les méthodes exactes exigent généralement des hypothèses relativement fortes que celles sous-jacentes à la théorie asymptotique standard, mais elles ont l'avantage d'être robustes aux problèmes comme celui des instruments faibles. En plus les méthodes exactes restent généralement valides asymptotiquement sous des hypothèses plus faibles similaires à celle de la théorie asymptotique. La thèse est constituée de 3 essais portant sur des thèmes différents de l'inférence structurelle.

Le premier essai traite de l'inférence statistique dans les modèles structurels linéaires. Le problème considéré est celui de l'inférence sur des transformations du vecteur des paramètres inconnus (par exemple sur des composantes du vecteur ou des combinaisons linéaires de ces composantes). Ce problème est important parce que, dans plusieurs cas, les méthodes usuelles permettent de tester des hypothèses spécifiant le vecteur entier des paramètres inconnus mais ne permettent pas de faire de l'inférence sur des transformations de ce vecteur. Ce défaut caractérise en particulier le test d'Anderson-Rubin ainsi que plusieurs tests robustes aux problèmes des instruments faibles proposés récemment dans la littérature [Wang-Zivot (1998), Kleibergen (2001a), Stock-Wright (2000)]. Une solution générale à ce problème est la technique de projection décrite dans Dufour (1990, 1997). Cette technique a été exploitée auparavant par Abdelkhalek-Dufour (1998) et Dufour-Kiviet (1998) dans des contextes différents et a donné des résultats satisfaisants. Cependant de façon générale, le calcul des projections est compliqué et nécessite l'utilisation de méthodes numériques.

Dans cet essai, nous dérivons une solution analytique complète au problème de construction d'intervalles de confiance basés sur la technique de projection à partir de régions de confiance obtenues par inversion de statistiques de type AR. La solution est fournie pour les intervalles de confiance pour des composantes du vecteur des paramètres inconnus et pour des combinaisons linéaires de ces composantes. Avant de calculer les projections, nous étudions les caractéristiques des régions de confiance basées sur l'inversion des statistiques de type AR sur la base de la géométrie des quadriques [voir Shilov (1961) et Pettofrezzo-Marcoantonio (1970)]. Cette solution qui peut être vue comme une extension des intervalles et des ellipsoïdes de confiance usuels, rend la technique de projection très attrayante, surtout dans des situations où aucune autre solution n'a été proposée antérieurement. Nous montrons aussi que les intervalles de confiance par projection peuvent être interprétés comme des intervalles de type Wald basés sur des estimateurs de type k-class [voir Davidson-MacKinnon (1993)] où les écarts-type sont corrigés en tenant compte du niveau de signification du test. Nous discutons également la relation entre ces intervalles de confiances et les

régions de confiance "à la Scheffé" largement utilisées en analyse de variance.

Dans le deuxième essai, nous considérons le problème d'inférence statistique dans un modèle structurel non linéaire. Nous proposons une approche simple pour construire des tests statistiques exacts. Cette approche basée sur des régressions artificielles généralise l'approche de Hartley (1964) pour les tests d'hypothèses et la construction de régions de confiance dans les régressions non linéaires et la procédure d'Anderson-Rubin spécifique aux modèles structurels à équations linéaires. La première généralisation vise à inclure des variables explicatives endogènes tandis que la seconde considère des modèles non linéaires. Nous abordons ensuite le problème d'optimalité des instruments utilisés pour le test. La notion d'optimalité des instruments que nous considérons diffère de celle typiquement connue dans la littérature [Amemiya (1977)] et qui consiste à minimiser la variance asymptotique d'un estimateur. Dans cet essai, l'optimalité est définie par rapport à la maximisation de la puissance d'un test. Nous calculons la fonction puissance du test et nous généralisons un vieux résultat de Revankar-Mallela (1972) qui ont calculé la fonction puissance du test AR dans le contexte restreint des modèles à équations simultanées linéaires. Par la suite, nous étudions la monotonocité de cette fonction puissance en fonction du nombre et de la valeur des instruments. Les instruments optimaux que nous dérivons dépendent de la valeur de l'alternative: nous parlons d' "instruments point-optimaux".

La distribution exacte du test est obtenue sous les hypothèses de normalité et d'indépendance des perturbations, mais nous montrons que le test reste valide asymptotiquement sous des hypothèses similaires à celles habituellement considérées dans la théorie asymptotique. La matrice des instruments optimaux est inconnue de façon générale puisqu'elle dépend des paramètres inconnus et/ou de la distribution des variables endogènes. Nous proposons une méthode basée sur la technique de "split-sample" [Dufour-Jasiak (1993), Angrist-Krueger (1995)] pour estimer cette matrice.

Le troisième essai s'inscrit dans la littérature qui étudie le problème des instruments faibles. Il s'agit de la sélection d'instruments. En effet, dans la recherche empirique le nombre potentiel d'instruments est généralement élevé. L'inclusion d'un

instrument faible cause des détériorations de biais dans l'estimation par IV et la perte de puissance dans les tests comme Anderson-Rubin [voir Phillips (1980), Buse (1992), et Kleibergen (2001a)]. Le chercheur est donc amené à sélectionner uniquement les instruments qui semblent pertinents. Très récemment deux études ont abordé ce problème. Hall-Peixe (2000) propose une méthode qui consiste à sélectionner les instruments sur la base de la corrélation canonique entre les instruments et les variables endogènes explicatives. La méthode consiste à sélectionner le sous-ensemble d'instruments qui minimise un critère basé sur cette corrélation canonique. Un inconvénient de cette méthode est qu'elle est assez lourde à appliquer en pratique, puisque avec un nombre l d'instruments potentiels, le nombre de choix possibles est $2^l - 1$. Donald-Newey (2000) propose un critère basé sur l'erreur quadratique moyenne approximative pour sélectionner des instruments pour l'estimation par la méthode de maximum de vraisemblance à information limitée (LIML), les doubles moindres carrés (2SLS), et la version jackknife de 2SLS [voir Angrist-Imbens-Krueger (1999)]. Les auteurs supposent donnée une matrice d'instruments ordonnés et se fixent comme objectif de déterminer le nombre k tel que les k premiers instruments optimisent le critère. Cette approche a un défaut important: si les instruments non pertinents correspondent aux premières colonnes de la matrice des instruments potentiels, le sous-ensemble d'instruments qui sera sélectionné va les inclure. Dans cet essai, nous considérons ce problème de sélection des instruments dans le cadre d'un modèle structurel à équations linéaires où la méthode 2SLS pour l'estimation et/ou le test AR sont utilisés. Nous étudions d'abord comment la matrice des instruments affecte la qualité de l'estimation et de l'inférence, et nous considérons ensuite la sélection d'instruments sur la base de ces résultats. Nous nous concentrons sur le cas d'un modèle à une seule variable explicative endogène. Ce cas est intéressant parce que c'est assez souvent le cas dans la recherche empirique, et en plus les distributions exactes des estimateurs et tests usuels sont relativement maniables. La méthode que nous proposons est basée sur la maximisation séquentielle du paramètre de concentration et est simple à appliquer dans la pratique. Nous utilisons les simulations Monte Carlo pour déterminer si son utilisation permet d'améliorer la qualité de l'estimation

et de l'inférence par rapport à la méthode qui consiste à utiliser tous les instruments potentiels. Nous faisons également une comparaison entre les différentes méthodes de sélection proposées dans la littérature.

Chapter 1

Projection-based statistical inference in linear structural models with possibly weak instruments

1. Introduction

One of the classic problems of econometrics consists in making inference on the coefficients of structural models. Such models typically involve endogenous explanatory variables (which can lead to endogeneity biases), the need to use instrumental variables, and the possibility that "structural parameters of interest" may not be identifiable. Recently, the statistical problems raised by structural modelling have received new attention in view of the observation that proposed instruments are often "weak", i.e. poorly correlated with the relevant endogenous variables, which correspond to situations where the structural parameters are close to being not identifiable (through the instruments used). The literature on so-called "weak instruments" problems is now considerable; see, for example, Nelson and Startz (1990a, 1990b), Buse (1992), Maddala and Jeong (1992), Bound, Jaeger, and Baker (1993, 1995), Angrist and Krueger (1995), Hall, Rudebusch, and Wilcox (1996), Dufour (1997), Shea (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz, and Nelson (1998), Startz, Nelson, and Zivot (1999), Perron (1999), Stock and Wright (2000), Dufour and Jasiak (2001), and Kleibergen (2001a, 2001b).

In such contexts, several papers have documented by simulation and approximate asymptotic methods the poor performance of standard asymptotically justified procedures [Nelson and Startz (1990a, 1990b), Buse (1992), Bound, Jaeger, and Baker (1993, 1995), Hall, Rudebusch, and Wilcox (1996), Staiger and Stock (1997), Zivot, Startz, and Nelson (1998), Dufour and Jasiak (2001)]. The main difficulty here is that the finite-sample distributions of the relevant statistics (in particular, test statistics) are very sensitive to unknown nuisance parameters; indeed, they can exhibit an arbitrary large sensitivity to such parameters. Further, limiting distributions are non-standard when identification conditions do not hold, while usual large-sample approximations do not converge uniformly, so that the latter may be arbitrarily inaccurate in finite samples even when identification and standard regularity conditions obtain. The fact that standard asymptotic theory can be arbitrarily inaccurate in finite samples (of any size) is shown rigorously in Dufour (1997), where it is observed

that valid confidence intervals in a standard linear structural equations model must be unbounded with positive probability and Wald-type statistics have distributions which can deviate arbitrarily from their large-sample distribution (even when identification holds). The fact that both finite-sample and large-sample distributions exhibit strong dependence on nuisance parameters has also been demonstrated, such as finite-sample distributional theory [see Choi and Phillips (1992)] and local to non-identification asymptotics [see Staiger and Stock (1997) and Wang and Zivot (1998)].

As a result, it appears especially important in such problems to build tests and confidence sets based on properly pivotal (or boundedly pivotal) functions, as well as to study inference procedures from a finite-sample perspective. The fact that tests should be based on statistics whose distributions can be bounded and that confidence sets should be obtained from pivotal statistics is, of course, a requirement of basic statistical theory [see Lehmann (1986)]. In the framework of linear simultaneous equations and in view of weak instrument problems, the importance of using pivotal functions for statistical inference has been recently reemphasized by several authors [see Dufour (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz, and Nelson (1998), Startz, Nelson, and Zivot (1999), Dufour and Jasiak (2001) and Kleibergen (2001a, 2001b)]. In particular, this suggests that confidence sets should be built by inverting likelihood ratio (LR) and Lagrange multiplier (LM) type statistics, as opposed to the more usual method which consists in inverting Wald-type statistics (such as asymptotic t-ratios).

Useful pivotal functions are however difficult to find in structural models. The oldest and still the most important one is the Anderson and Rubin (1949, henceforth AR) statistic. The latter allows one to test an hypothesis which fixes the value of the full vector of the endogenous explanatory variables coefficients in a linear structural equation; under usual parametric assumptions (error Gaussianity, instrument strict exogeneity) the distribution of the statistic is a central Fisher distribution, while under weaker (standard) assumptions it is asymptotically chi-square, irrespective of the presence of weak instruments. Extensions and variants of this method have also been suggested in Wang and Zivot (1998), Dufour and Taamouti (2000a), Dufour and

Jasiak (2001), and Kleibergen (2001a).

An important practical shortcoming of these methods is that they are designed to test hypotheses of the form $H_0 : \beta = \beta_0$, where β is the vector of all coefficients of the endogenous explanatory variables. In particular these statistics do not allow to test linear and nonlinear restrictions on the vector β. A general solution to this problem is the *projection* technique described in Dufour (1990, 1997) and Dufour and Jasiak (2001).

The basic idea behind the latter approach is quite simple. Let θ be a multidimensional parameter vector for which we can build a confidence set $C_\theta(\alpha)$ with level $1-\alpha : \mathsf{P}[\theta \in C_\theta(\alpha)] \geq 1-\alpha$. Now consider a transformation of interest $g(\theta)$ which takes its values in $\mathbb{R}^m$. For example, $g(\theta)$ could be one of the components of θ. Then it is easy to see that the image set $g[C_\theta(\alpha)] = \{g(\theta) \in \mathbb{R}^m : \theta \in C_\theta(\alpha)\}$ is a confidence set with level $1-\alpha$ for $g(\theta)$, i.e. $P\big[g(\theta) \in g[C_\theta(\alpha)]\big] \geqslant P[\theta \in C_\theta(\alpha)] \geqslant 1-\alpha$. Such methods are also exploited in Abdelkhalek and Dufour (1998) and Dufour and Kiviet (1998) for completely different models. In general, however, the calculation of $g[C_\theta(\alpha)]$ is not simple and may require using costly numerical methods [as done, for example, in Abdelkhalek and Dufour (1998), Dufour and Kiviet (1998) or Dufour and Jasiak (2001)].

In this paper, we study some general geometrical features of AR-type confidence sets and we provide a complete close-form solution to the problem of building projection-based confidence sets from such confidence sets. We first observe that AR-type confidence sets can be described as **quadrics** [see Shilov (1961, Chapter 11) and Pettofrezzo and Marcoantonio (1970)], a class of geometric figures which covers as special cases the more usual confidence intervals and confidence ellipsoids but also includes hyperboloids and paraboloids. In particular we give a simple necessary and sufficient condition under which such confidence sets are bounded (which indicates we do not have an identification problem). We use the projection technique to build confidence sets for components of the vector of unknown parameters and for linear combinations of these components. Using these results, we then derive simple explicit expressions for projection-based confidence intervals in the case of individual

structural coefficients (or linear transformations of these coefficients). Consequently, no search by nonlinear methods is anymore required. The explicit calculation of the confidence sets thus makes the projection approach very attractive, especially in situations where no other solution has been proposed to date in the literature. Note also the intervals obtained in this way can be viewed as generalizations of Scheffé-type simultaneous confidence intervals [see Scheffé (1959), Savin (1984) and Dufour (1989)].

The study of the characteristics of the confidence sets before projecting them enables one to study the form of the projection set before an explicit calculation is made. In particular, we give the general necessary and sufficient condition under which AR-type confidence sets are bounded. This condition is easy to check in practice. We also show that, when the projection-based confidence intervals are bounded, they may be interpreted as confidence intervals based on k-class estimators [for a discussion of k-class estimators, see Davidson and MacKinnon (1993, page 649)] where the "standard error" is corrected in a way that depends on the level of the test. The confidence interval for a linear combination of the parameters $w'\beta$ takes the usual form $[w'\hat{\beta} - \hat{\sigma} z_\alpha; w'\hat{\beta} + \hat{\sigma} z_\alpha]$ with $\hat{\beta}$ a k-class type estimator of β. We further discuss in this case the relationship of such sets with Scheffé-type simultaneous confidence sets, which are widely used in analysis of variance.

The methods discussed in this work are evaluated and compared on the basis of Monte Carlo simulations. In particular we analyze the conservatism of the projection-based confidence sets.

To illustrate the projection approach, we present two empirical applications. In the first one, we study the relationship between standards of living and openness in the context of an equation previously considered by Frankel and Romer (1999). In the second one, we study returns to scale and externalities in various industrial sectors of the U.S. economy, using a production function specification previously considered by Burnside (1996).

In Section 2, we present the background model and statistical inference methods on the coefficient vector of the explanatory endogenous variables. Section 3 presents

the simultaneous confidence sets. In Section 4, we discuss some general properties of quadric confidence sets and provide a simple necessary and sufficient condition under which such sets are bounded. Section 5 provides explicit projection-based confidence intervals for individual structural parameters and linear transformations of these parameters. In Section 6, we present the results of Monte Carlo simulations, and Section 7 presents the empirical applications. Finally, Section 8 concludes.

2. Framework

We consider here the standard simultaneous equations model (SEM):

$$y = Y\beta + X_1\gamma + u\,, \tag{2.1}$$

$$Y = X_2\Pi_2 + X_1\Pi_1 + V\,, \tag{2.2}$$

where y and Y are $T \times 1$ and $T \times G$ matrices of endogenous variables, X_1 and X_2 are $T \times k_1$ and $T \times k_2$ matrices of exogenous variables, β and γ are $G \times 1$ and $k_1 \times 1$ vectors of unknown coefficients, Π_2 and Π_1 are $k_2 \times G$ and $k_1 \times G$ matrices of unknown coefficients, u is a vector of structural disturbances, and V is a $T \times G$ matrix of reduced-form disturbances; we also assume the vectors $(u_t,\, V_t')$, $t = 1, \ldots, T$, are independent and identically distributed with mean zero and covariance matrix Σ, with $\det(\Sigma) \neq 0$. Further,

$$X = [X_1,\, X_2] \text{ is a full-column rank } T \times k \text{ matrix} \tag{2.3}$$

where $k = k_1 + k_2$. Finally, to get a finite-sample distributional theory for the test statistics, we shall use the following assumptions:

$$[u,\, V] \text{ and } X \text{ are independent;} \tag{2.4}$$

$$(u_t,\, V_t')' \overset{i.i.d.}{\sim} N(0,\, \Sigma)\,,\ t = 1, \ldots, T. \tag{2.5}$$

In such a model, we are generally interested in making inference on β and γ. In Dufour (1997), it is shown that if the model is unidentified (the matrix Π_2 is of reduced rank) any valid confidence set for β and γ must be unbounded with positive probability. This is due to the fact that such a model may be unidentified and holds indeed even if identification restrictions are imposed. This result explains many recent findings about the performance of standard asymptotic statistics when the instruments X_2 are weakly correlated with the endogenous explanatory variables Y. The usual approach that consists in inverting Wald-type statistics to obtain confidence sets is not valid in these situations since the resulting confidence sets are bounded with probability 1. This is related to the fact that finite-sample distributions of such statistics are not pivotal and follow distributions which depend heavily on nuisance parameters.

Choi and Phillips (1992) considered the same model where they suppose that a subset of parameters are not identified. They derive exact and asymptotic distributions of the instrumental variables estimator and the Wald statistic. The analytic expressions obtained are complex and are different from commonly known ones. Staiger and Stock (1997) also considered the same model but assumed that the elements of the matrix Π_2 tend to 0 as T increases ($\Pi_2 = C/\sqrt{T}$, where C is a fixed matrix). They derive the asymptotic distributions of different statistics including the two stage least squares (2SLS), the limited information maximum likelihood (LIML) and the Wald statistic based on these estimators. In conformity with the results in Dufour (1997), these distributions depend on nuisance parameters and are not pivotal. Wang and Zivot (1998) derived [under the same assumption as Staiger and Stock, i.e. $\Pi_2 = C/\sqrt{T}$] the asymptotic distributions of likelihood ratio (LR) and Lagrange multiplier (LM) type statistics based on maximum likelihood and GMM estimation methods. As before, these distributions depend on nuisance parameters and are not pivotal. These derivations are interesting in the way that they allow us to understand the poor performance of asymptotic approximations reported in previous works but they do not solve the statistical inference problem in these models.

A first solution to this problem [see Dufour (1997) and Staiger and Stock (1997)] consists in using the Anderson-Rubin statistic [Anderson and Rubin (1949)]. This test is based on the simple idea that if β is specified, model (2.1)-(2.2) can be reduced to a simple linear regression equation. More precisely, if we consider the hypothesis

$$H_0 : \beta = \beta_0$$

in equation (2.1), we can write:

$$y - Y\beta_0 = X_2\theta_2 + X_1\theta_1 + \varepsilon \tag{2.6}$$

where $\theta_2 = \Pi_2(\beta - \beta_0)$, $\theta_1 = \gamma + \Pi_1(\beta - \beta_0)$ and $\varepsilon = u + V(\beta - \beta_0)$. Equation (2.6) satisfies all the conditions of the linear regression model. We can test H_0 by testing $H_0' : \theta_2 = 0$. Under Assumptions (2.3) to (2.5), the usual F-test is given by:

$$\mathrm{AR}(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(X)](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k)} \sim F(k_2, T - k) \tag{2.7}$$

where for any full rank matrix B, $M(B) = I - P(B)$ and $P(B) = B(B'B)^{-1}B'$ is the projection matrix on the space spanned by the columns of B.

A confidence set for β with level $1 - \alpha$ can be obtained by inverting the statistic $\mathrm{AR}(\beta_0)$:

$$C_\beta(\alpha) = \{\beta_0 : \mathrm{AR}(\beta_0) \leq F_\alpha(k_2, T - k)\} \tag{2.8}$$

where $F_\alpha(k_2, T-k)$ is the $1-\alpha$ quantile of the Fdistribution with k_2 and $T-k$ degrees of freedom. This confidence set is exact and does not require any identification assumption. When $G = 1$ [see Dufour and Jasiak (2001)], $C_\beta(\alpha)$ is unbounded if $F(\Pi_2 = 0) < F_\alpha$, where $F(\Pi_2 = 0)$ is the F-test for $H_0 : \Pi_2 = 0$ in equation (2.2). Further, Monte Carlo simulations [Maddala (1974), Dufour and Jasiak (2001)] indicate that the AR based test behaves well in terms of power. This test also remains asymptotically valid under weaker distributional assumptions, in the sense that the asymptotic null distribution of $\mathrm{AR}(\beta_0)$ is $\chi^2_{k_2}/k_2$ [see Dufour and Jasiak (2001) and

Staiger and Stock (1997)].

In a similar context, Wang and Zivot (1998) also proposed two alternative statistics. The first one is an LR−type statistic and the second is an LM−type statistic. These two statistics follow $\chi^2_{k_2}$ distributions asymptotically when the model is exactly identified ($k_2 = G$), and are bounded by a $\chi^2_{k_2}$ distribution when the model is over-identified ($k_2 > G$). To test $H_0 : \beta = \beta_0$, these statistics are:

$$\mathrm{LR}_{\mathrm{LIML}}(\beta_0) = T[\ln(k(\beta_0)) - \ln[k(\widehat{\beta}_{\mathrm{LIML}})]\,, \tag{2.9}$$

$$\mathrm{LM}_{\mathrm{2SLS}}(\beta_0) = \frac{T(y - Y\beta_0)'P[P[M(X_1)X_2]Y](y - Y\beta_0)}{(y - Y\beta_0)'M(X_1)(y - Y\beta_0)}\,, \tag{2.10}$$

where

$$k(\beta_0) = \frac{(y - Y\beta_0)'M(X_1)(y - Y\beta_0)}{(y - Y\beta_0)'M(X)(y - Y\beta_0)}\,.$$

Asymptotic and conservative confidence sets for β can be obtained by inverting these statistics.

A common shortcoming of all these tests is that they require one to specify the entire vector β. In particular, they do not allow for general hypotheses of the form $H_0 : g(\beta) = 0$, where $g(\beta)$ may be any transformation of β, such as $g(\beta) = \beta^i - \beta^i_0$, where β^i is any scalar component of β.

In this paper, we deal with this problem by studying the characteristics of the confidence sets obtained by inverting such statistics, and then using them to derive confidence sets for the components of β or linear combinations of these components. We first show that the confidence sets based on the statistics AR, LR and LM can be written in a simple form parametrized by a matrix A, a vector b and a scalar c. These sets (replacing the inequality by an equality) are known as **quadrics** [Shilov (1961, Chapter 11), Pettofrezzo and Marcoantonio (1970, Chapters 9-10)]. We will then use mathematical results on these sets to study the different possible forms as functions of A, b and c, and we will derive analytic expressions for projection-based confidence intervals in the case of linear transformations of model parameters.

3. Simultaneous quadric confidence sets in SEM

Consider first the statistic AR. A simple algebraic calculation shows that the inequality

$$\mathrm{AR}(\beta_0) \leq F_\alpha(k_2, T-k)$$

may be written in the following simple form:[1]

$$\beta_0' A \beta_0 + b'\beta_0 + c \leq 0 \tag{3.1}$$

where $A = Y'HY$, $b = -2Y'Hy$, $c = y'Hy$ and

$$H \equiv H_{\mathrm{AR}} = M(X_1) - \left[1 + \frac{k_2 F_\alpha(k_2, T-k)}{T-k}\right] M(X)\,. \tag{3.2}$$

We can thus write:

$$C_\beta(\alpha) = \{\beta_0 : \beta_0' A \beta_0 + b'\beta_0 + c \leq 0\}\,. \tag{3.3}$$

If β is scalar, this set is the solution of a quadratic inequation:

$$C_\beta(\alpha) = \{\beta_0 : a\beta_0^2 + b\beta_0 + c \leq 0\}. \tag{3.4}$$

Depending on the values of a, b and c, this set may take several forms (a closed interval, a semi-open interval, a union of two semi-open intervals, the set $\mathbb{R}$ of all possible values, or the empty set); see Dufour and Jasiak (2001), and Zivot, Startz, and Nelson (1998).

If we use the statistics $\mathrm{LR}_{\mathrm{LIML}}(\beta_0)$ or $\mathrm{LM}_{\mathrm{2SLS}}(\beta_0)$ instead of AR, we get analogous confidence sets which only differ through the H matrix. For $\mathrm{LR}_{\mathrm{LIML}}(\beta_0)$, this matrix takes the form

$$H_{\mathrm{LR}} = M(X_1) - \exp[\chi^2_{k_2}(\alpha)/T]\, k(\widehat{\beta}_{\mathrm{LIML}})\, M(X) \tag{3.5}$$

[1] This form is also given in Wang and Zivot (1998).

while, for $\mathrm{LM}_{\mathrm{2SLS}}(\beta_0)$, it is

$$H_{\mathrm{LM}} = P\big[P[M(X_1)X_2]Y\big] - [\chi^2_{k_2}(\alpha)/T]\, M(X_1)\,. \tag{3.6}$$

For the AR and LR statistics, the matrix A may be written:

$$A = Y'M(X_1)Y - [1 + f(\alpha)]\, Y'M(X)Y$$

where $f(\alpha) = k_2 F_\alpha(k_2, T-k)/(T-k)$ for the AR statistic and $f(\alpha) = \exp[\chi^2_{k_2}(\alpha)/T]k(\widehat{\beta}_{\mathrm{LIML}}) - 1$ for the LR statistic. Clearly A is symmetric and a typical diagonal element of this matrix is

$$A_{ii} = Y_i'M(X_1)Y_i - (1 + f(\alpha))Y_i'M(X)Y_i\,, \tag{3.7}$$

which is a **corrected** difference between the sum of squared residuals from the regression of Y_i on X_1 and the sum of squared residuals from the regression of Y_i on $X = [X_1, X_2]$. This difference may be viewed as a measure of the importance of X_2 in explaining Y_i, *i.e.* the relevance of X_2 as an instrument for Y_i. A necessary condition for matrix A to be positive definite is that the instruments X_2 should provide sufficient additional explanatory power for Y (with respect to X_1). Similarly, $c = y'Hy$ is a corrected difference between the sum of squared residuals from the regression of y on X_1 and the sum of squared residuals from the regression of y on $X = [X_1, X_2]$. For the vector b, a typical element is given by

$$b_i = -2\{[M(X_1)Y_i]'[M(X_1)y] - (1 + f(\alpha))[M(X)Y_i]'[M(X)y]\}. \tag{3.8}$$

The first term [multiplied by $-1/(2T)$] is the sample covariance between the residuals of the regression of Y_i on X_1 and the residuals of the regression of y on X_1, while the second term gives the same covariance with X_1 replaced by $X = [X_1, X_2]$.

4. Geometry of quadric confidence sets

The locus of points that satisfy an equation of the form

$$\beta' A\beta + b'\beta + c = 0\,, \tag{4.1}$$

where A is a symmetric $G \times G$ matrix, b is a $G \times 1$ vector and c is a scalar, is known in the mathematical literature as a *quadric* surface [Shilov (1961, Chapter 11), Pettofrezzo and Marcoantonio (1970)]. Consequently, we shall call a confidence set of the form

$$C_\beta = \{\beta : \beta' A\beta + b'\beta + c \leq 0\} \tag{4.2}$$

a *quadric confidence set.* A quadric is characterized by the sum a quadratic form $(\beta' A\beta)$ and an affine transformation $(b'\beta + c)$. Depending on the values of A, b and c, it may take several forms. In this section, we examine some general properties of quadric confidence sets, especially the conditions under which such sets are bounded or unbounded. In particular, we will see that the eigenvalues of the A matrix play a central role in these properties and that larger eigenvalues are associated with more "concentrated" (or "smaller") quadric confidence sets. For these reasons, we call A the *concentration matrix* of the quadric.

In the sequel of this section, it will be convenient to distinguish between two basic cases: the one where A is nonsingular, and the one where it is singular. We adopt the convention that an empty set is bounded.

4.1. Nonsingular concentration matrix

If A is nonsingular, we can write:

$$\begin{aligned} \beta' A\beta + b'\beta + c &= \left(\beta + \frac{1}{2}A^{-1}b\right)' A\left(\beta + \frac{1}{2}A^{-1}b\right) - \left(\frac{1}{4}b'A^{-1}b - c\right) \\ &= (\beta - \widetilde{\beta})' A(\beta - \widetilde{\beta}) - d \end{aligned} \tag{4.3}$$

where $\widetilde{\beta} = -\frac{1}{2}A^{-1}b$ and $d = \frac{1}{4}b'A^{-1}b - c$. Since A is a real symmetric matrix, we can write $A = P'DP$ where P is an orthogonal matrix and D is a diagonal matrix whose elements are the eigenvalues of A. Inequality (3.1) may then be reexpressed as

$$\lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d \tag{4.4}$$

where the λ_i's are the eigenvalues of A and $z = P(\beta - \widetilde{\beta})$. The transformation $z = P(\beta - \widetilde{\beta})$ represents a translation followed by a rotation of β, so it is clear we have:

$$C_\beta \text{ is bounded } \Leftrightarrow C_z \text{ is bounded} \tag{4.5}$$

where

$$\begin{aligned} C_\beta &\equiv \{\beta : \beta' A\beta + b'\beta + c \leq 0\} = \{\beta : (\beta - \widetilde{\beta})' A(\beta - \widetilde{\beta}) \leq d\} && (4.6)\\ &= \{\beta : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d \text{ and } z = P(\beta - \widetilde{\beta})\}, && (4.7)\\ C_z &\equiv \{z : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d\}. && (4.8) \end{aligned}$$

Again it will be convenient to distinguish between three cases according to the signs of the eigenvalues of A, namely:

1. all the eigenvalues of A are positive ($\lambda_i > 0$, $i = 1, \ldots, G$), i.e. A is positive definite;

2. all the eigenvalues of A are negative ($\lambda_i < 0$, $i = 1, \ldots, G$), i.e. A is negative definite;

3. A has both positive and negative values, i.e. A is neither positive nor negative definite.

4.1.1. Positive definite concentration matrix

If $\lambda_i > 0$, $i = 1, \ldots, G$, the inequality (4.4) can be reexpressed as

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_G}{\gamma_G}\right)^2 \leq d, \tag{4.9}$$

where $\gamma_i = \sqrt{1/\lambda_i}$, $i = 1, \ldots, G$. If $d = 0$, we have $C_z = \{0\}$ and C_β reduces to $\{\widetilde{\beta}\}$. If $d < 0$, C_z and C_β are empty. If $d > 0$, C_z is the area inside an ellipsoid, hence it is a compact set, and similarly for C_β. Consequently, C_z and C_β are bounded.

4.1.2. Negative definite concentration matrix

If $\lambda_i < 0$, $i = 1, \ldots, G$, the set C_z is the set of all values of z that satisfy

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_G}{\gamma_G}\right)^2 \geq -d \tag{4.10}$$

where $\gamma_i = \sqrt{-1/\lambda_i}$, or equivalently, the set *not inside* the open ellipsoid defined by

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_G}{\gamma_G}\right)^2 < -d\,. \tag{4.11}$$

C_z and C_β are thus unbounded sets. In particular, if $d \geq 0$, we have $C_\beta = C_z = \mathbb{R}^G$.

4.1.3. Concentration matrix not positive or negative definite

If A has both positive and negative eigenvalues, we can assume, without loss of generality, that $\lambda_i > 0$ for $i = 1, \ldots, p$, and $\lambda_i < 0$, for $i = p+1, \ldots, G$, where $1 \leq p < G$. Inequality (4.4) may then be rewritten:

$$\left(\frac{z_1}{\gamma_1}\right)^2 + \cdots + \left(\frac{z_p}{\gamma_p}\right)^2 - \left(\frac{z_{p+1}}{\gamma_{p+1}}\right)^2 - \cdots - \left(\frac{z_G}{\gamma_G}\right)^2 - d \leq 0 \tag{4.12}$$

where p is the number of positive eigenvalues of A, $\gamma_i = \sqrt{1/\lambda_i}$ for $i = 1, \ldots, p$, and $\gamma_i = \sqrt{-1/\lambda_i}$ for $i = p+1, \ldots, G$.

In this case, the quadric (4.1) takes the form of an hyperboloid or a paraboloid

[see Shilov (1961, Chapter 11)]. In all cases, C_z and C_β are unbounded. This is easy to see as follows: for arbitrary given values of $z_1, , \ldots , z_p$ and d, it is clear that inequality (4.12) will hold if any of the values z_i, $p+1 \leq i \leq G$, is small enough (as $|z_i| \to \infty$). Consequently, each component of z is unbounded in C_z and similarly for each component of β in C_β.

4.2. Singular concentration matrix

We now consider the case where A is singular. Since A is real symmetric, there exists an orthogonal matrix P such that $A = P'DP$, where D is a diagonal matrix with the eigenvalues of A as diagonal elements. If we assume that A has r eigenvalues different from 0 ($r < G$), we have (see the details in the Appendix):

$$\beta' A\beta + b'\beta + c = \sum_{i=1}^{r} \lambda_i z_i^2 + \sum_{i=r+1}^{G} \delta_i z_i - d \tag{4.13}$$

where λ_i are the eigenvalues of A that are different from 0, $\delta = Pb$, $d = -c + \sum_{i=1}^{r} \delta_i^2/(4\lambda_i)$, $z = P\beta + \mu$ and

$$\mu_i = \begin{cases} \delta_i/(2\lambda_i), & \text{if } \lambda_i \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{4.14}$$

In the new space given by the transformation $z = P\beta + \mu$, C_z may take many forms following the number of non-zero eigenvalues and their signs. However, this set will always be unbounded. From (4.13) it is clear that we can make any z_i, $i = r+1, \ldots , G$, arbitrarily large with an opposite sign of its coefficient δ_i, the inequality (4.4) will hold.

4.3. Necessary and sufficient condition for bounded quadric confidence set

We can now deduce the conditions under which C_β is bounded. According to results in Dufour (1997), a valid confidence set $C_\beta(\alpha)$ for β (with level $1-\alpha$) in model (2.1)-(2.5) must be unbounded with positive probability for any parameter configuration, a probability that should be large (close to $1-\alpha$) when the matrix Π_2 does not have full rank (or is close to have full column rank). Given the complicated expressions of the random matrix A, the random vector b and the random scalar c, it seems difficult to evaluate this probability. In the following proposition, we give an easy-to-verify necessary and sufficient condition for a confidence set of the form C_β to be bounded.

Proposition 4.1 NECESSARY AND SUFFICIENT CONDITION FOR BOUNDED CONFIDENCE SET. *The set C_β in (4.2) is bounded if and only if the matrix A is positive definite.*

Proofs are provided in the Appendix. It is of interest to note here that the case where A is singular is unlikely to be met with AR-type confidence sets such as those described in Section 3, because in this case we have $A = Y'HY$, where Y and H, are a $T \times G$ and a $T \times T$ matrices respectively. If Y follows an absolutely continuous distribution (as assumed in Section 2), A will be nonsingular with probability one as soon as the rank of H is greater than or equal to G.

Finally, we note that the above results also apply to the problem of building joint confidence sets for β and γ. This can be done by using an appropriate extension of the AR procedure [see Dufour and Jasiak (2001)]. From (2.1) - (2.2), we can write:

$$y - Y\beta_0 - X_1\gamma_0 = X_2\theta_2 + X_1\theta_1 + \xi \tag{4.15}$$

where $\theta_2 = \Pi_2(\beta - \beta_0)$, $\theta_1 = \Pi_1(\beta - \beta_0) + \gamma - \gamma_0$ and $\xi = V(\beta - \beta_0) + u$. We can test

$$H_0 : (\beta, \gamma) = (\beta_0, \gamma_0) \tag{4.16}$$

by testing $H_0' : \theta_1 = 0$ and $\theta_2 = 0$ and obtain a joint confidence set for β and γ by inverting the corresponding F-statistic. We obtain after a similar simple algebraic calculation the same form as before:

$$C_{(\beta_0',\gamma_0')'}(\alpha) = \{(\beta_0', \gamma_0')' : (\beta_0', \gamma_0')A_1(\beta_0', \gamma_0')' + b_1'(\beta_0', \gamma_0')' + c_1 \leq 0\} \tag{4.17}$$

where $A_1 = [Y, X_1]'H_1[Y, X_1]$, $b_1 = -2[Y, X_1]'H_1 y$, $c_1 = y'H_1 y$ and

$$H_1 = I - \left[1 + \frac{k}{T-k}F_{k,T-k}(\alpha)\right] M(\bar{X}). \tag{4.18}$$

5. Confidence sets for transformations of β

5.1. The projection approach

The projection technique is a general approach that may be applied in different contexts. Given a confidence set $C_\theta(\alpha)$ with level $1-\alpha$ for the vector of parameters θ, this method enables one to deduce confidence sets for general transformations g in $\mathbb{R}^m$ of this vector. Since:

$$x \in E \Rightarrow g(x) \in g(E)$$

for any set E, we have

$$\mathsf{P}[\theta \in C_\theta(\alpha)] \geq 1-\alpha \Rightarrow \mathsf{P}\left[g(\theta) \in g\left[C_\theta(\alpha)\right]\right] \geq 1-\alpha, \tag{5.1}$$

where $g\left[C_\theta(\alpha)\right] = \{x \in \mathbb{R}^m : \exists\, \theta \in C_\theta(\alpha),\ g(\theta) = x\}$. Hence $g(\theta) \in g\left[C_\theta(\alpha)\right]$ is a conservative confidence set for $g(\theta)$ with level $1-\alpha$.

If $g(\theta)$ is scalar, the projection-based confidence set is not necessarily an interval. In Dufour (1997) [see also Abdelkhalek and Dufour (1998), Wang and Zivot (1998), Stock and Wright (2000), and Dufour and Jasiak (2001)], it is suggested that we can obtain an interval by taking:

$$g_L(\alpha) = \inf\{g(\theta_0), \theta_0 \in C_\theta(\alpha)\}, \quad g_U(\alpha) = \sup\{g(\theta_0), \theta_0 \in C_\theta(\alpha)\},$$

so that

$$\mathrm{P}[g_L(\alpha) \leq g(\theta) \leq g_U(\alpha)] \geqslant 1 - \alpha\,. \tag{5.2}$$

Hence $[g_L(\alpha),\, g_U(\alpha)]\backslash\{-\infty,\, +\infty\}$ is a confidence interval with level $1-\alpha$ for $g(\theta)$, where it is assumed that $-\infty$ and $+\infty$ are not admissible. This interval is not bounded if $g_L(\alpha)$ or $g_U(\alpha)$ is infinite.

It is worth noting that we can obtain simultaneous confidence sets for any number of transformations of β: $g_1(\beta),\, g_2(\beta),\, \ldots,\, g_n(\beta)$. The set $C_{g_1(\beta)}(\alpha) \times C_{g_2(\beta)}(\alpha) \times \cdots \times C_{g_n(\beta)}(\alpha)$ where $C_{g_i(\beta)}(\alpha)$ is the projection-based confidence set for $g_i(\beta)$, $i = 1, \ldots, n$, is a simultaneous confidence set for $\big(g_1(\beta),\, g_2(\beta),\, \ldots,\, g_n(\beta)\big)$ with level greater than or equal to $1-\alpha$.

In this section, we build confidence sets for $g(\beta)$ by "projecting" the set $C_\beta(\alpha)$.[2] We study two particular transformations: $g(\beta) = w'\beta$ (a linear combination of the components of β) and $g(\beta) = \beta_i$ (the projection on the axis β_i).[3] We also show that calculating $g\,[C_\theta(\alpha)]$ by minimizing and maximizing $g(\beta)$ over $C_\beta(\alpha)$ results in a total loss of information when the two optimization problems have unbounded solutions in a manner that $g_L(\alpha) = -\infty$ and $g_U(\alpha) = +\infty$ when in fact the appropriate projection gives unbounded sets but strictly included in $\mathbb{R}$, hence the importance of studying the set $C_\beta(\alpha)$ before choosing the way to project it.

One should notice here that if our aim is only to test $H_0 : g(\beta) = 0$, we can easily deduce from $C_\beta(\alpha)$ a conservative test, it consists of rejecting H_0 when all vectors β_0 that satisfy H_0 are rejected by the AR test, or equivalently when the minimum of $\mathrm{AR}(\beta_0)$ subject to the constraint (s.c.) $g(\beta) = 0$ is larger than $F_\alpha(k_2, T-k)$:

$$\min_{g(\beta)=0} \mathrm{AR}(\beta) \geq F_\alpha(k_2, T-k). \tag{5.3}$$

[2]Given that $C_\beta(\alpha)$ and $C_{(\beta',\delta')'}(\alpha)$ have the same form, the projection of the later set may be calculated in the same way by replacing β by $(\beta', \delta')'$.

[3]We suppose that $G \geq 2$. The case $G = 1$ is simple and is considered in Dufour and Jasiak (2001).

5.2. Projection-based confidence sets for scalar linear transformations

We consider now a general confidence set of the form

$$C_\beta(\alpha) = \{\beta_0 : \beta_0' A \beta_0 + b'\beta_0 + c \leq 0\} \tag{5.4}$$

where c and d are real scalars, A is a symmetric $G \times G$ matrix, and b is a $G \times 1$ vector. By definition, the associated projection-based confidence interval for the scalar function $g(\beta) = w'\beta$ is:

$$C_{w'\beta}(\alpha) \equiv g[C_\beta(\alpha)] = \{w'\beta_0 : \beta_0' A \beta_0 + b'\beta_0 + c \leq 0\}. \tag{5.5}$$

To study the characteristics of $C_{w'\beta}(\alpha)$ and its close-form, we shall distinguish again between the case where A is nonsingular and the case where it is singular.

5.2.1. Nonsingular concentration matrix

In this case all the eigenvalues of A are different from 0, and we shall distinguish 3 cases depending on the number of negative eigenvalues:

1. all the eigenvalues of A are positive (i.e., A is positive-definite);
2. A has one negative eigenvalue;
3. A has at least two negative eigenvalues.

We shall consider again the transformation $z = P(\beta - \widetilde{\beta})$, then $C_{w'\beta}(\alpha)$ may be written:

$$C_{w'\beta}(\alpha) = \{w'\beta_0 : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d \text{ and } z = P(\beta_0 - \widetilde{\beta})\}.$$

Let $a = Pw$, then

$$w'\beta = w'P'P\beta = w'P'P(\beta - \widetilde{\beta}) + w'P'P\widetilde{\beta}$$

$$= a'z + w'\widetilde{\beta} \tag{5.6}$$

and define

$$C_{a'z} = \{a'z : \lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 \leq d\}. \tag{5.7}$$

From equation (5.6), it is clear that

$$C_{a'z} = \mathbb{R} \quad \Leftrightarrow \quad C_{w'\beta}(\alpha) = \mathbb{R}\ .$$

When A is positive definite, $C_\beta(\alpha)$ is a bounded set and, correspondingly, its image $g[C_\beta(\alpha)]$ by the continuous function $g(\beta) = w'\beta$ is also bounded [see Abdelkhalek and Dufour (1998, Proposition 2)]. The following proposition provides an explicit form for the projection-based confidence set $C_{w'\beta}(\alpha)$.

Proposition 5.1 PROJECTION-BASED CONFIDENCE SETS FOR LINEAR TRANSFORMS WHEN A IS POSITIVE DEFINITE. *Suppose the matrix A in* (5.4) *is positive definite and let $w \in \mathbb{R}^G \backslash \{0\}$. Then, if $d \geq 0$, where $d = \frac{1}{4}b'A^{-1}b - c \geq 0$, the projection-based confidence set for $w'\beta$ is given by the interval*

$$C_{w'\beta}(\alpha) = \left[w'\widetilde{\beta} - \sqrt{d(w'A^{-1}w)}\ ,\ w'\widetilde{\beta} + \sqrt{d(w'A^{-1}w)}\ \right] \tag{5.8}$$

where $\widetilde{\beta} = -\frac{1}{2}A^{-1}b$. Further, if $d < 0$, then $C_{w'\beta}(\alpha)$ is empty.

In the special case where $w = e_i = (\delta_{1i}, \delta_{2i}, \ldots, \delta_{Gi})'$, $\delta_{ji} = 1$ if $j = i$ and $\delta_{ji} = 0$ if $j \neq i$, the set $C_{w'\beta}(\alpha)$ is a confidence interval for the component β_i. This set is given by the following corollary, which is a direct consequence of Proposition **5.1**.

Corollary 5.2 PROJECTION-BASED SET INTERVAL FOR AN INDIVIDUAL COEFFICIENT. *Suppose the matrix A in* (5.4) *is positive definite, then, if $d \equiv \frac{1}{4}b'A^{-1}b - c \geq 0$, the projection-based confidence set for any component β_i of β is given by the interval*

$$C_{\beta_i}(\alpha) = \left[\ \widetilde{\beta}_i - \sqrt{dA_{ii}^{-1}}\ ,\ \widetilde{\beta}_i + \sqrt{dA_{ii}^{-1}}\ \right]$$

where $\widetilde{\beta}_i = -A_{i.}^{-1}b/2$ *is the* i*-th element of* $\widetilde{\beta} = -\frac{1}{2}'A^{-1}b$, $A_{i.}^{-1}$ *is the* i*-th row of* A^{-1}, A_{ii}^{-1} *is the* i*-th element of the diagonal of* A^{-1}, *and* $A_{ii}^{-1} > 0$. *Further, if* $d < 0$, *then* $C_{\beta_i}(\alpha)$ *is empty.*

It is interesting to notice the relation with Scheffé-type confidence sets. The confidence set for β is based on the F-test of $H_0 : \theta_2 = \Pi_2(\beta - \beta_0) = 0$ in the regression equation:

$$y - Y\beta_0 = X_2\theta_2 + X_1\theta_1 + \varepsilon.$$

So H_0 is not rejected means [Scheffé (1959)] that the hypotheses $H_0(a) : a'\theta_2 = 0$ are not rejected for any $k_2 \times 1$ vector a, with a (conservative) decision that rejects $H_0(a)$ when $|t(a)| > S$, $t(a)$ is the student statistic for $H_0(a)$ and $S = \sqrt{k_2 F_{k_2,T-k}(\alpha)}$. Since $a'\theta_2 = a'\Pi_2(\beta - \beta_0)$, this is equivalent to say that the hypothesis $H_0'(w) : w'\beta = w'\beta_0$ is not rejected for any $G \times 1$ vector w. The projection based confidence set for $w'\beta$ is merely the Scheffé confidence set based on the conservative test of $H_0(a) : a'\theta_2 = 0$ where a is such that $w = \Pi_2 a$. So if $\beta_0 \in C_\alpha(\beta)$ then $w'\beta_0 \in C_\alpha(w'\beta)\ \forall\ w$. On the other hand, if H_0 is rejected, there is at least one hypothesis $H_0(a)$ that is rejected. In particular we will have $\max_a |t(a)| > S$. The vector a that corresponds to the maximum is given by [see Savin (1984)]:[4]

$$a^* = \frac{V^{-1}(\hat{\theta}_2 - \theta_2)}{\sqrt{(\hat{\theta}_2 - \theta_2)'V^{-1}(\hat{\theta}_2 - \theta_2)}}$$

where

$$V = [I_{k_2}, 0](\bar{X}'\bar{X})^{-1}[I_{k_2}, 0]' = [X_2'M(X_1)X_2]^{-1},$$

and $\hat{\theta}_2$ is the least squares estimator of θ_2. Hence one of the combinations responsible for the rejection of H_0, is $a^{*\prime}\theta_2 = a^{*\prime}\Pi_2(\beta - \beta_0)$.

Let us now consider the case where A has exactly one negative eigenvalue. Our basic result on this case is given by the following proposition.

Proposition 5.3 PROJECTION-BASED CONFIDENCE SETS FOR LINEAR TRANS-

[4] a^* is the solution verifying the normalisation $a'V^{-1}a = 1$.

FORMS WHEN A HAS ONE NEGATIVE EIGENVALUE. *Let $w \in \mathbb{R}^G \backslash \{0\}$ and suppose the matrix A in* (5.4) *is nonsingular with exactly one negative eigenvalue. Then, if $d < 0$ and $w'A^{-1}w < 0$, the projection-based confidence set for $w'\beta$ is given by :*

$$C_{w'\beta}(\alpha) = \left]-\infty\,,\; w'\widetilde{\beta} - \sqrt{d(w'A^{-1}w)}\right] \cup \left[w'\widetilde{\beta} + \sqrt{d(w'A^{-1}w)}\,,\; +\infty\right[\,. \qquad (5.9)$$

Further, if $d \geq 0$, we have $C_{w'\beta}(\alpha) = \mathbb{R}$.

It is interesting to note here that $C_{w'\beta}(\alpha)$ can remain informative, even if it is unbounded. In particular, if we want to test $H_0 : w'\beta = r$ and consider as a decision rule one that rejects H_0 if $r \notin C_{w'\beta}(\alpha)$, H_0 will be rejected for all values of r in the interval

$$\left[w'\widetilde{\beta} - \sqrt{d(w'A^{-1}w)}\,,\; w'\widetilde{\beta} + \sqrt{d(w'A^{-1}w)}\right]\,. \qquad (5.10)$$

On the other hand, notice that if we try to obtain this confidence set by computing $g_L(\alpha) = \inf\{w'\beta_0 : \beta_0 \in C_\beta(\alpha)\}$ and $g_U(\alpha) = \sup\{w'\beta_0 : \beta_0 \in C_\beta(\alpha)\}$, we will obtain $\mathbb{R}$, an uniformative set, while in fact the true projection-based confidence set is strictly included in $\mathbb{R}$.

Finally, we consider the case where A has at least two negative eigenvalues. We cover here the case where the matrix A is negative definite or not positive nor negative definite but with at least 2 negative eigenvalues. In this case the projection-based confidence set for any linear combination of the components of β is equal to the real line, thus uniformative. This is stated in the following proposition.

Proposition 5.4 PROJECTION-BASED CONFIDENCE SETS FOR LINEAR TRANSFORMS WHEN A HAS MORE THAN ONE NEGATIVE EIGENVALUE. *Let $w \in \mathbb{R}^G \backslash \{0\}$. If the matrix A in* (5.4) *is nonsingular and admits at least two negative eigenvalues, then $C_{w'\beta}(\alpha) = \mathbb{R}$.*

5.2.2. Singular concentration matrix

Given the great number of cases to distinguish (see Section 4.2), we use the suggestion in Dufour (1997) by solving the two optimization problems and consider only

transformations of the type $g(\beta) = \beta_i$, $i = 1, ..., G$. Even if we will not use directly the reduced form of $C_\beta(\alpha)$ in this case, it is important to clarify the relation between this set and its projections on the axes β_i and the axes z_i. The following lemma (which is also valid in the case of nonsingular matrix A) gives this relation.

Lemma 5.5 EQUIVALENCE BETWEEN CONFIDENCE SET AND PROJECTION-BASED CONFIDENCE INTERVAL BOUNDEDNESS. *The following three statements are equivalent:*

(*i*) $C_\beta(\alpha) = \{\beta_0 : \beta_0' A\beta_0 + b'\beta_0 + c \leq 0\}$ *is bounded;*

(*ii*) *the projection of* $C_\beta(\alpha)$ *on any axis* β_i *is bounded;*

(*iii*) *the projection of* $C_\beta(\alpha)$ *on any axis* z_i *where* $z = (z_1, \ldots, z_G)' = P\beta + \mu$, *is bounded,* P *is any orthogonal matrix of order* G *and* μ *is any* $G \times 1$ *vector.*

The statements (ii) and (iii) may be grouped by saying that the projection-based confidence set for any linear combination of the components of β is bounded. To calculate a confidence set for a component of β, say β_1, we consider the two problems:

$$\text{(A)} \quad \min_\beta \{e_1'\beta\} \quad \text{s.c. } \beta' A\beta + b'\beta + c \leq 0\,,$$
$$\text{(B)} \quad \max_\beta \{e_1'\beta\} \quad \text{s.c. } \beta' A\beta + b'\beta + c \leq 0\,,$$

where $e_1' = (1, 0, .., 0)$. Set

$$A = \begin{pmatrix} a_{11} & A_{21}' \\ A_{21} & A_{22} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_{(1)} \end{pmatrix}, \tag{5.11}$$

where a_{11} is a scalar, A_{21} is a $(G-1) \times 1$ vector, A_{22} a $(G-1) \times (G-1)$ matrix, b_1 is a scalar and $b_{(1)}$ is a $(G-1) \times 1$ vector. We shall distinguish between the following cases:

$$(a) \quad a_{11} > 0\,, \tag{5.12}$$
$$(b) \quad a_{11} = 0, A_{21} = \mathbf{0} \text{ and } b_1 > 0\,, \tag{5.13}$$

$$(c) \quad a_{11} = 0, A_{21} = \mathbf{0} \text{ and } b_1 < 0\,. \tag{5.14}$$

We have the following result.

Lemma 5.6 REDUCTION OF INEQUALITY TO EQUALITY CONSTRAINTS. *If* (a) *or* (c) *is true, then if a finite solution exists for the minimization problem, the constraint must be binding.*
If (a) *or* (b) *is true, then if a finite solution exists for the maximization problem, the constraint must be binding.*
If none of (a), (b) *and* (c) *is true then the solutions of the two optimization problems are infinite.*

This proposition does not mean that finite solutions exist under assumptions (a), (b) or (c), but means that if finite solutions exist we can calculate them by using the Lagrange method for problems with equality constraints. By seeking for the solutions of these problems we will determine the additional conditions that ensure the existence of finite solutions and hence resulting in closed or semi-closed intervals. These conditions and the resulting confidence sets are given in the following proposition.

Proposition 5.7 CONFIDENCE INTERVAL WITH SINGULAR CONCENTRATION MATRIX. *Suppose the matrix* A *is singular and, whenever* A_{22} *is nonsingular, set* $\Delta_{11} \equiv \frac{1}{4} b'_{(1)} A_{22}^{-1} b_{(1)}$ *and* $\Delta_{21} \equiv A'_{21} A_{22}^{-1} b_{(1)}$. *Then a confidence interval of level greater than or equal to* $1-\alpha$ *for* β_1 *can be obtained as follows:*

(a) *if* $a_{11} > 0$,

$$\begin{aligned}
C_{\beta_1}(\alpha) &= \left]-\infty\,,\ (c-\Delta_{11})\,/\,(-b_1+\Delta_{21})\right]\,, && \text{if } A_{22} \text{ is p.d. and } b_1\text{-}\Delta_{21} > 0\,,\\
&= \left[(c-\Delta_{11})\,/\,(-b_1+\Delta_{21})\ ,\ +\infty\right[\,, && \text{if } A_{22} \text{ is p.d. and } b_1\text{-}\Delta_{21} < 0\,,\\
&= \left[\frac{-b_1-\sqrt{b_1^2-4a_{11}c}}{2a_{11}}\,,\ \frac{-b_1+\sqrt{b_1^2-4a_{11}c}}{2a_{11}}\right]\,, && \text{if } A_{22} \text{ is p.s.d. and } A_{21} = b_{(1)} = 0\,,\\
&= \mathbb{R}\,, && \text{otherwise;}
\end{aligned} \tag{5.15}$$

(*b*) *if* $a_{11} = 0$, $A_{21} = \mathbf{0}$ *and* $b_1 > 0$,

$$\begin{aligned} C_{\beta_1}(\alpha) &=]-\infty\,,\ -(c-\Delta_{11})/b_1]\,, && \text{if } A_{22} \text{ is p.d.,} \\ &=]-\infty\,,\ -c/b_1]\,, && \text{if } A_{22} \text{ is p.s.d. and } b_{(1)} = 0\,, \\ &= \mathbb{R}\,, && \text{otherwise;} \end{aligned} \tag{5.16}$$

(*c*) $a_{11} = 0, A_{21} = \mathbf{0}$ *and* $b_1 < 0$,

$$\begin{aligned} C_{\beta_1}(\alpha) &= [-(c-\Delta_{11})/b_1\,,\ +\infty[\,, && \text{if } A_{22} \text{ is p.d.,} \\ &= [-c/b_1\,,\ +\infty[\,, && \text{if } A_{22} \text{ is p.s.d. and } b_{(1)} = 0\,, \\ &= \mathbb{R}\,, && \text{otherwise.} \end{aligned} \tag{5.17}$$

Note that in the case where A_{22} is p.s.d. and $A_{21} = b_{(1)} = 0$, we have:

$$C_\beta(\alpha) = \{\beta : a_{11}\beta_1^2 + b_1\beta_1 + c + \beta'_{(1)}A_{22}\beta_{(1)} \leq 0\}.$$

Since A_{22} is positive semi-definite, $a_{11}\beta_1^2 + b_1\beta_1 + c \leq 0$ (otherwise $C_\beta(\alpha)$ is empty). Taking $\beta_{(1)} = k\bar{\beta}_{(1)}$ where $\bar{\beta}_{(1)}$ is such that $\bar{\beta}'_{(1)}A_{22}\bar{\beta}_{(1)} = 0$ and k arbitrary large, we can see that $C_\beta(\alpha)$ is unbounded even if its projection on the β_1 axis is a closed interval.[5]

It is important to recall that when the confidence interval obtained by solving the two previous optimization problems is closed or semi-closed it also corresponds to the true projection of $C_\beta(\alpha)$. The only case where the projection may differ from the solution of the two problems is when this solution is $\mathbb{R}$.

The results presented in this section are important for two main reasons. First, they allow one to obtain confidence sets in situations where no other solutions has been

[5] Since $C_{\beta_1}(\alpha)$ is closed, it also corresponds to the projection of $C_\beta(\alpha)$ on β_1 axis.

proposed to date in the literature, and second the explicit expressions found avoid one the use of costly numerical methods as used in the papers cited previously. This is much more important given the nature of the problems to be solved numerically. We tried many of the standard software as GAUSS and GAMS, and they seem to have difficulties to find the solutions, unless the starting point is chosen near the solution (which is naturally unknown). However, Fortran-based IMSL routines appear to perform quite well.

5.3. A Wald-type interpretation of the projection-based confidence sets

When the eigenvalues of the matrix A are positive and the projection-based confidence set for $w'\beta$ is bounded, it is interesting to note that the form of this confidence set (see proposition **5.1**) is similar to the standard form: $(\hat{\theta} \pm \hat{\sigma} z(\alpha))$.

Since $\theta = w'\beta$, the corresponding estimator of β is

$$\tilde{\beta} = -\frac{1}{2} A^{-1} b.$$

The estimated variance of the estimator should be a constant (say $\hat{\sigma}^2$) times the matrix A^{-1}, $\hat{\sigma}^2 A^{-1}$, and since the confidence interval is of level greater than or equal $1-\alpha$, $\sqrt{d}/\hat{\sigma}$ should correspond to a quantile of an order greater than or equal $1-\alpha$ of the statistic $\left|(w'\tilde{\beta} - w'\beta)/[\hat{\sigma}^2(w'A^{-1}w)]^{1/2}\right|$.[6] Replacing A and b by their expressions, the estimator $\tilde{\beta}$ may be written:

$$\tilde{\beta} = (Y'HY)^{-1} Y'Hy.$$

$\tilde{\beta}$ may be interpreted as an instrumental variables estimator. Indeed, multiplying (2.1) by $(HY)'$, we get

$$Y'Hy = Y'HY\beta + Y'Hu.$$

[6] This is not rigourous language since d is random.

Taking the matrix HY as a matrix of instrumental variables for Y, we get:

$$\hat{\beta}_{\text{IV}} = (Y'HY)^{-1}Y'Hy = \tilde{\beta}.$$

HY is asymptotically uncorrelated with the disturbances u under the following assumption (5.18). Moreover, when $C_\beta(\alpha)$ is obtained from inverting the AR statistic, then under the usual assumptions,

$$\left(\frac{\bar{X}'\bar{X}}{T}, \frac{\bar{X}'u}{T}, \frac{\bar{X}'V}{T}\right) \xrightarrow[T\to\infty]{\text{p}} (Q_{\bar{X}\bar{X}}, 0, 0)\,, \tag{5.18}$$

$$\frac{\bar{X}'u}{\sqrt{T}} \xrightarrow[T\to\infty]{L} N(0, \sigma_u^2 Q_{\bar{X}\bar{X}})\,, \tag{5.19}$$

it is easy to show that if Π_2 is of full rank. Then

$$\sqrt{T}(\tilde{\beta} - \beta) \xrightarrow[T\to\infty]{L} N\big[0, \sigma_u^2 \operatorname*{plim}_{T\to\infty}\big(\frac{1}{T}A\big)^{-1}\big] \tag{5.20}$$

with

$$\operatorname*{plim}_{T\to\infty}\frac{1}{T}A = \Pi_2'\big[\,Q_{X_2X_2} - Q_{X_2X_1}Q_{X_1X_1}^{-1}Q_{X_2X_1}'\big]\Pi_2 \tag{5.21}$$

$$Q_{X_iX_j} = \operatorname*{plim}_{T\to\infty}\frac{1}{T}X_i'X_j\,,\ i,j = 1,\,2\,. \tag{5.22}$$

On developing the expression of $\tilde{\beta}$, we may also write:

$$\tilde{\beta} = \{Y'[M(X_1) - (1 + f(\alpha))M(\bar{X})]Y\}^{-1}Y'[M(X_1) - (1 + f(\alpha))M(\bar{X})]y.$$

This is the expression of the well-known Theil's k-class estimator [see Davidson and MacKinnon (1993, page 649)] with $k = 1 + f(\alpha)$, and since $f(\alpha)$ tends to 0 when T becomes large, $\tilde{\beta}$ is asymptotically equivalent to the two stage least squares estimator. The later may be written:

$$\hat{\beta}_{\text{2SLS}} = \{Y'[M(X_1) - M(\bar{X})]Y\}^{-1}Y'[M(X_1) - M(\bar{X})]y.$$

Hence when Π_2 is of full rank and the eigenvalues of A are positive, the projection-based confidence sets for $w'\beta$ may be interpreted as a Wald-type confidence intervals based on the statistic (which is asymptotically pivotal):

$$T = \frac{w'\tilde{\beta} - w'\beta}{\sqrt{\hat{\sigma}_u^2(w'A^{-1}w)}}.$$

6. Monte Carlo evaluation

In this section we study the projection based statistical inference on the basis of Monte Carlo simulations. We first study the characteristics, especially the degree of conservatism, of the projection-based confidence sets and then we compare the confidence sets obtained on the basis of different statistics, the Anderson-Rubin statistic (AR) given by (2.7), the asymptotic AR statistic (ARS) given by (2.7) but without assumption (2.5) (it follows asymptotically a $\chi^2_{k_2}/k_2$ distribution) and the LM statistic proposed by Wang and Zivot (1998) and given by (2.10). The statistics AR and ARS are valid for any matrix Π_2 in equation (2.2) while the distribution of the statistic LM is derived under the local to zero assumption ($\Pi_2 = C/\sqrt{T}$, where C is a fixed matrix). For comparison purposes we consider the later hypothesis, while for the study of the characteristics of the confidence sets based on the statistic AR, we consider Π_2 as a fixed matrix.

6.1. Characteristics of AR-type confidence sets

The data generating process is:

$$y = Y_1\beta_1 + Y_2\beta_2 + \gamma X_1 + u\,, \tag{6.1}$$

$$(Y_1, Y_2) = X_2\Pi_2 + X_1\Pi_1 + (V_1, V_2)\,, \tag{6.2}$$

where $\beta_1 = \frac{1}{2}$, $\beta_2 = 1$, $\gamma = 2$, $\Pi_1 = (0.1, 0.2)$, $(u_t, V_{1t}, V_{2t})' \sim i.i.d.\ N(0, \Sigma)$ and

$$\Sigma = \begin{pmatrix} 1 & .2 & .2 \\ .2 & 1 & .2 \\ .2 & .2 & 1 \end{pmatrix}. \tag{6.3}$$

The correlation coefficient r between u and V_i $(i = 1, 2)$ is set equal to 0.2, the variables Y_1 and Y_2 are endogenous, thus the instrumental variables X_2 are necessary.

We consider 3 different matrices Π_2, $\Pi_2 = 0$, Π_2 has rank 1, and Π_2 has full rank. The number of instruments (k_2) vary from 2 to 40. All simulations are based on 10.000 replications. The results are given in Table 1 covering two sample sizes ($T = 50$ and $T = 100$) and two significance levels ($\alpha = 5\%$ and $\alpha = 10\%$).

The third column of Table 1 gives the simulated confidence level of the confidence set for $\beta = (\beta_1, \beta_2)'$ based on the statistic AR. The following columns give some characteristics of the projection-based confidence set for β_1.

The first observation is that the coverage rate of the projection-based confidence sets for β_1 decreases as k_2 increases and tends to the exact confidence level $1 - \alpha$ of the confidence set for β. [7] **Hence the projection-based confidence sets become less conservative as the number of relevant instruments increases.** This suggests use of a number of relevant instruments as large as possible. But on the other hand, as noted by Dufour and Taamouti (2000a) and Kleibergen (2001a), a large number of instruments will induce loss of power for the Anderson-Rubin test for β. Obviously, this should not be true in the extreme case of $\Pi_2 = 0$, X_2 vanishes from equation (6.2).

The proportion of unbounded confidence sets is always greater than $1 - \alpha$ when Π_2 does not have full rank. This is predictable according to the results in Dufour (1997). It is natural when Π_2 does not have full rank to get an unbounded confidence set, β is not identified in this case and the set of possible values is large. Moreover, we observe that this proportion tends to $1 - \alpha$ as k_2 increases. On the other hand,

[7] Recall that theoretically, this rate is always greater than or equal to the confidence level of the set from which the projection is done.

when Π_2 has full rank, the frequency of unbounded confidence sets is nearly zero.

The proportion of empty confidence sets is near 0 when Π_2 has reduced rank and is always smaller than α otherwise. This also is natural since an empty confidence set may be interpreted as a rejection of the model [Dufour (1997)] while in fact our data is generated by the model itself.

In Section 5.2 we showed that when A is nonsingular (which is true with probability 1 when T gets large) and the projection-based confidence set is unbounded, it takes two possible forms, either $\mathbb{R}$, or the union of two semi-closed intervals $]-\infty, c_1] \cup [c_2, \infty[$. When it has the second form, the unbounded confidence set is informative since it excludes many values of β_1 and induces the rejection of all hypotheses that assign to β_1 one of these excluded values. As an illustration, the last column of Table 1 gives the proportion of confidence sets that are different from $\mathbb{R}$ and excludes the value 0, thus inducing the rejection of $H_0 : \beta_1 = 0$.

Column 7 in Table 1 gives the proportion of confidence sets equal to $\mathbb{R}$ and thus uninformative. This rate is nearly 0 when Π_2 has full rank and is decreasing of k_2 when Π_2 does not have full rank.

If we compare the results for the two sample sizes, $T = 50$ and $T = 100$, the general observation is that the results are similar for the two sizes. This is expected given that the AR test considered here is exact. Other simulations not presented here show that the number of explanatory exogenous variables (X_1) have no apparent effect on the characteristics of these confidence sets.

6.2. Comparison on the basis of the test statistic

In this section we consider the same model as before but we replace Π_2 by $C/\sqrt{T}$ as assumed in Staiger and Stock (1997) and Wang and Zivot (1998). We compare projection-based confidence sets obtained from inverting a statistic which is pivotal whatever the number of instruments (namely AR) and a statistic that is boundedly pivotal, we choose the LM statistic in the second case. The LM statistic has an asymptotic distribution following $\chi^2_{k_2}$ when the model is exactly identified ($k_2 = 2$) and is bounded by a $\chi^2_{k_2}$ in the case of over-identification. Unlike the statistic AR,

is not based on assumption (2.5). For this reason we also consider the asymptotic AR statistic (henceforth ARS) which is distributed as $\chi^2_{k_2}/k_2$ [Dufour and Jasiak (2001)]. Obviously, since the statistics LM and ARS have asymptotic distributions, the sample size will matter. For this reason, we consider three different sizes $T = 50$, $T = 100$, and $T = 200$.

Table 2.*c* presents the results for $C = 0$, Table 2.*b* presents the results for a matrix C with components c_{ij} such that $1 < c_{ij} < 5$, and Table 2.*a* with c_{ij} such that $10 \leq c_{ij} \leq 20$. The results for AR based confidence sets are obviously identical to those presented in Table 1. The general observation is that the statistic LM turns to give confidence sets much more conservative than those based on AR or ARS, and unlike the AR statistic, the degree of conservatism of the resulting confidence sets increases with the number of instruments k_2. The coverage rate of the confidence sets based on LM statistic is always greater than 98.5% and approaches rapidly 100% as k_2 increases.[8] This is predictable since the LM based confidence sets are doubly conservative, by majorization of the distribution of LM and by projection.

For ARS based confidence sets, they are similar to those based on AR when the number of instruments is small and T is large. However, as k_2 increases, the coverage rate decreases and gets below 95%. This may be explained by the fact that when k_2 increases, the number of unknown parameters increases too and the asymptotic approximation becomes less accurate.[9]

As we may expect the high coverage rate of the LM based confidence sets induces power loss for the test that rejects $H_0 : \beta_1 = \beta_1^0$ when the projection-based confidence set for β_1 excludes β_1^0. This is shown in Table 3 and Figure 1 where we present the results of simulations of $\mathsf{P}[\text{reject } H_0 : \beta_1 = 1/2 | \beta_1 = \beta_1^i]$ with a decision rule consisting of rejecting H_0 if $\frac{1}{2}$ is excluded from the confidence set for β_1. The theoretical size is 95%. The value of the alternative varies from $-\frac{1}{2}$ to $\frac{3}{2}$ with increments of $\frac{1}{10}$.

For $k_2 = 2$, the three tests have the same power but as k_2 increases, it appears clearly that the LM based test have less power and tends to reject few often. On the

[8] The theoretical confidence level in Table 2 is 95%.

[9] The number of parameters in the model, excluding the covariances is $k_2G + k_1G + k_1 + G$.

other hand, when k_2 increases, the test ARS appears to be powerful but in fact its size becomes greater than α.

7. Empirical applications

In this section we illustrate the statistical inference methods discussed in the previous sections through two empirical applications related to important issues in the macroeconomics literature. The first one concerns the relation between growth and trade examined through cross-country data on a large sample of countries, while the second application is about the returns to scale and externality spillovers in U.S. industry.

7.1. Trade and growth

A large number of cross-country studies in the macroeconomics literature have looked at the relationship between standards of living and openness. Recent literature includes Frankel and Romer (1996, 1999), Harrison (1996), Mankiw, Romer, and Weil (1992) and the survey of Rodrik (1995). Despite the great effort that has been devoted to studying this issue, there is little persuasive evidence concerning the effect of openness on income even if many studies conclude that openness has been conductive to higher growth.

Estimating the impact of openness on income through a cross-country regression raises two basic difficulties. The first is to come up with an appropriate indicator of openness. The most commonly used one is the trade share (ratio of imports or exports to GDP). The second problem is the endogeneity of this indicator. Frankel and Romer (1999) argue that the trade share should be viewed as an endogenous variable, and similarly for the other indicators such as trade policies.

As a solution to this problem, Frankel and Romer (1999) propose to use instrumental variables method to estimate the income-trade regression. The equation studied is given by

$$y_i = a + bT_i + c_1 N_i + c_2 A_i + u_i \tag{7.1}$$

where y_i is log income per person in country i, T_i is the trade share (measured as the ratio of imports and exports to GDP), and N_i and A_i are log population and area. The trade share T_i is considered as endogenous and to instrument for it, the authors use an instrument constructed on the basis of geographic characteristics [see Frankel and Romer (1999, equation (6), page 383)].

The data used lists for each country, the trade share in 1985, the area and population (1985), and its income per person in 1985.[10] The authors focus on two samples. The first is the full 150 countries covered by the Penn World Table, and the second sample is the 98-country sample considered by Mankiw, Romer, and Weil (1992).

In this paper we consider the sample of 150 countries. For this sample, it is not clear how "weak" the instruments are. The F-statistic of the first stage regression

$$T_i = \alpha + \beta Z_i + \gamma_1 N_i + \gamma_2 A_i + \varepsilon_i \tag{7.2}$$

is about 13 [see Frankel and Romer (1999, Table 2, page 385)].

To draw inference on the coefficients of the structural equation (7.1), we can use the Anderson-Rubin method in two ways. First if we are interested only in the coefficient of trade share, we can invert the AR test for $H_0 : b = b_0$ to obtain a quadratic confidence set for b. On the other hand, if we want to build confidence sets for the other parameters of (7.1), we must first use the AR test to obtain a joint confidence set for b and each one of the other parameters and then use the projection approach to obtain confidence sets for each one of these parameters.[11] As assumed in the literature, the observations are considered to be homoskedastic and uncorrelated but not necessarily normal, we use the asymptotic AR test with a χ^2 distribution. The results are as follows.

The 95% quadratic confidence set for the coefficient of trade share, b is given by:

$$C_b(\alpha) = \{b : 0.963b^2 - 4.754b + 1.274 \leq 0\}$$

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

[11] We can not use the AR test to build directly confidence sets for the coefficients of the exogenous variables.

which corresponds to the closed interval

$$C_b(\alpha) = [0.284\,,\ 4.652]\,. \tag{7.3}$$

The p-value of the Anderson-Rubin test for $H_0 : b = 0$ is 0.0244, this means a significant positive impact of trade on income. The instrumental variables estimation of this coefficient is 1.97 with a standard error of 0.99, yielding the confidence interval

$$(\widehat{b}_{\text{IV}} \pm 2\hat{\sigma}_{\hat{b}_{\text{IV}}}) = [-0.01\,,\ 3.95], \tag{7.4}$$

which is not very different from the AR based confidence set. Interestingly, in contrast with (7.3), (7.4) does not exclude zero and may suggest that b is not significantly different form zero.

The joint confidence sets for the coefficient of trade share and each one of the other coefficients of (7.1) are given in Table 4A. All the confidence sets are bounded, a natural outcome since we do not have a serious problem of identification in this model. From this confidence sets we can obtain projection-based confidence sets for each one of the parameters. Using Proposition **5.1**, we obtain the results presented in Table 4B. Even if 0 is included in the confidence sets for the coefficient of openness it is likely that the true value of the coefficient is positive. AR-projection-based confidence sets are conservative so when the level of the joint confidence set is 95% it is likely that the level of the projection is close to 98% (see the simulations in Section 6), but if we compare them to those obtained from t-statistics, they are not really larger.

7.2. Returns to scale and externality spillovers in U.S. industry

One of the widely studied problems in recent macroeconomic literature is the extent of returns to scale and externalities in the U.S. industry. Recent work on these issues includes Hall (1990), Caballero and Lyons (1989, 1992), Basu and Fernald (1995, 1997) and Burnside (1996). The results of these researches and many others

have important implications on many fields of macroeconomics, such as growth and business cycle models.

Burnside (1996) presents a short survey of different specifications of the production function adopted in this literature. One of these specifications considers the following equation:

$$Y_{it} = F(K_{it}, L_{it}, E_{it}, M_{it}) \tag{7.5}$$

where, for each industry i and each period t, Y_{it} is the gross output, K_{it} is the amount of capital services used, L_{it} is the amount of labor, E_{it} is energy used, and M_{it} is the quantity of materials. If we assume that F is a differentiable function and homogeneous of degree ρ, we get the following regression equation [see Burnside (1996)]:

$$\Delta y_{it} = \rho \Delta x_{it} + \Delta a_{it} \tag{7.6}$$

where Δy_{it} is the growth rate of the output, Δx_{it} is a weighted average of the inputs and Δa_{it} represents technological changes.[12] In this specification, ρ is the coefficient that measures the extent of returns to scale. Returns to scale are increasing, constant or decreasing depending on whether $\rho > 1$, $\rho = 1$ or $\rho < 1$.

To identify simultaneously the effects of externalities between industries, Caballero and Lyons (1992) added to the previous regression equation the aggregated industrial output as a measure of this effect. Burnside (1996) suggested a variable based on inputs rather than output, arguing by the fact that the first measure may induce spurious externalities for industries with a large output. Adopting the later suggestion, the previous regression equation becomes:

$$\Delta y_{it} = \rho \Delta x_{it} + \eta \Delta x_t + \Delta a_{it} \tag{7.7}$$

where Δx_t is the cost shares weighted average of the Δx_{it} [Burnside (1996,-equation (2.8))]. The coefficient η measures the externalities effect.

To estimate this equation, Hall (1990) proposed a set of instruments that was

[12] The weights are the production cost shares of each input.

used in most subsequent researches. These instruments include the growth rate of military purchases, the growth rate of world oil price, a dummy variable representing the political party of the president of Unites States and one lag of each of these variables. Estimation methods used include ordinary least squares, two stages least squares and three stages least squares.

The regressions are performed using panel data on two-digit SIC (Standard Industrial Classification) code level manufacturing industries. This classification includes 21 industries. The data set is described in detail by Jorgenson, Gollop, and Fraumeni (1987) and contains information on gross output, labor input, stock of capital, energy use, and materials inputs.

These regressions are interesting as an application for the statistical inference methods developed in this paper because the instruments used appear to be weak and may induce identification problems. These instruments have been studied in detail by Burnside (1996) who showed on the basis of calculations of R^2 and partial R^2 [Shea (1997)], that these instruments are weak. A valid method to draw inference on ρ (returns to scale) and η (externalities) then consists in using an extension of the Anderson-Rubin approach [as suggested in Dufour and Jasiak (2001)] to build a joint confidence set for $(\rho, \eta)'$ and then build through projection individual confidence intervals for ρ and η.[13]

Given this identification problem, we expect unbounded, but valid confidence sets. Using the same data set as Burnside (1996), we find the results presented in Table 5. This table presents the 2SLS and the confidence sets for the returns to scale coefficients and externalities coefficients for the 21 U.S. manufacturing industries over the period 1953-1984. The projection based confidence sets are obtained from joint confidence sets[14] for (ρ, η) of level 90%.

The average estimation over all industries of the coefficients ρ and η are of the same order as those obtained by Burnside (1996).[15] For confidence sets, as expected only 7

[13] As reported in Caballero and Lyons (1989), there is no evidence of serial correlation from either the Durbin-Watson statistic or the Ljung-Box Q statistic.

[14] We used χ^2 as asymptotic distribution for the Anderson-Rubin statistic instead of the Fisher distribution valid under normality and independence assumption.

[15] The small differences may be due to the use of TSLS instead of 3SLS.

among 21 are bounded. For industries 19 (stone, clay and glass) and 26 (instruments), the returns to scale are increasing. For industry 15 (chemicals), the returns to scale are decreasing. For industries 9 (Textile mill products), 12 (furniture and fixtures), 13 (paper and allied), and 23 (electrical machinery) the hypothesis of constant returns to scale is rejected with a significance level smaller than or equal 10%. For industry 10 (apparel) the confidence set is empty which may be explained by the fact that the data does not support the model. For industries 7 (Food and kindred products), 8 (Tobacco), 11 (Lumber and wood), 16 (Petroleum and coal products), 17 (Rubber and miscellaneous plastics), 18 (Leather), and 24 (Motor vehicles), the confidence sets are equal to $\mathbb{R}$ and thus give no information on ρ and η.

8. Conclusion

Many researches in recent econometrics literature have shown that the problem of weak instruments is wider than is expected. The development of new valid statistical inference methods is one of the priorities of ongoing research. Many tests have been proposed for use in the case of uncertainty about the validity of instruments, but unfortunately they have in general the drawback of requiring the specification of all the vector of parameters associated with the endogenous variables of the model. The projection technique is a solution to this drawback. However its use is not simple in general and may require using costly numerical methods.

In this paper we provided a complete close-form solution to the problem of building projection-based confidence sets from Anderson-Rubin-type confidence sets in linear simultaneous equations models. We used the projection technique to build confidence sets for components of the vector of unknown parameters and for linear combinations of these components. We also showed that, when the projection-based confidence intervals are bounded, they may be interpreted as confidence intervals based on k-class estimators where the "standard error" is corrected in a way that depends on the level of the test. Our Monte Carlo simulations show that the projection-based confidence sets are not very conservative and may be informative even when they are

unbounded.

Finally, the results of this paper are valid for all models where the confidence sets of the form (5.4) may be met. Such models include those considered by Dufour and Jasiak (2001), and some of the nonlinear models considered in Dufour and Taamouti (2000a).

9. Appendix: Proofs

Proof of equation (4.13) A is a real symmetric matrix, there exists an orthogonal matrix P such that $A = P'DP$, where D is a diagonal matrix with diagonal elements given by the r nonzero eigenvalues and $G - r$ zero eigenvalues of A. Setting $x = P\beta$, and $\delta = Pb$, we can then write:

$$\begin{aligned}
\beta' A\beta + b'\beta + c &= \beta' P'DP\beta + b'P'P\beta + c = (P\beta)'D(P\beta) + (Pb)'(P\beta) + c \\
&= \sum_{i=1}^{r}\lambda_i x_i^2 + \sum_{i=1}^{G}\delta_i x_i + c = \sum_{i=1}^{r}\lambda_i(x_i^2 + \frac{\delta_i x_i}{\lambda_i}) + \sum_{i=r+1}^{G}\delta_i x_i + c \\
&= \sum_{i=1}^{r}\lambda_i(x_i + \frac{\delta_i}{2\lambda_i})^2 + \sum_{i=r+1}^{G}\delta_i x_i + c - \sum_{i=1}^{r}\frac{\delta_i^2}{4\lambda_i} \\
&= \sum_{i=1}^{r}\lambda_i z_i^2 + \sum_{i=r+1}^{G}\delta_i z_i - d
\end{aligned}$$

where $x = P\beta$, $\delta = Pb$, $d = -\big(c - \sum_{i=1}^{r}\delta_i^2/(4\lambda_i)\big)$ and $z = P\beta + u$, with $u_i = \delta_i/(2\lambda_i)$, if $\lambda_i \neq 0$, and $u_i = 0$ otherwise. □

Proof of Proposition 4.1 The proof of this proposition is a direct consequence of the study of the characteristics of C_β in Section 4. If the eigenvalues of A are all positive (see Section 4.1.1), C_β is either the area inside of an ellipsoid, a single point or an empty set, hence it is bounded. If the eigenvalues of A are negative, C_β is the outside of an ellipsoid (see Section 4.1.2). If they are different from 0 but of different signs, C_β is always unbounded whatever the sign of d (see Section 4.1.3). If A is singular, from (4.13), C_β is unbounded. □

Proof of Proposition 5.1 If $d < 0$, C_β and thus $C_{w'\beta}$ are empty (see Section 4.1.1). If $d = 0$, we have $C_\beta = \{\widetilde{\beta} = -\frac{1}{2}w'A^{-1}b\}$, hence $C_{w'\beta} = \{w'\widetilde{\beta}\}$ and (5.8) holds. We shall now concentrate on the case where $d > 0$.

Since A is positive definite, $C_\beta(\alpha)$ is bounded, its projection on any axis will give a closed interval ($C_\beta(\alpha)$ is connected). To calculate its projection, we consider the

following two problems:

$$\begin{aligned} &(\mathrm{A}) \quad \min_{\beta} w'\beta \quad \text{s.c. } \beta \in C_\beta(\alpha)\,, \\ &(\mathrm{B}) \quad \max_{\beta} w'\beta \quad \text{s.c. } \beta \in C_\beta(\alpha)\,. \end{aligned}$$

Since $C_\beta(\alpha)$ is closed and bounded and $w'\beta$ is linear, the solutions to these two problems are not interior (see Lemma **5.6**), hence the constraint is binding. The Lagrangian is:

$$\mathcal{L} = w'\beta - \lambda(\beta' A\beta + b'\beta + c)$$

The first order conditions are:

$$\frac{\partial \mathcal{L}}{\partial \beta} = w - 2\lambda A\beta - \lambda b = 0\,, \tag{9.1}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -(\beta' A\beta + b'\beta + c) = 0\,. \tag{9.2}$$

From (9.1), we get

$$\beta = \frac{1}{2}A^{-1}(-b + \frac{w}{\lambda})\,. \tag{9.3}$$

Replacing β in (9.2), we obtain:

$$\lambda^2 = \frac{w'A^{-1}w}{4d}$$

where $d = \frac{1}{4}(b'A^{-1}b) - c$. Since A is positive definite, A^{-1} is also positive definite, we have $w'A^{-1}w > 0$ and the values of β that solve the problems (A) and (B) must satisfy:

$$\beta^* = \frac{1}{2}A^{-1}\big(-b \mp w\sqrt{\frac{4d}{w'A^{-1}w}}\big) = -\frac{1}{2}A^{-1}b \mp A^{-1}w\sqrt{\frac{d}{w'A^{-1}w}},$$

yielding the following minimum and maximum :

$$w'\beta^* = -\frac{1}{2}w'A^{-1}b \mp \sqrt{d(w'A^{-1}w)}. \tag{9.4}$$

The Hessian is then:

$$H_e = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial \lambda^2} & \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta'} \\ \frac{\partial^2 \mathcal{L}}{\partial \lambda \partial \beta} & \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta'} \end{pmatrix} = \frac{-1}{\lambda} \begin{pmatrix} 0 & w' \\ w & 2\lambda^2 A \end{pmatrix}$$

[where we use from (9.1): $2A\beta + b = \frac{w}{\lambda}$]. Since A is positive definite, the second order conditions will hold by taking λ negative for the minimization problem and λ positive for the maximization problem.[16] A confidence interval of level greater than or equal $1 - \alpha$ for $w'\beta$, is given by:

$$C_{w'\beta}(\alpha) = [-\frac{1}{2} w' A^{-1} b - \sqrt{dw' A^{-1} w} \ , \ -\frac{1}{2} w' A^{-1} b + \sqrt{dw' A^{-1} w}] \, .$$

□

Proof of Proposition 5.3 Consider the optimization problems:

$$\text{(A)} \quad \min_{\beta} w'\beta \quad \text{s.c. } \beta' A \beta + b'\beta + c \leq 0 \, ,$$

$$\text{(B)} \quad \max_{\beta} w'\beta \quad \text{s.c. } \beta' A \beta + b'\beta + c \leq 0 \, ,$$

and assume that the matrix A has one negative eigenvalue (λ_G) and $G - 1$ positive eigenvalues. Since A is not positive definite, the global minimum and maximum of these problems are infinite. To show that the confidence set takes the form given in Proposition **5.3**, it suffices to show that the optimization problems admit finite solutions $\beta^*_{\min}$ and $\beta^*_{\max}$ corresponding to local extrema. In this case, the projection on the axis $w'\beta$ takes the form $]-\infty \, , \ w'\beta^*_{\max}] \cup [w'\beta^*_{\min} \, , \ +\infty[$.

First, changing the argument β by $z = P(\beta + \frac{1}{2} A^{-1} b)$, the previous optimization

[16] For $\lambda < 0$ (alternatively, $\lambda > 0$), H_e equals a positive scalar times the Hessian of the problem $\min_x \lambda^2(x'Ax + b'x + c)$ s.c. $w'x = 0$ (alternatively, $\max_x$ -$\lambda^2(x'Ax + b'x + c)$ s.c. $w'x = 0$) ; see Boot (1964, page 26).

problems may be written:[17]

$$\min_z \left(\max_z\right) \; a'z$$
$$\text{s.c. } \lambda_1 z_1^2 + \cdots + \lambda_{G-1} z_{G-1}^2 - |\lambda_G| z_G^2 - d \leq 0$$

with $a = Pw$ and λ_G is assumed (without loss of generality) to be the negative eigenvalue of A. Since $d < 0$, for the constraint to be verified we should have $z'Dz = \lambda_1 z_1^2 + \cdots + \lambda_{G-1} z_{G-1}^2 - |\lambda_G| z_G^2 < 0$. On the other hand since we are looking for local extrema, the solution will be such that $a'z$ is positive for the minimization problem and negative for the maximization problem, otherwise we can always improve the solution (by multiplying $z_1, ..., z_G$ by a scalar $1 + \varepsilon$, with ε positive). On the same hand, the constraint must be binding otherwise we can always improve locally the solution by multiplying $z_1, \ldots, z_G$ by $1 - \varepsilon$. Hence we may use the Lagrangian to solve these problems for the local extrema:

$$\mathcal{L} = a'z - \alpha(z'Dz - d)\,.$$

The first order conditions are:

$$\frac{\partial \mathcal{L}}{\partial z} = a - 2\alpha Dz = 0\,, \tag{9.5}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = -(z'Dz - d) = 0\,. \tag{9.6}$$

Combining these two equations we get:[18]

$$\alpha^2 = a'D^{-1}a/(4d)\,. \tag{9.7}$$

Since $d < 0$, the problem admits a finite solution if $w'A^{-1}w = a'D^{-1}a < 0$. In this case the solution verifying the FOC is given by: $z^* = \frac{1}{2\alpha} D^{-1}a$. For the second order

[17] P is an orthogonal matrix such that $A = P'DP$, where D is diagonal.
[18] Since A is nonsingular, D is also nonsingular.

conditions, the Hessian is

$$H_e = \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial \alpha^2} & \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial z'} \\ \frac{\partial^2 \mathcal{L}}{\partial \alpha \partial z} & \frac{\partial^2 \mathcal{L}}{\partial z \partial z'} \end{pmatrix} = \begin{pmatrix} 0 & -\frac{1}{\alpha}a' \\ -\frac{1}{\alpha}a & 2\alpha D \end{pmatrix} = -\frac{1}{\alpha} \begin{pmatrix} 0 & a_1 & a_2 & \dots & a_G \\ a_1 & 2\alpha^2\lambda_1 & 0 & \dots & 0 \\ a_2 & 0 & 2\alpha^2\lambda_2 & \dots & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ a_G & 0 & 0 & \dots & 2\alpha^2\lambda_G \end{pmatrix}.$$

The bordered principal minors of H_e are:

$$B_i = (-1)^i 2^{i-1} \alpha^{i-3} \lambda_1 \lambda_2 \dots \lambda_i \left(\frac{a_1^2}{\lambda_1} + \dots + \frac{a_i^2}{\lambda_i}\right), \; i = 2, \dots, G.$$

Since $\lambda_1, \dots, \lambda_{G-1}$ are positive and $(\frac{a_1^2}{\lambda_1} + \dots + \frac{a_G^2}{\lambda_G}) = a'D^{-1}a < 0$, the second order conditions hold for the minimization problem $(B_i < 0, i = 2, \dots, G)$ for $\alpha = -\sqrt{a'D^{-1}a/(4d)}$ and for the maximization problem $[(-1)^i B_i > 0, i = 2, \dots, G]$ for $\alpha = \sqrt{a'D^{-1}a/(4d)}$. From (9.5) and (9.7), the solutions of the previous problems are given by

$$z^*_{\min} = -\left(\frac{d}{a'D^{-1}a}\right)^{1/2} D^{-1}a\,, \quad z^*_{\max} = \left(\frac{d}{a'D^{-1}a}\right)^{1/2} D^{-1}a\,.$$

Replacing z by $P(\beta + \frac{A^{-1}b}{2})$ and a by Pw the solutions are:

$$w'\beta^*_{\min} = -\frac{1}{2}w'A^{-1}b + \sqrt{d(w'A^{-1}w)}\,, \quad w'\beta^*_{\max} = -\frac{1}{2}w'A^{-1}b - \sqrt{d(w'A^{-1}w)}.$$

Hence the projection on the axis $w'\beta$ yields:

$$]-\infty, \; -\frac{1}{2}w'A^{-1}b - \sqrt{d(w'A^{-1}w)}] \cup [-\frac{1}{2}w'A^{-1}b + \sqrt{d(w'A^{-1}w)}\,, \; +\infty[\,.$$

If $d \geq 0$, we will now show that $C_{a'z} = \mathbb{R}$ where $C_{a'z}$ is given in (5.7). To see this, let x be any real number and assume (without loss of generality) that $a_{i_1} \neq 0$ where $i_1 \in \{1, 2, \dots, G\}$. If $i_1 = G$, we can choose z as follows: $z_G = x/a_G$ and $z_i = 0, i < G$.

Since $\lambda_G < 0$ and $d \geq 0$, z then verifies: $a'z = x$ and $\lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 = \lambda_1 z_1^2 \leq d$. If $i_1 \neq G$, then we can choose z as follows: (1) if $a_G \neq 0$, we set $z_{i_1} = \sqrt{d/\lambda_{i_1}}$, $z_G = (x - a_{i_1} z_{i_1})/a_G$, and $z_j = 0$, for $j \neq i_1$ and $j \neq G$; (2) if $a_G = 0$, we set $z_{i_1} = x/a_{i_1}$ and z_G such that $\lambda_G z_G^2 \leq d - (x/a_{i_1})^2$, and $z_j = 0$, for $j \neq i_1$ and $j \neq G$ (where z_G does exist since $\lambda_G < 0$). In both these cases, z then verifies: $a'z = x$ and $\lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_G z_G^2 = \lambda_1 z_1^2 \leq d$. All values of x can thus be attained in $C_{a'z}$, hence $C_{w'\beta}(\alpha) = C_{a'z} = \mathbb{R}$. □

Proof of Proposition 5.4 We need to show that $C_{a'z} = \mathbb{R}$. To see this, assume that λ_{i_1} and λ_{i_2} are the two negative eigenvalues of the matrix A, and (without loss of generality) that the first component of a is $a_1 \neq 0$. Let x be any real number. We will show that $x \in C_{a'z}$.

If λ_{i_1} or λ_{i_2} is associated with z_1 (say it is λ_{i_1}), we can set:

$$z_1 = \frac{x - a_{i_2} z_{i_2}}{a_1}, \; z_i = 0, \text{ for } i \neq i_1 \text{ and } i > 1,$$
$$\text{and } z_{i_2} \text{ such that } \lambda_1 \left(\frac{x - a_{i_2} z_{i_2}}{a_1} \right)^2 + \lambda_{i_2} z_{i_2}^2 \leq d .$$

Since λ_{i_1} and λ_{i_2} are negative, z_{i_2} does exist. The vector z verifies (4.4) and $a'z = x$, hence $x \in C_{a'z}(\alpha)$. We conclude that $C_{w'\beta}(\alpha) = C_{a'z} = \mathbb{R}$.

If none of λ_{i_1} and λ_{i_2} is associated with z_1, we can set:

$$z_1 = x/a_1 \, , \; z_i = 0, \text{ for } i \neq i_1, \; i \neq i_2 \text{ and } i > 1,$$
$$\text{and } z_{i_1} \text{ and } z_{i_2} \text{ such that } \lambda_{i_1} z_{i_1}^2 + \lambda_{i_2} z_{i_2}^2 \leq d - \lambda_1 (x/a_1)^2 \text{ and } a_{i_1} z_{i_1} + a_{i_2} z_{i_2} = 0 .$$

Since λ_{i_1} and λ_{i_2} are negative, z_{i_1} and z_{i_2} will always exist. Thus $x \in C_{a'z}(\alpha)$. Again, we conclude that $C_{w'\beta}(\alpha) = C_{a'z} = \mathbb{R}$. □

Proof of Lemma 5.5 By definition, $C_\beta(\alpha)$ is bounded if there exists a positive real number a such that for any $\beta \in C_\beta(\alpha)$, $\beta'\beta = \sum_{i=1}^{G} \beta_i^2 \leq a$. The projection $C_{\beta_i}(\alpha)$ of $C_\beta(\alpha)$ on the axis β_i is bounded if there exists a real number a_i such that $\beta_i^2 \leq a_i$

for any β_i in $C_{\beta_i}(\alpha)$.

The equivalence between (i) and (ii) is obvious. On the other hand, if (iii) is true, (i) and (ii) are also true (take $P = I$ and $\mu = 0$). Suppose now that P is an orthogonal matrix and μ any vector and let $z = P\beta + \mu$, where β is any vector in $C_\beta(\alpha)$, we have:

$$z'z = (P\beta + \mu)'(P\beta + \mu) = \beta'\beta + 2\beta'P'\mu + \mu'\mu$$

where $P'P = I$. If $C_\beta(\alpha)$ is bounded, this expression is always bounded, hence (i) implies (iii). □

Proof of Lemma 5.6 Let $A = \begin{pmatrix} a_{11} & A'_{21} \\ A_{21} & A_{22} \end{pmatrix}$ and $b = (b_1, b'_{(1)})'$, and consider the following different possible cases:

(a) $a_{11} > 0$;

(b) $a_{11} = 0$, $A_{21} = 0$ and $b_1 > 0$;

(c) $a_{11} = 0$, $A_{21} = 0$ and $b_1 < 0$;

(d) $a_{11} = 0$, $A_{21} \neq 0$ and $b_1 > 0$;

(e) $a_{11} = 0$, $A_{21} \neq 0$ and $b_1 < 0$;

(f) $a_{11} = 0$ and $b_1 = 0$;

(g) $a_{11} < 0$.

If β^* is an interior finite solution of the minimization problem (the proof is similar for the maximization problem), then:

$$Q(\beta^*) = \beta^{*\prime}A\beta^* + b'\beta^* + c < 0\,.$$

The question to consider now is whether we can find $\beta^{**} = \beta^* + \Delta$, where $\Delta = (\delta, 0, .., 0)'$, $\delta < 0$ and $Q(\beta^{**}) = 0$.

We have:

$$(\beta^* + \Delta)'A(\beta^* + \Delta) + b'(\beta^* + \Delta) + c = a_{11}\delta^2 + (2\beta^{*\prime}A + b')\Delta + Q(\beta^*)$$

and

$$Q(\beta^{**}) = a_{11}\delta^2 + (2\beta^{*\prime}A_{.1} + b_1)\delta + Q(\beta^*) = 0 \tag{9.8}$$

where $A_{.1}$ is the first column of A : $A_{.1} = (a_{11}, A'_{21})'$.

If (a) is true, equation (9.8) has a positive discriminant $[(2\beta^{*\prime}A_{.1} + b_1)^2 - 4a_{11}Q(\beta^*) > 0$ since $a_{11} > 0$ and $Q(\beta^*) < 0]$. It has two different solutions, one is[19]:

$$\delta_1 = \frac{-(2\beta^{*\prime}A_{.1} + b_1) - \sqrt{(2\beta^{*\prime}A_{.1} + b_1)^2 - 4a_{11}Q(\beta^*)}}{2a_{11}} < 0\,.$$

If (c) is true, from (9.8): $\delta = -Q(\beta^*)/b_1 < 0$. Hence, if a finite solution exists, it must be such that the constraint is binding, otherwise we can improve the solution by replacing β^* by $\beta^* + \Delta$.

Under (b) or (d): $a_{11} = 0$ and $b_1 > 0$;

$$Q(\beta) = \beta_1(b_1 + 2\sum_{j=2}^{G} a_{1j}\beta_j) + \sum_{i=2}^{G} a_{ii}\beta_i^2 + 2\sum\sum_{1<i<j} a_{ij}\beta_i\beta_j + \sum_{i=2}^{G} b_i\beta_i + c \leq 0\,.$$

To see that the solution is infinite, choose $\beta_2 = \ldots = \beta_G = 0$ and β_1 arbitrarily small.

Under (e) $a_{11} = 0$, $A_{21} \neq 0$ and $b_1 < 0$.

$$Q(\beta) = \beta_1(b_1 + 2\sum_{j=2}^{G} a_{1j}\beta_j) + \sum_{i=2}^{G} a_{ii}\beta_i^2 + 2\sum\sum_{1<i<j} a_{ij}\beta_i\beta_j + \sum_{i=2}^{G} b_i\beta_i + c \leq 0\,.$$

As before, it suffices to choose $\beta_2, ..., \beta_G$ such that $b_1 + 2\sum_{j=2}^{G} a_{1j}\beta_j \geq 0$ and make β_1 tend to $-\infty$. Note that here the condition $A_{21} \neq 0$ is necessary.

Under (f): $a_{11} = b_1 = 0$.

$$Q(\beta) = 2\beta_1(\sum_{j=2}^{G} a_{1j}\beta_j) + \sum_{i=2}^{G} a_{ii}\beta_i^2 + 2\sum\sum_{1<i<j} a_{ij}\beta_i\beta_j + \sum_{i=2}^{G} b_i\beta_i + c \leq 0\,.$$

Choose $\beta_2, ..., \beta_G$ such that $\sum_{j=2}^{G} a_{1j}\beta_j \geq 0$, and β_1 can take any negative large value.

[19]Notice that $-(2\beta^{*\prime}A_{.1} + b_1)$ is always smaller than $\sqrt{(2\beta^{*\prime}A_{.1} + b_1)^2 - 4a_{11}Q(\beta^*)}$.

Under (g) : $a_{11} < 0$. Taking $\beta_2 = ... = \beta_G = 0$, the constraint becomes $a_{11}\beta_1^2 + b_1\beta_1 + c \leq 0$. In this case β_1 may be any arbitrary negative large value and still verify the constraint. □

Proof of Proposition 5.7 We shall study in turn the three cases considered by the proposition.

Case (a) _ We suppose first that $a_{11} > 0$.

If A_{22} is positive definite, the Lagrangian is

$$\mathcal{L} = e_1'\beta - \lambda(\beta' A\beta + b'\beta + c)$$

and the first order conditions are:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \beta} &= e_1 - \lambda(2A\beta + b) = 0\,, \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= -(\beta' A\beta + b'\beta + c) = -[\frac{\beta'}{2}(2A\beta + b) + \frac{b'\beta}{2} + c] \\
&= -[\frac{e_1'\beta}{2\lambda} + \frac{b'\beta + 2c}{2}] = 0,
\end{aligned}$$

hence

$$\lambda = \frac{-e_1'\beta}{b'\beta + 2c} = \frac{-\beta_1}{b'\beta + 2c}\,. \tag{9.9}$$

Replacing λ in the first equation, we get the system

$$2\begin{pmatrix} a_{11} & A_{21}' \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} \beta_1 \\ \beta_{(1)} \end{pmatrix} + \begin{pmatrix} b_1 \\ b_{(1)} \end{pmatrix} = \frac{e_1}{\lambda} = \begin{pmatrix} \frac{b'\beta + 2c}{-\beta_1} \\ 0 \end{pmatrix},$$

hence the equations system:

$$\begin{aligned}
2a_{11}\beta_1 + 2A_{21}'\beta_{(1)} + b_1 &= \frac{b'\beta + 2c}{-\beta_1} \\
2A_{21}\beta_1 + 2A_{22}\beta_{(1)} + b_{(1)} &= 0
\end{aligned}$$

or

$$\begin{aligned} -2a_{11}\beta_1^2 - 2A'_{21}\beta_{(1)}\beta_1 - b_1\beta_1 &= b_1\beta_1 + b'_{(1)}\beta_{(1)} + 2c \\ \beta_{(1)} &= \frac{-1}{2}A_{22}^{-1}[2A_{21}\beta_1 + b_{(1)}]\,. \end{aligned}$$

We then obtain a second order equation in β_1 :

$$\beta_1^2[a_{11} - A'_{21}A_{22}^{-1}A_{21}] + \beta_1[b_1 - \Delta_{21}] + c - \frac{b'_{(1)}A_{22}^{-1}b_{(1)}}{4} = 0$$

where $\Delta_{21} = A'_{21}A_{22}^{-1}b_{(1)}$. Since $\det(A) = 0$, and $\det(A) = \det(A_{22})[a_{11} - A'_{21}A_{22}^{-1}A_{21}]$, we have

$$[a_{11} - A'_{21}A_{22}^{-1}A_{21}] = 0\,.$$

We then have

$$\beta_1^* = (c - \Delta_{11})\,/\,(-b_1 + \Delta_{21})$$

where $\Delta_{11} = \frac{b'_{(1)}A_{22}^{-1}b_{(1)}}{4}$, provided that $b_1 - \Delta_{21} \neq 0$; otherwise, $C_{\beta_1}(\alpha) = \mathbb{R}$ or $\emptyset$. We also get

$$\lambda = 1/\,(b_1 - \Delta_{21})\,.$$

The Hessian is

$$H_e = \frac{-1}{\lambda}\begin{pmatrix} 0 & e'_1 \\ e_1 & 2\lambda^2 A \end{pmatrix}.$$

The bordered minors of H_e are:

$$B_i = \frac{2^{i-1}}{\lambda^3}(-\lambda)^i A_{22}(i,i),\ i = 2,\ldots,G$$

where $A_{22}(i,i)$ is the i-th principal minor of A_{22} (which is positive since A_{22} is assumed to be positive definite).

If $b_1 - \Delta_{21} > 0$, then $\lambda > 0$ and we have a maximum:

$$C_{\beta_1}(\alpha) = \,]-\infty\ ,\ (c - \Delta_{11})\,/\,(-b_1 + \Delta_{21})]\,.$$

If $b_1 - \Delta_{21} < 0$, then $\lambda < 0$ and we have a minimum:

$$C_{\beta_1}(\alpha) = [(c - \Delta_{11}) / (-b_1 + \Delta_{21}) \ , \ +\infty [\, .$$

If A_{22} is positive semi-definite (but not positive definite), there exists $\overline{\beta}_{(1)} \neq 0$, such that $\overline{\beta}'_{(1)} A_{22} \overline{\beta}_{(1)} = 0$. Let $\beta^* = (\beta_1, \alpha \overline{\beta}'_{(1)})'$, where α is a real number.

$$\begin{aligned} \beta^{*\prime} A \beta^* + b' \beta^* + c &= a_{11}\beta_1^2 + 2\alpha A'_{21}\overline{\beta}_{(1)}\beta_1 + b_1\beta_1 + \alpha b'_{(1)}\overline{\beta}_{(1)} + c \\ &= \alpha(2\beta_1 A'_{21}\overline{\beta}_{(1)} + b'_{(1)}\overline{\beta}_{(1)}) + a_{11}\beta_1^2 + b_1\beta_1 + c \, . \end{aligned}$$

If $A_{21} \neq 0$ or $b_{(1)} \neq 0$, we can find a number β_1 as large as we wish and that verifies the constraint. It suffices to choose α large and of opposite sign of $(2\beta_1 A'_{21}\overline{\beta}_{(1)} + b'_{(1)}\overline{\beta}_{(1)})$. If $A_{21} = 0$ and $b_{(1)} = 0$,

$$\beta' A \beta + b' \beta + c \leq 0 \Longleftrightarrow a_{11}\beta_1^2 + \beta_1 b_1 + c + \beta'_{(1)} A_{22} \beta_{(1)} \leq 0$$

Following Lemma **5.6**,

$$\begin{aligned} \beta_1^{\min} &= \min_{\beta_{(1)}} \left\{ \frac{-b_1 - \sqrt{b_1^2 - 4a_{11}(\beta'_{(1)} A_{22} \beta_{(1)} + c)}}{2a_{11}} \right\} , \\ \beta_1^{\max} &= \max_{\beta_{(1)}} \left\{ \frac{-b_1 + \sqrt{b_1^2 - 4a_{11}(\beta'_{(1)} A_{22} \beta'_{(1)} + c)}}{2a_{11}} \right\} . \end{aligned}$$

Since A_{22} is positive semi-definite, we get the confidence set:

$$\left[\frac{-b_1 - \sqrt{b_1^2 - 4a_{11}c}}{2a_{11}} , \frac{-b_1 + \sqrt{b_1^2 - 4a_{11}c}}{2a_{11}} \right]$$

provided that $b_1^2 - 4a_{11}c > 0$. If $b_1^2 - 4a_{11}c < 0$, then $C_\beta(\alpha) = \emptyset$.

Finally, if A_{22} is not positive semi-definite, there exists $\overline{\beta}_{(1)}$ such that $\overline{\beta}'_{(1)} A \overline{\beta}_{(1)} < 0$. It is easy to see in this case that the confidence set induced by the previous optimization problems is $\mathbb{R}$.

Case (b) _ Suppose now that $a_{11} = 0, A_{21} = 0$ and $b_1 > 0$. We have:

$$\begin{aligned} \beta' A\beta + b'\beta + c \leq 0 &\iff b_1\beta_1 + \beta'_{(1)}A_{22}\beta_{(1)} + b'_{(1)}\beta_{(1)} + c \leq 0 \\ &\iff \beta_1 \leq -\frac{1}{b_1}\left(\beta'_{(1)}A_{22}\beta_{(1)} + b'_{(1)}\beta_{(1)} + c\right) . \end{aligned}$$

It is clear that $\beta_1^{\min} = -\infty$ and

$$\beta_1^{\max} = \max_{\beta_{(1)}} \left\{ \frac{\beta'_{(1)}A_{22}\beta_{(1)} + b'_{(1)}\beta_{(1)} + c}{-b_1} \right\} = -\min_{\beta_{(1)}} \left\{ \frac{\beta'_{(1)}A_{22}\beta_{(1)} + b'_{(1)}\beta_{(1)} + c}{b_1} \right\}$$

If A_{22} is positive definite, $\beta^*_{(1)} = \frac{1}{2}A_{22}^{-1}b_{(1)}$ and

$$C_{\beta_1}(\alpha) = \left] -\infty, -\frac{1}{b_1}(c - \Delta_{11}) \right] .$$

If A_{22} is positive semi-definite and $b_{(1)} = 0$,

$$C_{\beta_1}(\alpha) =]-\infty,\ -c/b_1\,] .$$

Otherwise $C_{\beta_1}(\alpha) = \mathbb{R}$.

Case (c) _ Finally, suppose that $a_{11} = 0$, $A_{21} = 0$ and $b_1 < 0$. Thus,

$$\beta' A\beta + b'\beta + c = b_1\beta_1 + \beta'_{(1)}A_{22}\beta_{(1)} + b'_{(1)}\beta_{(1)} + c \leq 0\,.$$

In this case $\beta_1^{\max} = +\infty$, and

$$\beta_1^{\min} = \min_{\beta_{(1)}} \left\{ -\frac{1}{b_1}\left(\beta'_{(1)}A_{22}\beta_{(1)} + b'_{(1)}\beta_{(1)} + c\right) \right\} . \tag{9.10}$$

If A_{22} is positive definite,

$$C_{\beta_1}(\alpha) = \left[-\frac{1}{b_1}(c - \Delta_{11})\ ,\ +\infty \right[. \tag{9.11}$$

If A_{22} is positive semi-definite and $b_{(1)} = 0$, then $C_{\beta_1}(\alpha) = [\,-c/b_1\ ,\ +\infty\,[$. Otherwise

$C_{\beta_1}(\alpha) = \mathbb{R}$. □

Table 1 : Characteristics of *AR*-projection-based confidence sets

(a) : $T = 50, \alpha = 5\%$

			Confidence set (CS) for β_1				
Π	k_2	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	% of unbounded (#R) CS excluding 0
Full rank	2	94.65	98.29	0.15	0.00	0.09	33.33
	3	94.81	98.04	0.00	0.46	0.00	.
	4	94.87	97.53	0.00	0.93	0.00	.
	5	95.46	97.48	0.00	1.11	0.00	.
	6	94.71	97.09	0.00	1.41	0.00	.
	7	95.05	96.85	0.00	1.80	0.00	.
	8	95.09	97.03	0.00	1.70	0.00	.
	9	94.94	96.52	0.00	2.06	0.00	.
	10	94.63	96.42	0.00	2.18	0.00	.
	15	94.51	96.06	0.00	2.76	0.00	.
	20	94.68	95.85	0.00	3.20	0.00	.
	30	94.60	95.27	0.00	4.03	0.00	.
	40	95.11	95.56	0.00	3.99	0.00	.
rank(Π) = 1	2	95.04	98.64	98.59	0.00	95.00	35.10
	3	95.44	98.05	97.84	0.00	92.71	30.99
	4	95.22	97.62	97.40	0.02	90.82	32.52
	5	95.08	97.52	97.25	0.01	90.32	32.04
	6	94.93	97.09	96.90	0.01	88.70	31.14
	7	95.07	97.10	96.85	0.02	88.66	31.99
	8	95.36	96.86	96.71	0.05	88.09	31.90
	9	94.94	96.55	96.62	0.02	87.39	31.85
	10	94.87	96.67	96.31	0.04	87.10	28.99
	15	94.78	96.04	96.49	0.04	86.25	33.69
	20	95.29	96.20	96.00	0.07	84.45	31.29
	30	94.66	95.39	95.68	0.08	82.52	29.94
	40	94.82	95.23	95.47	0.08	79.85	24.97
$\Pi = 0$	2	95.71	99.96	99.97	0.00	99.84	23.08
	3	94.52	99.94	99.98	0.00	99.84	42.86
	4	95.06	99.93	99.96	0.00	99.74	31.82
	5	95.61	99.96	99.97	0.00	99.77	10.00
	6	94.50	99.91	99.92	0.00	99.66	15.38
	7	95.16	99.96	99.91	0.00	99.75	18.75
	8	95.27	99.89	99.94	0.00	99.60	32.35
	9	95.00	99.91	99.90	0.01	99.54	22.22
	10	94.72	99.92	99.98	0.00	99.71	14.81
	15	95.27	99.94	99.95	0.00	99.60	14.29
	20	95.06	99.94	99.91	0.00	99.60	25.81
	30	95.05	99.93	99.90	0.00	99.42	20.83
	40	95.08	99.88	99.92	0.00	99.40	21.15

Table 1 (cont'd): Characteristics of *AR*-projection-based confidence sets

(b) : $T = 50, \alpha = 10\%$

			Confidence set (CS) for β_1				
Π	k_2	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	% of unbounded (#R) CS excluding 0
Full rank	2	90.06	96.78	0.06	0.00	0.03	0.00
	3	91.18	95.67	0.00	0.91	0.00	.
	4	89.82	94.75	0.00	2.19	0.00	.
	5	90.23	94.30	0.00	2.76	0.00	.
	6	90.08	94.01	0.00	3.44	0.00	.
	7	90.01	93.42	0.00	3.87	0.00	.
	8	89.64	93.07	0.00	4.16	0.00	.
	9	90.44	93.51	0.00	4.09	0.00	.
	10	89.98	92.93	0.00	4.60	0.00	.
	15	89.49	91.66	0.00	6.18	0.00	.
	20	90.03	91.88	0.00	6.56	0.00	.
	30	89.65	90.84	0.00	7.90	0.00	.
	40	90.40	91.01	0.00	8.13	0.00	.
rank(Π) = 1	2	89.79	96.65	96.42	0.00	88.96	37.00
	3	90.00	95.61	95.52	0.01	85.63	93.18
	4	90.21	95.07	94.91	0.04	83.53	37.43
	5	89.74	94.52	94.09	0.08	81.57	36.10
	6	89.72	93.60	94.10	0.08	79.54	35.86
	7	90.33	93.62	93.91	0.07	79.61	37.41
	8	89.99	93.32	94.06	0.11	79.12	35.81
	9	89.89	92.99	93.28	0.10	77.69	36.43
	10	89.85	92.63	93.02	0.13	77.03	35.96
	15	90.21	92.26	92.61	0.24	75.22	36.28
	20	90.21	92.09	92.00	0.20	72.14	35.24
	30	89.81	91.05	91.21	0.22	69.37	34.20
	40	90.25	90.90	90.49	0.46	66.45	31.78
$\Pi = 0$	2	90.26	99.84	99.84	0.00	99.36	16.67
	3	89.53	99.81	99.89	0.00	99.35	37.04
	4	89.95	99.73	99.67	0.00	98.89	34.62
	5	89.93	99.67	99.73	0.01	98.93	25.00
	6	90.08	99.67	99.82	0.00	98.77	31.43
	7	89.67	99.76	99.66	0.00	98.92	27.03
	8	89.74	99.68	99.65	0.00	98.58	25.23
	9	89.35	99.64	99.73	0.00	98.54	36.94
	10	90.18	99.72	99.72	0.00	98.68	23.08
	15	90.03	99.64	99.79	0.00	98.60	23.52
	20	90.17	99.63	99.60	0.01	98.18	21.13
	30	89.88	99.65	99.58	0.01	98.09	20.81
	40	90.18	99.58	99.41	0.02	97.61	20.00

Table 1 (cont'd): Characteristics of *AR*-projection-based confidence sets

(c) : $T = 100, \alpha = 5\%$

			Confidence set (CS) for β_1				
Π	k_2	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	% of unbounded ($\#R$) CS excluding 0
Full rank	2	95.22	98.66	0.00	0.00	0.00	.
	3	95.33	98.11	0.00	0.44	0.00	.
	4	95.12	97.64	0.00	0.86	0.00	.
	5	94.92	97.47	0.00	1.30	0.00	.
	6	95.40	97.30	0.00	1.32	0.00	.
	7	94.86	96.70	0.00	1.62	0.00	.
	8	95.21	97.31	0.00	1.63	0.00	.
	9	94.73	96.66	0.00	1.88	0.00	.
	10	94.82	96.63	0.00	2.10	0.00	.
	15	94.94	96.26	0.00	2.66	0.00	.
	20	95.00	96.03	0.00	2.87	0.00	.
	30	94.97	95.94	0.00	3.27	0.00	.
	40	94.62	95.55	0.00	3.84	0.00	.
rank(Π) = 1	2	95.07	98.54	98.36	0.00	94.82	35.03
	3	94.92	97.92	97.81	0.00	92.83	38.15
	4	94.69	97.44	97.36	0.03	90.92	32.14
	5	95.04	97.51	97.56	0.00	90.69	32.89
	6	94.64	97.06	97.36	0.00	89.78	33.90
	7	95.22	97.12	97.22	0.01	89.71	35.82
	8	94.39	96.66	96.54	0.04	88.58	35.56
	9	94.77	96.75	96.87	0.05	88.24	35.69
	10	95.23	96.80	96.88	0.00	88.49	32.18
	15	94.98	96.33	96.49	0.04	86.95	31.34
	20	95.09	96.08	96.40	0.05	86.54	33.26
	30	94.80	95.87	96.11	0.03	84.62	32.81
	40	94.56	95.42	95.95	0.02	83.05	34.03
$\Pi = 0$	2	95.26	99.94	99.97	0.00	99.85	41.67
	3	95.25	99.98	99.93	0.00	99.72	9.52
	4	94.84	99.94	99.96	0.00	99.81	33.33
	5	95.28	99.92	99.98	0.00	99.73	28.00
	6	95.00	99.91	99.97	0.00	99.77	55.00
	7	94.94	99.98	99.95	0.00	99.73	13.64
	8	95.14	99.91	99.96	0.00	99.74	22.73
	9	95.21	99.94	99.90	0.00	99.56	11.76
	10	95.21	99.93	99.92	0.00	99.70	22.73
	15	94.62	99.90	99.91	0.00	99.61	16.67
	20	95.11	99.94	99.93	0.00	99.68	36.00
	30	94.76	99.86	99.93	0.00	99.52	21.95
	40	94.57	99.92	99.91	0.00	99.63	21.43

Table 1 (cont'd): Characteristics of *AR*-projection-based confidence sets

(d) : $T = 100, \alpha = 10\%$

			Confidence set (CS) for β_1				
Π	k_2	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	% of unbounded (#R) CS excluding 0
Full rank	2	90.23	96.74	0.00	0.00	0.00	.
	3	90.20	95.62	0.00	1.18	0.00	.
	4	89.73	94.97	0.00	2.20	0.00	.
	5	90.03	94.41	0.00	2.56	0.00	.
	6	90.05	94.12	0.00	3.05	0.00	.
	7	90.01	93.84	0.00	3.43	0.00	.
	8	89.57	93.05	0.00	4.12	0.00	.
	9	89.88	93.19	0.00	4.30	0.00	.
	10	90.28	93.11	0.00	4.51	0.00	.
	15	89.84	92.64	0.00	5.33	0.00	.
	20	89.61	91.74	0.00	6.21	0.00	.
	30	90.10	91.88	0.00	6.53	0.00	.
	40	89.69	90.83	0.00	8.04	0.00	.
rank(Π) =	2	90.02	97.08	96.73	0.00	89.72	35.81
	3	90.41	95.84	95.64	0.03	86.25	38.02
	4	90.17	94.96	94.80	0.05	83.65	34.44
	5	90.24	94.43	94.40	0.09	82.14	36.87
	6	89.84	93.94	94.20	0.09	80.37	35.43
	7	90.05	93.69	93.57	0.15	80.49	34.18
	8	90.30	93.76	93.28	0.11	78.85	36.80
	9	89.83	93.39	93.35	0.17	78.64	37.32
	10	89.82	92.82	92.91	0.17	77.83	38.92
	15	89.96	92.45	92.64	0.16	76.05	33.75
	20	90.04	92.05	92.12	0.20	74.69	35.06
	30	89.79	91.61	91.53	0.22	72.90	35.80
	40	90.33	91.53	91.22	0.31	72.09	35.81
$\Pi = 0$	2	90.41	99.87	99.92	0.00	99.49	30.23
	3	89.59	99.82	99.87	0.00	99.19	32.35
	4	90.18	99.76	99.76	0.00	98.92	28.57
	5	89.86	99.78	99.80	0.01	99.15	27.69
	6	90.25	99.76	99.78	0.00	99.08	31.43
	7	90.15	99.71	99.80	0.00	98.90	24.44
	8	90.25	99.76	99.65	0.00	98.79	23.26
	9	89.75	99.77	99.72	0.00	98.89	20.48
	10	90.34	99.70	99.75	0.00	98.83	29.35
	15	90.36	99.63	99.70	0.02	98.88	24.39
	20	90.04	99.69	99.71	0.00	98.71	27.00
	30	90.13	99.66	99.63	0.00	98.55	23.15
	40	89.92	99.66	99.55	0.02	98.09	30.14

Table 2 : Comparison between *AR* and *LM* based confidence sets

(a) : $10 < Cij < 20$

		AR					*ASYMPTOTIC AR*					*LM*				
T	k_2	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R
50	2	95.09	98.68	0.00	0.00	0.00	94.10	98.35	0.00	0.00	0.00	95.11	98.68	0.00	0.00	0.00
	3	94.99	97.63	0.00	0.59	0.00	93.87	97.02	0.00	0.71	0.00	98.38	99.55	0.00	0.00	0.00
	4	94.85	97.47	0.00	0.94	0.00	93.26	96.61	0.00	1.24	0.00	99.24	99.81	0.00	0.00	0.00
	5	95.30	97.33	0.00	1.17	0.00	93.25	96.22	0.00	1.74	0.00	99.77	99.95	0.00	0.00	0.00
	6	94.89	96.95	0.00	1.52	0.00	92.49	95.30	0.00	2.47	0.00	99.88	99.99	0.00	0.00	0.00
	7	94.85	96.77	0.00	1.68	0.00	92.08	95.00	0.00	2.57	0.00	99.95	99.99	0.00	0.00	0.00
	8	94.96	96.53	0.00	1.96	0.00	92.30	94.84	0.00	3.17	0.00	100.00	100.00	0.00	0.00	0.00
	9	95.28	97.05	0.00	1.80	0.00	91.69	94.39	0.00	3.53	0.00	99.99	99.99	0.00	0.00	0.00
	10	95.04	96.49	0.00	2.08	0.00	91.66	94.06	0.00	3.77	0.00	100.00	100.00	0.00	0.00	0.00
	15	94.91	96.13	0.00	2.71	0.00	88.92	91.67	0.00	6.14	0.00	100.00	100.00	0.00	0.00	0.00
	20	95.03	96.00	0.00	3.28	0.00	86.88	89.18	0.00	8.54	0.00	100.00	100.00	0.00	0.00	0.00
	30	95.18	95.82	0.00	3.58	0.00	80.57	82.53	0.00	15.48	0.00	100.00	100.00	0.00	0.00	0.00
	40	95.42	95.82	0.00	3.77	0.00	68.24	69.97	0.00	27.89	0.00	100.00	100.00	0.00	0.00	0.00
100	2	95.43	98.83	0.00	0.00	0.00	94.92	98.70	0.00	0.00	0.00	95.43	98.83	0.00	0.00	0.00
	3	95.14	98.02	0.00	0.53	0.00	94.55	97.80	0.00	0.61	0.00	98.29	99.53	0.00	0.00	0.00
	4	95.19	97.83	0.00	0.73	0.00	94.49	97.46	0.00	0.81	0.00	99.12	99.79	0.00	0.00	0.00
	5	95.32	97.63	0.00	1.04	0.00	94.49	97.20	0.00	1.27	0.00	99.74	99.96	0.00	0.00	0.00
	6	94.82	97.11	0.00	1.43	0.00	93.83	96.60	0.00	1.75	0.00	99.87	99.96	0.00	0.00	0.00
	7	95.00	97.01	0.00	1.58	0.00	93.85	96.37	0.00	1.95	0.00	99.92	99.99	0.00	0.00	0.00
	8	94.93	96.84	0.00	1.95	0.00	93.56	95.81	0.00	2.48	0.00	99.95	100.00	0.00	0.00	0.00
	9	94.83	96.72	0.00	1.93	0.00	93.37	95.62	0.00	2.70	0.00	99.99	99.99	0.00	0.00	0.00
	10	95.34	97.26	0.00	1.71	0.00	93.86	96.01	0.00	2.45	0.00	100.00	100.00	0.00	0.00	0.00
	15	95.28	96.67	0.00	2.23	0.00	92.88	94.87	0.00	3.59	0.00	100.00	100.00	0.00	0.00	0.00
	20	95.09	96.30	0.00	2.78	0.00	91.77	93.63	0.00	4.89	0.00	100.00	100.00	0.00	0.00	0.00
	30	95.01	96.01	0.00	3.17	0.00	90.22	91.85	0.00	6.57	0.00	100.00	100.00	0.00	0.00	0.00
	40	95.04	95.80	0.00	3.50	0.00	87.81	89.42	0.00	9.15	0.00	100.00	100.00	0.00	0.00	0.00
200	2	94.92	98.59	0.00	0.00	0.00	94.76	98.53	0.00	0.00	0.00	94.93	98.59	0.00	0.00	0.00
	3	95.13	98.11	0.00	0.38	0.00	94.75	98.00	0.00	0.40	0.00	98.20	99.55	0.00	0.00	0.00
	4	95.09	97.78	0.00	0.84	0.00	94.70	97.55	0.00	0.90	0.00	99.11	99.79	0.00	0.00	0.00
	5	94.97	97.47	0.00	1.22	0.00	94.60	97.29	0.00	1.31	0.00	99.60	99.96	0.00	0.00	0.00
	6	95.13	97.42	0.00	1.31	0.00	94.67	97.10	0.00	1.43	0.00	99.90	99.98	0.00	0.00	0.00
	7	94.83	96.96	0.00	1.65	0.00	94.21	96.66	0.00	1.85	0.00	99.87	99.97	0.00	0.00	0.00
	8	94.89	96.85	0.00	1.59	0.00	94.32	96.35	0.00	1.93	0.00	99.97	100.00	0.00	0.00	0.00
	9	95.15	97.05	0.00	1.60	0.00	94.51	96.55	0.00	1.79	0.00	99.99	99.99	0.00	0.00	0.00
	10	95.14	97.01	0.00	1.74	0.00	94.38	96.49	0.00	2.07	0.00	100.00	100.00	0.00	0.00	0.00
	15	94.93	96.45	0.00	2.40	0.00	93.78	95.36	0.00	3.08	0.00	100.00	100.00	0.00	0.00	0.00
	20	94.82	96.16	0.00	2.85	0.00	93.43	95.05	0.00	3.70	0.00	100.00	100.00	0.00	0.00	0.00
	30	95.29	96.30	0.00	2.88	0.00	93.17	94.66	0.00	4.18	0.00	100.00	100.00	0.00	0.00	0.00
	40	95.05	95.83	0.00	3.36	0.00	92.31	93.42	0.00	5.33	0.00	100.00	100.00	0.00	0.00	0.00

Table 2 (cont'd) : Comparison between *AR* and *LM* based confidence sets

(b) : $1 < Cij < 5$

		AR					*ASYMPTOTIC AR*					*LM*				
T	k_2	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R
50	2	94.70	99.19	97.70	0.00	93.45	93.89	98.93	97.09	0.00	92.22	94.74	99.20	97.71	0.00	92.45
	3	95.13	98.84	96.36	0.01	90.50	93.77	98.48	95.35	0.03	88.09	98.12	99.71	96.50	0.00	92.49
	4	95.04	98.20	95.71	0.01	88.00	93.35	97.52	93.94	0.04	84.63	99.03	99.85	96.11	0.00	92.00
	5	95.06	98.27	94.22	0.03	85.41	93.00	97.48	92.07	0.04	80.69	99.58	99.94	94.95	0.00	90.95
	6	94.80	98.01	93.31	0.09	84.12	92.39	96.89	90.62	0.15	78.50	99.82	99.98	94.35	0.00	90.63
	7	95.45	98.14	93.30	0.08	83.28	93.01	96.87	89.90	0.14	76.85	99.92	99.98	94.50	0.00	91.09
	8	94.83	97.53	91.39	0.04	80.46	91.95	95.78	87.07	0.16	72.76	99.97	100.00	93.06	0.00	89.61
	9	94.45	97.24	91.24	0.10	79.15	91.24	95.23	86.54	0.26	70.29	99.99	99.99	93.60	0.00	90.04
	10	95.31	97.49	91.48	0.06	79.99	91.67	95.16	85.78	0.28	69.62	99.98	100.00	93.89	0.00	90.99
	15	94.89	96.78	87.06	0.19	72.31	89.38	92.91	78.37	0.65	56.04	100.00	100.00	93.10	0.00	90.73
	20	94.96	96.53	85.86	0.24	70.63	86.84	90.33	70.30	1.48	47.53	100.00	100.00	95.15	0.00	93.61
	30	95.16	96.37	88.29	0.23	72.32	80.56	84.20	64.26	3.24	36.48	100.00	100.00	99.89	0.00	99.80
	40	95.16	96.97	93.05	0.10	77.87	69.64	73.86	55.30	7.87	22.71	100.00	100.00	100.00	0.00	100.00
100	2	94.49	99.28	97.55	0.00	93.58	94.03	99.20	97.29	0.00	92.80	94.49	99.29	97.55	0.00	92.58
	3	94.68	98.65	96.30	0.01	90.06	94.08	98.47	95.76	0.01	88.85	97.80	99.65	96.40	0.00	91.95
	4	95.12	98.31	95.67	0.01	88.28	94.12	97.99	94.92	0.01	86.63	99.04	99.86	95.87	0.00	91.91
	5	94.87	98.16	93.33	0.02	84.19	93.69	97.62	92.16	0.02	81.58	99.44	99.94	93.61	0.00	89.12
	6	95.26	97.91	92.60	0.05	83.02	94.14	97.27	91.13	0.07	79.97	99.66	99.88	93.10	0.00	88.56
	7	95.18	97.66	92.27	0.06	81.86	94.00	97.12	90.47	0.06	78.49	99.91	99.98	92.95	0.00	88.58
	8	95.42	97.49	90.42	0.05	79.52	93.92	96.72	88.26	0.09	75.78	99.91	99.97	91.18	0.00	86.80
	9	94.83	97.24	89.66	0.11	78.21	93.39	96.48	87.44	0.25	73.98	99.93	100.00	90.74	0.00	86.97
	10	94.85	97.30	90.10	0.08	78.48	93.32	96.35	87.84	0.14	73.62	99.99	100.00	91.27	0.00	87.61
	15	95.05	96.99	84.21	0.18	70.22	92.41	95.22	79.18	0.43	62.55	100.00	100.00	86.96	0.00	83.19
	20	94.95	96.61	81.18	0.18	65.97	91.85	94.14	73.95	0.41	55.76	100.00	100.00	85.29	0.00	81.59
	30	94.72	96.16	80.16	0.33	64.34	90.06	92.16	68.92	1.00	48.50	100.00	100.00	88.15	0.00	85.85
	40	94.78	95.92	78.73	0.29	62.53	88.40	90.63	63.81	1.53	42.33	100.00	100.00	91.42	0.00	89.84
200	2	95.24	99.25	97.42	0.00	93.35	95.04	99.21	97.22	0.00	93.08	95.24	99.25	97.42	0.00	93.35
	3	95.14	98.61	96.06	0.03	89.45	94.72	98.52	95.81	0.04	88.91	97.90	99.67	96.09	0.00	91.49
	4	94.91	98.44	95.35	0.02	87.81	94.54	98.26	94.96	0.02	87.20	98.95	99.82	95.38	0.00	91.15
	5	94.90	98.16	92.50	0.02	83.57	94.52	97.96	91.93	0.03	82.54	99.51	99.88	92.71	0.00	87.94
	6	94.92	97.81	91.92	0.05	82.37	94.37	97.62	91.24	0.05	81.01	99.69	99.94	92.17	0.00	87.82
	7	95.07	97.83	91.84	0.03	81.60	94.35	97.50	90.99	0.03	79.72	99.83	99.97	92.16	0.00	87.43
	8	95.09	97.39	89.69	0.06	79.02	94.40	97.01	88.70	0.07	77.07	99.90	99.97	89.97	0.00	85.60
	9	94.80	97.34	89.55	0.11	77.73	94.11	96.95	88.15	0.12	75.41	99.91	99.99	90.04	0.00	85.75
	10	95.13	97.27	89.64	0.06	78.51	94.42	96.72	88.25	0.07	76.25	99.97	100.00	90.33	0.00	86.31
	15	95.24	96.99	82.40	0.23	68.98	94.30	96.38	79.81	0.32	65.32	100.00	100.00	83.92	0.00	79.56
	20	95.04	96.70	79.21	0.26	64.93	93.63	95.71	75.42	0.42	59.73	100.00	100.00	81.52	0.00	77.56
	30	94.50	96.04	75.47	0.40	60.22	92.47	94.39	70.68	0.74	52.52	100.00	100.00	79.10	0.00	75.86
	40	95.05	96.32	72.81	0.45	57.85	92.28	94.09	64.50	1.05	47.70	100.00	100.00	78.39	0.00	75.41

Table 2 (cont'd): Comparison between *AR* and *LM* based confidence sets

(c) : $Cij = 0$

		AR					*ASYMPTOTIC AR*					*LM*				
T	k_2	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R	Coverage rate for β	Coverage rate for β_1	% of unbounded CS	% of empty CS	% of CS = R
50	2	94.74	99.98	99.98	0.00	99.83	93.70	99.95	99.95	0.00	99.74	94.75	99.98	99.98	0.00	99.83
	3	94.90	99.93	99.99	0.00	99.79	93.66	99.88	99.92	0.00	99.60	96.73	99.99	99.99	0.00	99.91
	4	95.12	99.94	99.99	0.00	99.74	93.39	99.90	99.95	0.00	99.49	97.91	99.99	99.99	0.00	99.90
	5	95.07	99.94	99.91	0.00	99.64	93.20	99.88	99.86	0.00	99.41	98.59	100.00	99.92	0.00	99.83
	6	95.08	99.92	99.97	0.00	99.73	92.69	99.86	99.88	0.00	99.44	99.01	99.99	99.99	0.00	99.94
	7	94.83	99.94	99.92	0.00	99.65	91.98	99.84	99.83	0.00	99.29	99.25	100.00	99.94	0.00	99.92
	8	94.96	99.96	99.92	0.00	99.67	91.88	99.86	99.78	0.00	99.20	99.39	100.00	99.96	0.00	99.94
	9	95.07	99.87	99.92	0.00	99.67	92.13	99.78	99.77	0.02	99.01	99.72	99.99	99.96	0.00	99.93
	10	94.84	99.94	99.99	0.00	99.71	91.29	99.82	99.87	0.00	99.09	99.72	100.00	100.00	0.00	99.98
	15	95.32	99.95	99.97	0.00	99.66	89.86	99.58	99.58	0.00	98.28	99.94	100.00	100.00	0.00	100.00
	20	94.96	99.94	99.84	0.00	99.55	88.39	99.33	99.20	0.02	97.00	100.00	100.00	100.00	0.00	100.00
	30	95.04	99.91	99.90	0.00	99.44	80.51	98.21	98.25	0.10	93.33	100.00	100.00	100.00	0.00	100.00
	40	94.74	99.94	99.85	0.00	99.35	68.15	94.09	94.39	0.58	82.93	100.00	100.00	100.00	0.00	100.00
100	2	94.73	99.97	99.99	0.00	99.89	94.22	99.96	99.98	0.00	99.87	94.73	99.97	99.99	0.00	99.89
	3	94.92	99.95	99.95	0.00	99.84	94.31	99.93	99.92	0.00	99.80	96.55	99.98	99.95	0.00	99.88
	4	95.37	99.95	99.96	0.00	99.74	94.71	99.91	99.95	0.00	99.67	97.74	99.99	99.96	0.00	99.83
	5	95.51	99.93	99.98	0.00	99.72	94.58	99.89	99.97	0.00	99.58	98.39	99.99	99.98	0.00	99.89
	6	95.19	99.98	99.94	0.00	99.81	94.18	99.93	99.91	0.00	99.80	98.59	100.00	99.94	0.00	99.91
	7	95.35	99.94	99.95	0.00	99.75	94.10	99.91	99.93	0.00	99.69	98.94	99.99	99.95	0.00	99.93
	8	94.74	99.89	99.93	0.00	99.72	93.43	99.85	99.89	0.00	99.53	99.23	99.99	99.95	0.00	99.91
	9	95.14	99.92	99.92	0.00	99.81	93.65	99.84	99.89	0.00	99.36	99.48	100.00	99.93	0.00	99.87
	10	95.21	99.96	99.95	0.00	99.86	93.50	99.92	99.86	0.00	99.51	99.55	100.00	99.96	0.00	99.91
	15	94.62	99.89	99.93	0.00	99.81	91.97	99.77	99.82	0.00	99.16	99.85	100.00	99.97	0.00	99.94
	20	95.02	99.94	99.90	0.00	99.57	91.87	99.78	99.75	0.00	99.06	99.97	100.00	99.97	0.00	99.95
	30	95.08	99.89	99.95	0.00	99.81	90.32	99.54	99.67	0.01	98.41	100.00	100.00	99.98	0.00	99.98
	40	94.93	99.92	99.91	0.01	99.83	88.10	99.38	99.43	0.02	97.66	100.00	100.00	100.00	0.00	100.00
200	2	94.75	99.99	99.95	0.00	99.83	94.52	99.99	99.95	0.00	99.82	94.76	99.99	99.95	0.00	99.83
	3	94.92	99.96	99.96	0.00	99.85	94.52	99.96	99.96	0.00	99.82	96.46	99.97	99.96	0.00	99.89
	4	94.57	99.94	99.95	0.00	99.74	94.26	99.93	99.95	0.00	99.73	97.18	100.00	99.95	0.00	99.91
	5	95.27	99.90	99.94	0.00	99.66	94.82	99.89	99.92	0.00	99.61	98.20	100.00	99.94	0.00	99.85
	6	95.19	99.95	99.97	0.00	99.68	94.60	99.91	99.95	0.00	99.62	98.47	100.00	99.97	0.00	99.90
	7	94.90	99.98	99.94	0.00	99.82	94.43	99.97	99.92	0.00	99.76	99.03	100.00	99.96	0.00	99.88
	8	95.23	99.96	99.93	0.00	99.77	94.57	99.96	99.93	0.00	99.75	99.13	99.99	99.93	0.00	99.90
	9	94.79	99.96	99.95	0.00	99.78	94.05	99.94	99.93	0.00	99.69	99.22	100.00	99.97	0.00	99.92
	10	95.33	99.94	99.90	0.00	99.65	94.49	99.93	99.89	0.00	99.64	99.45	100.00	99.92	0.00	99.86
	15	95.08	99.91	99.93	0.00	99.64	94.00	99.90	99.91	0.00	99.54	99.77	100.00	99.95	0.00	99.89
	20	95.20	99.90	99.93	0.00	99.59	93.75	99.87	99.82	0.00	99.29	99.95	100.00	99.96	0.00	99.95
	30	94.91	99.85	99.87	0.00	99.57	92.76	99.82	99.80	0.01	99.32	100.00	100.00	99.92	0.00	99.91
	40	94.70	99.88	99.93	0.00	99.54	91.73	99.79	99.78	0.00	98.84	99.99	100.00	99.99	0.00	99.98

Table 3: Power of tests induced by projection-based confidence sets

H_0: $\beta_1 = 1/2$

(*a*) $k_2 = 2$ *et* $T = 100$

β_1	*AR*	*ARS*	*LM*
-0.5	1.00	1.00	1.00
-0.4	1.00	1.00	1.00
-0.3	1.00	1.00	1.00
-0.2	1.00	1.00	1.00
-0.1	0.99	0.99	0.99
0	0.97	0.97	0.97
0.1	0.88	0.88	0.88
0.2	0.64	0.66	0.64
0.3	0.31	0.32	0.31
0.4	0.07	0.07	0.07
0.5	0.01	0.02	0.01
0.6	0.07	0.07	0.07
0.7	0.28	0.29	0.28
0.8	0.59	0.61	0.59
0.9	0.83	0.84	0.83
1	0.94	0.94	0.94
1.1	0.98	0.98	0.98
1.2	0.99	0.99	0.99
1.3	1.00	1.00	1.00
1.4	1.00	1.00	1.00
1.5	1.00	1.00	1.00

(*b*) $k_2 = 4$ *et* $T = 100$

β_1	*AR*	*ARS*	*LM*
-0.5	1.00	1.00	1.00
-0.4	1.00	1.00	1.00
-0.3	1.00	1.00	1.00
-0.2	1.00	1.00	1.00
-0.1	1.00	1.00	1.00
0	0.99	0.99	0.98
0.1	0.93	0.94	0.88
0.2	0.72	0.74	0.58
0.3	0.33	0.35	0.19
0.4	0.08	0.09	0.02
0.5	0.02	0.03	0.00
0.6	0.07	0.08	0.02
0.7	0.29	0.31	0.17
0.8	0.60	0.62	0.48
0.9	0.84	0.86	0.77
1	0.95	0.95	0.93
1.1	0.99	0.99	0.98
1.2	1.00	1.00	0.99
1.3	1.00	1.00	1.00
1.4	1.00	1.00	1.00
1.5	1.00	1.00	1.00

(*c*) $k_2 = 8$ *et* $T = 100$

β_1	*AR*	*ARS*	*LM*
-0.5	1.00	1.00	1.00
-0.4	1.00	1.00	1.00
-0.3	1.00	1.00	1.00
-0.2	1.00	1.00	1.00
-0.1	1.00	1.00	1.00
0	1.00	1.00	1.00
0.1	1.00	1.00	0.99
0.2	0.97	0.98	0.87
0.3	0.67	0.71	0.32
0.4	0.15	0.18	0.01
0.5	0.03	0.04	0.00
0.6	0.14	0.16	0.01
0.7	0.61	0.65	0.29
0.8	0.94	0.95	0.81
0.9	1.00	1.00	0.98
1	1.00	1.00	1.00
1.1	1.00	1.00	1.00
1.2	1.00	1.00	1.00
1.3	1.00	1.00	1.00
1.4	1.00	1.00	1.00
1.5	1.00	1.00	1.00

(*d*) $k_2 = 4$ *et* $T = 200$

β_1	*AR*	*ARS*	*LM*
-0.5	1.00	1.00	1.00
-0.4	1.00	1.00	1.00
-0.3	1.00	1.00	1.00
-0.2	1.00	1.00	1.00
-0.1	1.00	1.00	1.00
0	0.99	0.99	0.99
0.1	0.94	0.95	0.91
0.2	0.73	0.74	0.62
0.3	0.35	0.36	0.22
0.4	0.08	0.08	0.02
0.5	0.02	0.02	0.00
0.6	0.07	0.08	0.02
0.7	0.29	0.30	0.18
0.8	0.62	0.63	0.51
0.9	0.87	0.87	0.81
1	0.96	0.96	0.94
1.1	0.99	0.99	0.98
1.2	1.00	1.00	1.00
1.3	1.00	1.00	1.00
1.4	1.00	1.00	1.00
1.5	1.00	1.00	1.00

Table 4. Confidence sets for the coefficients of the Frankel-Romer income-trade equation

A. Bivariate joint confidence sets (size = 95%)

θ	Joint confidence set (95%)
(b, c_1)	$\theta' \begin{pmatrix} 1.78 & -16.36 \\ -16.36 & 257.85 \end{pmatrix} \theta + \begin{pmatrix} -2.23, & -34.50 \end{pmatrix} \theta + 0.19 \leq 0$
(b, c_2)	$\theta' \begin{pmatrix} 3.83 & -34.58 \\ -34.58 & 386.87 \end{pmatrix} \theta + \begin{pmatrix} -10.6, & 69.17 \end{pmatrix} \theta + 2.13 \leq 0$
(b, a)	$\theta' \begin{pmatrix} 38.41 & 33.34 \\ 33.35 & 29.52 \end{pmatrix} \theta + \begin{pmatrix} -611.55, & -537.47 \end{pmatrix} \theta + 2445.58 \leq 0$

B. Projection-based individual confidence intervals (size $\geq$ 95%)

Coefficient	Projection-based confidence sets	IV-based Wald-type confidence sets
Openness	[−0.21, 6.18]	[−0.01, 3.95]
Population	[−0.01, 0.52]	[−0.01, 0.37]
Area	[−0.14, 0.49]	[−0.11, 0.29]
Constant	[2.09, 9.38]	[0.56, 9.36]

Table 5. Confidence sets for the returns to scale and externality coefficients in different U.S. industries (size $\geq 90\%$)

	Returns to scale		Externalities	
Industry*	2SLS	Confidence set	2SLS	Confidence set
7	0.99	$\mathbb{R}$	-0.06	$\mathbb{R}$
8	1.06	$\mathbb{R}$	0.28	$\mathbb{R}$
9	0.61	$]-\infty, 0.56] \cup [2.23, +\infty[$	0.2	$\mathbb{R}$
10	1.09	$\emptyset$	-0.05	$\emptyset$
11	0.86	$\mathbb{R}$	-0.08	$\mathbb{R}$
12	1.13	$]-\infty, 0.58] \cup [1.77, +\infty[$	-0.01	$]-\infty, -0.73] \cup [0.55, +\infty[$
13	0.54	$]-\infty, 0.74] \cup [4.56, +\infty[$	0.61	$]-\infty, -4.51] \cup [0.45, +\infty[$
14	0.93	$[-1.2, 4.23]$	0.23	$[-0.11, 1.05]$
15	0.22	$[-7.36, 0.54]$	1.06	$[0.85, 11.7]$
16	0.34	$\mathbb{R}$	0.29	$\mathbb{R}$
17	1.29	$\mathbb{R}$	-0.31	$\mathbb{R}$
18	0.39	$\mathbb{R}$	0.01	$\mathbb{R}$
19	1.21	$[1, 3.34]$	-0.03	$[-3.16, 0.15]$
20	0.79	$[0.46, 1.01]$	0.42	$[-0.37, 1.51]$
21	0.8	$]-\infty, 2.25] \cup [1.15, +\infty[$	0.3	$]-\infty, -0.13] \cup [4.21, +\infty[$
22	1.16	$[0.73, 1.81]$	0.02	$[-1.41, 0.76]$
23	1.17	$]-\infty, 0.29] \cup [2.47, \infty[$	0.05	$]-\infty, 1.16] \cup [1.72, +\infty[$
24	1.23	$\mathbb{R}$	-0.12	$\mathbb{R}$
25	1.07	$[0.64, 1.55]$	0.1	$[-0.36, 1.6]$
26	1.38	$[1.19, 3.29]$	-0.07	$[-1.5, 0.38]$
27	1.5	$]-\infty, -88.7] \cup [0.48, +\infty[$	-0.51	$]-\infty, 0.12] \cup [102.1, +\infty[$
Mean	0.94		0.11	

(*) 7:Food and kindred products; 8:Tobacco; 9:Textile mill products; 10:Apparel; 11:Lumber and wood; 12:Furniture and fixtures; 13:Paper and allied; 14:Printing; publishing; and allied; 15:Chemicals; 16:Petroleum and coal products; 17:Rubber and misc. plastics; 18:Leather; 19:Stone, clay, glass; 20:Primary metal; 21:Fabricated metal; 22:Machinery, non-electrical; 23:Electrical machinery; 24:Motor vehicles; 25:Transportation equipment & ordnance; 26:Instruments; 27:Misc. manuf.

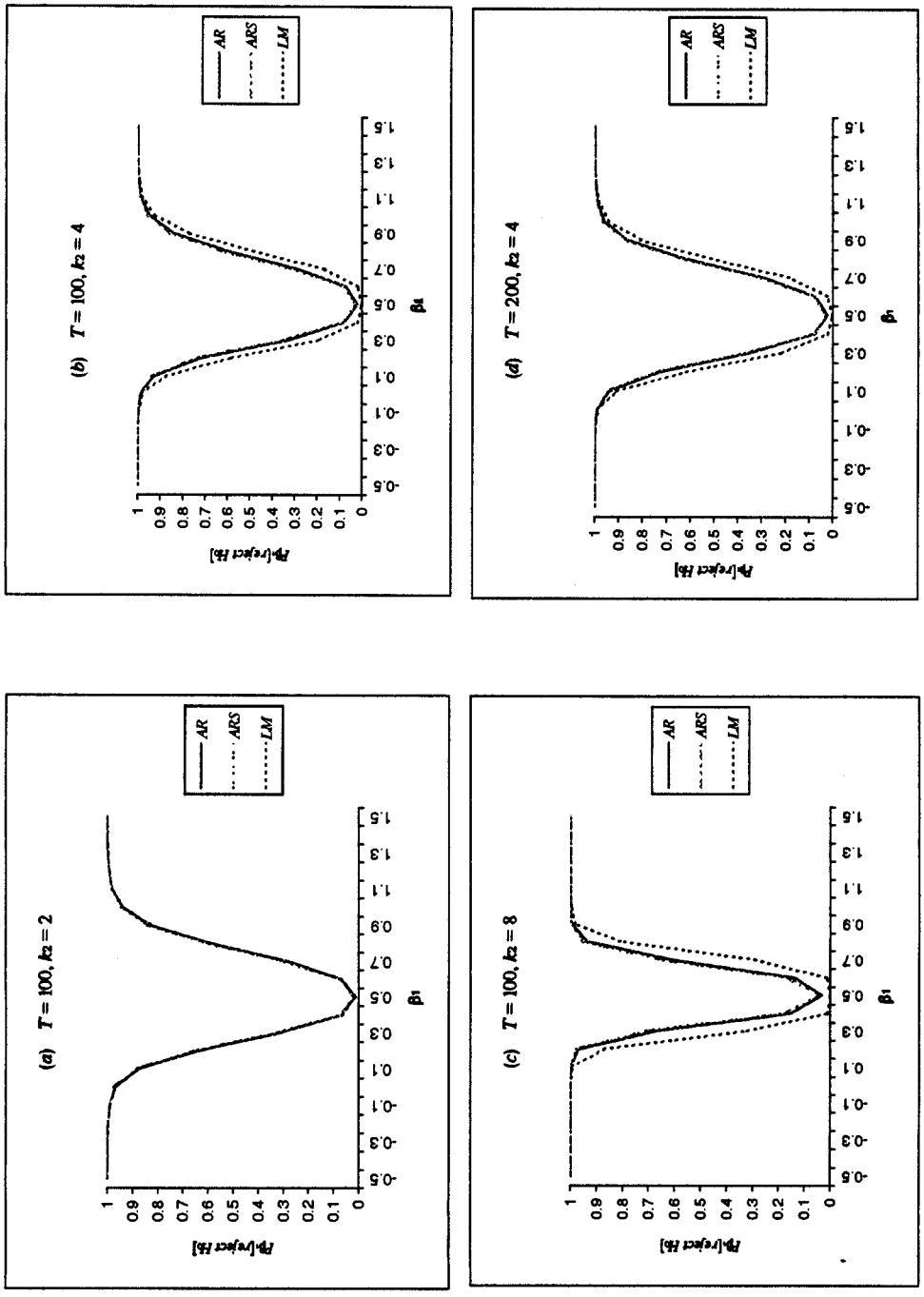


Figure 1: Power of tests induced by projection-based confidence sets
Ho: $\beta_1 = 1/2$

Chapter 2

Point-optimal instruments and generalized Anderson-Rubin procedure in nonlinear models

1. Introduction

In nonlinear econometric models, the statistical tools developed to date in the literature to infer on unknown parameters or on the specification of the model are typically based on asymptotic approximations. Exact inference methods are rare outside of the class of linear models. However, asymptotic approximations are not always valid. The problem is on one hand their validity in small samples and, on the other hand, even if the sample size is large, many recent works [Dufour (1997), Staiger and Stock (1997), Wang and Zivot (1998), Stock and Wright (2000)] showed that when the model has identification problems, the usual asymptotic approximations are not valid. The problem is of the worst kind, false results are accompanied by reported confidence intervals which lend an appearance of great precision. Exact inference methods do not solve the identification problem, but they remain valid in the sense of controlling the levels of tests and confidence sets irrespective of the identification status and sample size. Moreover, tests which are valid for small samples under restrictive assumptions, remain generally valid asymptotically under less restrictive assumptions.

In this paper we consider the statistical inference problem in a model specified by a structural nonlinear equation involving a set of exogenous variables, a set of endogenous variables and a vector of unknown parameters. We propose an exact and simple test which may be inverted to obtain confidence sets for the vector of unknown parameters. This test, which is based on an artificial regression [see Davidson and MacKinnon (1998)] generalizes the approach of Hartley (1964) for hypothesis testing and confidence sets construction in nonlinear regressions and the Anderson and Rubin (1949) procedure in linear structural models. The first generalization is done in the sense of allowing for endogenous explanatory variables, and the second generalization is done in the sense of allowing for nonlinear structural models and more general type of instruments (instruments that depend on the null and alternative hypotheses).

This test does not however avoid a common drawback of the majority of tests robust to identification problems [Dufour (1997), Wang and Zivot (1998), Stock and Wright (2000)]. It requires the specification of the entire vector of parameters, and

it is not adapted to general restrictions on the unknown parameters. One solution to this problem is the projection technique as it is used in Dufour (1990, 1997). This technique allows one to deduce confidence sets for transformations of the unknown parameters from a confidence set for the entire vector. It has been studied in detail in Dufour and Taamouti (2000b) in linear simultaneous equations models.

We also study here the problem of building optimal instruments for testing purposes. The optimal instruments are those that maximize the power of the proposed test. This contrasts with the optimality of instruments well known in the econometric literature [Amemiya (1977)] and which refers to the minimization of the asymptotic variance of an estimator. We derive the power function of the proposed test and generalize a result by Revankar and Mallela (1972) who gave the power function of the Anderson-Rubin test in a more restricted framework. For a given alternative hypothesis we derive the optimal matrix of instruments for testing the simple hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. This belongs to **point-optimal** testing theory [King (1988)]. The maximization of the power function for each alternative allows one to derive the power envelope [King (1988) and Dufour and King (1991)], i.e. the highest achievable power for each point in the alternative for the class of tests under consideration. The resulting matrix of optimal instruments is related, although different, to standard results on IV (and GMM) estimation. This matrix corresponds to $E(\frac{\partial f}{\partial \theta'})$ where f is the function specifying the equation and θ is the vector of unknown parameters [Amemiya (1977), Gallant (1987), Newey (1990, 1993)]. In the literature, this matrix is derived from the minimization of the asymptotic variance of the three stage least squares estimator whereas in this paper it is derived from the maximization of the power function of a test. One of the main differences between the two optimal matrices is that our optimal instruments depend on the alternative hypothesis, yielding **point-optimal instruments**.

The matrix of optimal instruments cannot be used directly in practice since it is unknown in general, it depends on the unknown parameters of the model and on the unspecified distribution of the endogenous variables. As a solution to this problem we use the split-sample technique [Dufour and Jasiak (1993)]. This technique allows

one under the assumption of i.i.d. disturbances to get an exogenous estimation of the optimal instruments and hence to preserve the exact distribution of the test.

The exact distribution of the test is derived under the assumptions of independence and normality of the disturbances, but we show that the test remains valid asymptotically under weaker assumptions similar to those usually assumed to derive asymptotic distributions.

The paper is organized as follows. In section 2, we present the background model and the proposed statistic. In section 3, we derive the point-optimal instruments and deal with its use in practice. Section 4 presents a small Monte Carlo study. In section 5, we show the asymptotic validity of the proposed test without the independence and normality assumptions. Finally, section 6 concludes.

2. Model specification and hypotheses testing

We consider the following equation of interest:

$$f_t(y_t, x_t, \theta) = u_t, \quad t = 1, \ \ldots \ , T, \tag{2.1}$$

where f_t, $t = 1, \ldots, T$ are scalar functions that may have a different form for each observation, θ is a $k \times 1$ vector of unknown parameters, y_t is an $n \times 1$ vector of endogenous variables, x_t is a vector of exogenous variables, and u_t is a disturbance with mean zero. This equation may be viewed as the equation of interest in a model that may include several other equations. It may be estimated by usual estimation methods such as nonlinear two stages least squares [NL2SLS, Amemiya (1974, 1975)] or the generalized method of moments [GMM, Hansen (1982)].

In this paper we consider statistical inference on the vector of unknown parameters θ. Our main objective is to test statistical hypotheses and build confidence sets for θ or for transformations $g(\theta)$, where g is some real function. The usual method for dealing with this problem is to estimate θ by GMM (or NL2SLS) and then use usual asymptotic procedures to test and construct confidence sets for θ [see Newey and West

(1987), Gallant (1987)]. In the presence of identification problems, this approach may have serious problems due to the fact that asymptotic approximations are no longer valid even for large sample sizes [Dufour (1997), Staiger and Stock (1997), Stock and Wright (2000)]. In this paper we propose exact inference methods that are valid even for small samples. Despite the fact that these methods are generally based on more restrictive assumptions compared to asymptotic methods, they are nevertheless reliable and valid especially when we ignore if the model under consideration presents identification problems (which is always the case). To obtain exact distributions of the test statistics, we shall use the following assumption:

$$u_t \overset{i.i.d.}{\sim} N(0,\sigma^2) \quad , \quad t = 1, \ \ldots \ , T. \tag{2.2}$$

2.1. Statistical inference on θ

Consider first an hypothesis of the form

$$H_0 : \theta = \theta_0 \ .$$

If H_0 is true, we have $f_t(y_t, x_t, \theta_0) = u_t$. Thus, if we regress $f_t(y_t, x_t, \theta_0)$ on any exogenous variable or set of exogenous variables we should obtain regression coefficients that are close to zero. Let z_t be a vector of exogenous variables of dimension $l \geq k$, then the coefficients of the regression

$$f_t(y_t, x_t, \theta_0) = z_t'\gamma + \varepsilon_t \tag{2.3}$$

should be close to zero. Hence we can test H_0 by testing

$$H_0' : \gamma = 0. \tag{2.4}$$

z_t can be viewed as a vector of instruments, which may include x_t.

It is important to note that regression (2.3) should be understood as a computational device rather than a statistical model. However, it may also be a simple

transformation of the initial equation of the model. To understand this, consider a linear regression

$$y_t = x_t'\theta + u_t \tag{2.5}$$

and suppose we wish to test $H_0 : \theta = \theta_0$. In this case, we have:

$$f_t(y_t, x_t, \theta) = y_t - x_t'\theta.$$

On subtracting $x_t'\theta_0$ from each side of (2.5), we get:

$$y_t - x_t'\theta_0 = x_t'(\theta - \theta_0) + u_t$$

or

$$f_t(y_t, x_t, \theta_0) = x_t'\gamma + u_t$$

where $\gamma = \theta - \theta_0$. If we choose $z_t = x_t$ (the vector of exogenous variables), (2.3) is equivalent to (2.5). But if we choose a vector of variables $z_t \neq x_t$, (2.3) may be interpreted as an artificial regression [see Mackinnon (1992)].

To test the hypothesis H_0', the F-test is given by

$$T(\theta_0, f) = \frac{f(Y, X, \theta_0)'P(Z)f(Y, X, \theta_0)/l}{f(Y, X, \theta_0)'M(Z)f(Y, X, \theta_0)/(T - l)} \tag{2.6}$$

where $f(Y, X, \theta_0)$ is a $T \times 1$ vector of components $f_t(y_t, x_t, \theta_0)$, $Z = [z_1, z_1, \ldots, z_T]'$, $P(Z) = Z(Z'Z)^{-1}Z'$ is the projection matrix on the space spanned by the columns of Z, and $M(Z) = I - P(Z)$. If the variables z_t are such that

$$z_t \text{ and } u_s \text{ are independent}, \forall s, t, \tag{2.7}$$

and if the matrix Z has full column rank, then under Assumption (2.2) and H_0, $T(\theta_0, f)$ follows a central Fisher distribution with degrees of freedom l and $T - l$. If the hypotheses of normality and independence of the disturbances do not hold, this statistic may not follow an F-distribution. However, if the sample size is sufficiently large, the test $T(\theta_0, f)$ is valid even if the normality and independence hypotheses do

not hold. Its asymptotic distribution under H_0 is given by χ^2_l/l (see Section 5).

This approach is a generalization of the method proposed many years ago by Hartley (1964). This author considered the problem of building confidence sets in nonlinear regression models

$$f_t(y_t, x_t, \theta) = y_t - g(x_t, \theta). \tag{2.8}$$

He constructed a pivotal statistic from an orthogonal decomposition of the sum of squared errors. It is clear there is an infinite number of possible decompositions, and the author's suggestion as will be shown in the next section is not the optimal one.

If we consider the linear regression, we have

$$f_t(y_t, x_t, \theta_0) = y_t - x_t'\theta_0 \ ,$$

and it is easy to see that for $Z = X$, $T(\theta_0, f)$ turns out to be the usual F-statistic:

$$T(\theta_0, f) = \frac{(S_w - S_\Omega)/l}{S_\Omega/(T-l)}$$

where S_w is the sum of squared residuals from the OLS estimation under H_0 and S_Ω is the sum of unconstrained OLS squared residuals.

If we consider a linear simultaneous equations model with $z_t = [x_{2t}', x_{1t}']'$ as a vector of instruments:

$$y_{1t} = y_{2t}'\theta_1 + x_{1t}'\theta_2 + u_t \tag{2.9}$$

$$y_{2t}' = x_{2t}'\pi + x_{1t}'\phi + V_t' \tag{2.10}$$

where $y_t = (y_{1t}, y_{2t}')'$ is the vector of endogenous variables, and $\theta = (\theta_1', \theta_2')'$ is the vector of structural coefficients, the regression equation (2.3) reduces to

$$y_{1t} - (y_{2t}', x_{1t}')\theta_0 = z_t'\gamma + \varepsilon_t \tag{2.11}$$

and $T(\theta_0, f)$ coincides with the test proposed by Anderson and Rubin (1949) for hypotheses specifying all the vector of structural coefficients. Hence, this method also generalizes the Anderson-Rubin test proposed in the context of linear simultaneous equations model to more general possibly nonlinear models.

The approach proposed here can be further generalized if we replace equation (2.1) by a more general one, namely

$$f_t(y_t, x_t, \theta) - g_t(x_t, \theta)'\beta = u_t \tag{2.12}$$

where θ is the parameter of interest, β is another $p \times 1$ vector of unknown parameters and $g_t(x_t, \theta)$ is a $p \times 1$ vector which depends on the exogenous variables x_t and θ. In this case, instead of (2.3), we consider the regression

$$f_t(y_t, x_t, \theta_0) = g_t(x_t, \theta_0)'\beta + z_t'\gamma + \varepsilon_t \tag{2.13}$$

and we test $H_0 : \gamma = 0$. The corresponding F-statistic is given by

$$T(\theta_0, f, g) = \frac{f(Y, X, \theta_0)'\left[M[G(X, \theta_0)] - M[G(X, \theta_0), Z]\right] f(Y, X, \theta_0)/l}{f(Y, X, \theta_0)' M[G(X, \theta_0), Z] f(Y, X, \theta_0)/(T - l - p)}$$

where $G(X, \theta_0) = [g_1(x_1, \theta_0), \ldots, g_T(x_T, \theta_0)]'$. If Assumptions (2.2) and (2.7) hold, $T(\theta_0, f, g)$ follows under H_0, a central F-distribution with degrees of freedom l and $T - l$. This specification covers linear simultaneous equations models and allows tests specifying only the coefficients of the endogenous explanatory variables. For convenience reasons we will focus in what follows, on model (2.1). The extension to (2.12) is straightforward. Below we shall call $T(\theta_0, f)$ and $T(\theta_0, f, g)$ **generalized Hartley-Anderson-Rubin** ($GHAR$) test statistics.

If we are interested in building confidence sets for θ, it is easy to derive from $T(\theta_0, f)$ a confidence set for θ, it is given by

$$C_\theta(\alpha) = \{\theta_0 : \; T(\theta_0, f) \leq F_{l,T-l}(\alpha)\} \tag{2.14}$$

where $F_{l,T-l}(\alpha)$ is the $1-\alpha$ quantile of the central F-distribution with degrees of freedom l and $T-l$. It is important to notice here that this confidence set is exact and does not require an identification assumption.

2.2. Inference on transformations of θ

As the Anderson-Rubin test and many other tests robust to identification proposed recently [Wang and Zivot (1998), Stock and Wright (2000)], a drawback of the $GHAR$ test is that it requires the specification of the whole vector of the unknown parameters. It does not allow one to draw inference on sub-vectors of θ or on general transformations of it. A solution to this problem is the projection technique described in Dufour (1990, 1997). This approach is based on the fact that if

$$P[\theta \in C_\theta(\alpha)] \geqslant 1-\alpha,$$

then,

$$P\big[g(\theta) \in g[C_\theta(\alpha)]\big] \geqslant 1-\alpha$$

and hence $g[C_\theta(\alpha)]$ is a confidence set with level $1-\alpha$ for $g(\theta)$. Despite the simple idea behind this approach, its applications in the context of computable general equilibrium models [Abdelkhalek and Dufour (1998)], or in the context of simultaneous equations models [Dufour and Jasiak (2001), Dufour and Taamouti (2000b)] give good results.

In general, the projections are calculated using numerical methods. However, if the instruments Z do not depend on θ and if the function f_t is linear in θ (nonlinearity in variables only):

$$f_t(y_t, x_t, \theta) = h_{1t}(y_t, x_t)'\theta + h_{2t}(y_t, x_t),$$

the set (2.14) can be written

$$C_\theta(\alpha) = \{\theta_0 : \ \theta_0' A\theta_0 + b'\theta_0 + c \leq 0\} \tag{2.15}$$

where $A = h_1(Y,X)'Bh_1(Y,X)$, $b = -2h_1(Y,X)'Bh_2(Y,X)$, $c = h_2(Y,X)'Bh_2(Y,X)$, $h_1(Y,X)$ is a $T \times k$ matrix with i-th row given by $h_{1t}(y_t, x_t)'$, $h_2(Y,X)$ is the $T \times 1$ vector with the i-th component given by $h_{2t}(y_t, x_t)$, $B = P(Z) - r(\alpha)M(Z)$ and $r(\alpha) = lF_{l,T-l}(\alpha)/(T-l)$. The confidence sets of this form were extensively studied in Dufour and Taamouti (2000b). We showed how to calculate explicitly confidence sets for linear combinations of the components of θ using (2.15). On the other hand, in accordance with Dufour (1997) result's, the confidence set (2.15) may be unbounded with positive probability. Boundedness obtains when the eigenvalues of the stochastic matrix A are positive [Dufour and Taamouti (2000b)].

3. Point-optimal instruments

To evaluate the performance of the test $T(\theta_0, f)$, we should calculate its power, i.e. the probability of rejecting H_0 when it is false. This power depends on the instruments Z used in the regression (2.3), and our main goal is to determine which matrix Z of instruments maximizes the power of the test $T(\theta_0, f)$. Amemiya (1977) was the first to consider the problem of choosing optimal instruments in the context of nonlinear simultaneous equations models. The optimality criterion considered was the minimization of the asymptotic variance of the nonlinear three stage least squares (NL3SLS) estimator [see also Gallant (1987) and Hansen (1982)]. The matrix of optimal instruments he found is the expectation of the derivative of f with respect to the vector of parameters. In general this optimal matrix of instruments cannot be used in practice because it depends on unknown parameters as well as on the conditional distribution of the endogenous variables. Newey (1990, 1993) considered the problem of estimating the optimal instrumental variables and proposed nonparametric methods to achieve this goal.

The optimality results of Amemiya (1977) and subsequent researches [Gallant (1987), Newey (1990, 1993)] are based on the asymptotic distributions of the underlying estimators and are obtained by minimizing the asymptotic variance of these estimators. Here, we wish to maximize the power function of a test. The power of

the test $T(\theta_0, f)$ against an alternative $H_1 : \theta = \theta_1$, is defined by

$$\pi(\theta_1) = P_{\theta_1}[T(\theta_0, f) \geq F_{l,T-l}(\alpha)].$$

Under H_1, $T(\theta_0, f)$ does not follow anymore a central F-distribution. To determine how power depends on the instruments Z, we need to determine the distribution of $T(\theta_0, f)$ under H_1. To derive this distribution we shall consider the following additional assumption:

$$f_t,\ t = 1, \ldots, T \text{ are continously differentiable functions of } \theta. \tag{3.1}$$

Under this assumption, we can use the mean value theorem and get:

$$f_t(y_t, x_t, \theta_0) = f_t(y_t, x_t, \theta_1) + \frac{\partial f_t(y_t, x_t, \bar{\theta}_t)}{\partial \theta'}(\theta_0 - \theta_1) \tag{3.2}$$

where $\bar{\theta}_t$ is such that $\bar{\theta}_t = p_t\theta_0 + (1 - p_t)\theta_1$, and $p_t = p_t(y_t, x_t, \theta_0, \theta_1) \in [0, 1]$. Writing the theorem under this form implicitly assumes that x_t and y_t are parameters of the function $f(\theta)$. $\bar{\theta}_t$ depends on θ_0 and θ_1, but also, via p_t, on x_t and y_t. To determine the distribution of $T(\theta_0, f)$, we distinguish between three main cases:

(1) the derivative of f_t, $t = 1, \ldots, T$, with respect to θ does not depend on the endogenous variables y_t, i.e. it is exogenous;

(2) the derivative of f_t, $t = 1, \ldots, T$, with respect to θ is a linear function of y_t;

(3) the general case where f_t, $t = 1, \ldots, T$, may be any real continuously differentiable function of θ.

3.1. The derivative of f_t is exogenous

In this case, the function $f_t(y_t, x_t, \theta)$ may be written :

$$f_t(y_t, x_t, \theta) = G_{1t}(x_t, \theta) + G_{2t}(y_t, x_t). \tag{3.3}$$

This form covers linear regression models, where

$$f_t(y_t, x_t, \theta) = y_t - x_t\theta \ ,$$

as well as nonlinear regression models where

$$f_t(y_t, x_t, \theta) = y_t - g(x_t, \theta).$$

Moreover, if $G_{1t}(x_t, \theta)$ is linear in θ,

$$G_{1t}(x_t, \theta) = G_{11t}(x_t)'\theta + G_{12t}(x_t),$$

inverting the test $T(\theta_0, f)$ yields a confidence set of the form (2.15). On replacing f_t by its expression (3.3) in equation (3.2), we obtain:

$$G_{1t}(x_t, \theta_0) + G_{2t}(y_t, x_t) = G_{1t}(x_t, \theta_1) + G_{2t}(y_t, x_t) + \frac{\partial G_{1t}(x_t, \bar{\theta}_t)}{\partial \theta'}(\theta_0 - \theta_1)$$

or

$$G_{1t}(x_t, \theta_0) - G_{1t}(x_t, \theta_1) = \frac{\partial G_{1t}(x_t, \bar{\theta}_t)}{\partial \theta'}(\theta_0 - \theta_1). \tag{3.4}$$

It is clear from this equation that $\bar{\theta}_t$ depends only on θ_0, θ_1 and x_t. It does not depend on the endogenous variables, and conditional on X it is fixed. Now we are able to derive the distribution of the statistic $T(\theta_0, f)$ under H_1, it is given by the following proposition. Proofs are presented in the Appendix.

Proposition 3.1 DISTRIBUTION OF GHAR STATISTICS WHEN DERIVATIVES ARE EXOGENOUS. *Under the Assumptions* (2.1) (2.2), (2.7), (3.1), *and* (3.3), *suppose* $\theta = \theta_1$. *Then, conditional on* X, *the statistic* $T(\theta_0, f)$ *defined in* (2.6) *follows a doubly noncentral Fisher distribution with degrees of freedom* l *and* $T - l$ *and noncentrality parameters* λ_1 *and* λ_2 *given by*

$$\lambda_1 = \frac{1}{\sigma^2}(\theta_0 - \theta_1)'H(\bar{\theta})'P(Z)H(\bar{\theta})(\theta_0 - \theta_1),$$

$$\lambda_2 = \frac{1}{\sigma^2}(\theta_0 - \theta_1)' H(\bar{\theta})' M(Z) H(\bar{\theta})(\theta_0 - \theta_1),$$

$$H(\bar{\theta}) = [\frac{\partial f_1(y_1, x_1, \bar{\theta}_1)}{\partial \theta}, \ldots, \frac{\partial f_T(y_T, x_T, \bar{\theta}_T)}{\partial \theta}]',$$

and $\bar{\theta} = [\bar{\theta}_1, \ldots, \bar{\theta}_T]$, $\bar{\theta}_t$ *is an implicit function of* θ_0, θ_1, *and* x_t *given by* (3.4). *We note hereafter this distribution* $F_{l,T-l}(\lambda_1, \lambda_2)$.

Following this proposition, we can write the power function as

$$\pi(\theta_1) = P[F_{l,T-l}(\lambda_1, \lambda_2) \geq c(\alpha)]$$

where $c(\alpha) = F_{l,T-l}(\alpha)$ is a constant equal to the quantile of order $1 - \alpha$ of a central F-distribution with l and $T - l$ as degrees of freedom. If we apply the result of Proposition **3.1** to a linear regression $[f_t(y_t, x_t, \theta) = y_t - x_t'\theta]$, we obtain the power function of the usual F-test ($Z = X$) for an hypothesis $H_0 : \theta = \theta_0$. In this case $\lambda_2 = 0$ and λ_1 reduces to

$$\lambda_1 = \frac{1}{\sigma^2}(\theta_0 - \theta_1)' X' X (\theta_0 - \theta_1).$$

If $\theta_1 = \theta_0$, the noncentrality parameters reduce to 0 and the power function of the test equals the test level α whatever the instruments Z used in the artificial regression.

When $\theta_1 \neq \theta_0$, we wish to select the instruments Z such that this probability is as large as possible. This involves selecting the number of instruments and their values. We first consider the question of the value of Z assuming a fixed number of columns, $l \geq k$. So, we will focus on matrices of instruments of dimension $T \times l$. We solve the problem by studying the variation of the power function $\pi(\theta_1)$ with respect to the noncentrality parameters λ_1 and λ_2 [Johnson, Kotz, and Balakrishnan (1994, page 443)]. This is done in the following lemma which, to best of our knowledge, has not been stated elsewhere in the literature.

Lemma 3.2 MONOTONOCITY OF THE NONCENTRAL F DISTRIBUTION WITH RESPECT TO NONCENTRALITY PARAMETERS. *If* $F_{k_1,k_2}(\lambda_1, \lambda_2)$ *has a doubly noncentral Fisher distribution with degrees of freedom* k_1 *and* k_2 *and noncentrality parameters* λ_1

and λ_2, *then its distribution function* $F(y|\ k_1, k_2, \lambda_1, \lambda_2)$ *is an increasing function of* λ_2 *and a decreasing function of* λ_1.

In our context, we have:

$$k_1 = l,\ k_2 = T - l,$$

$$\lambda_1 = \frac{1}{\sigma^2}(\theta_0 - \theta_1)'H(\bar{\theta})'P(Z)H(\bar{\theta})(\theta_0 - \theta_1),$$

$$\lambda_2 = \frac{1}{\sigma^2}(\theta_0 - \theta_1)'H(\bar{\theta})'H(\bar{\theta})(\theta_0 - \theta_1) - \lambda_1.$$

So, maximizing the power $\pi(\theta_1)$ is equivalent to maximizing λ_1. From this we can characterize the optimal instrument matrix for testing $\theta = \theta_0$ against $\theta = \theta_1$. This is given in the following proposition.

Proposition 3.3 OPTIMAL INSTRUMENTS WHEN DERIVATIVES ARE EXOGENOUS. *Under the assumptions of Proposition* **3.1**, *a necessary and sufficient condition for a* $T \times l$ *matrix* Z $(l \geq k)$ *to maximize the power of* $T(\theta_0, f)$ *among all full rank* $T \times l$ *matrices is that the space spanned by the columns of* Z *include the space spanned by the columns of* $H(\bar{\theta})$.

Recall that

$$H(\bar{\theta}) = \left[\frac{\partial f_1(y_1, x_1, \bar{\theta}_1)}{\partial \theta}, \ldots, \frac{\partial f_T(y_T, x_T, \bar{\theta}_T)}{\partial \theta}\right]'.$$

Proposition **3.3** means that the projection of each column of $H(\bar{\theta})$ on the space spanned by the columns of Z yields the same column:

$$P(Z)H(\bar{\theta}) = H(\bar{\theta}).$$

In this case, the matrix Z may be written as

$$Z = [H(\bar{\theta}), C]D \tag{3.5}$$

where D is a nonsingular $l \times l$ matrix, and C is a matrix of dimension $T \times (l-k)$. This means that k columns of Z are identified up to a bijective transformation. When Z is given as in (3.5), the second noncentrality parameter λ_2 vanishes. The power function can be written:

$$\pi(\theta_1) = P[F_{l,T-l}(\lambda_1, 0) \geq c]$$

where $\lambda_1 = (\theta_0 - \theta_1)'H(\bar{\theta})'H(\bar{\theta})(\theta_0 - \theta_1)/\sigma^2$ does not depend on l.

The second step consists in choosing the optimal number of instruments. This is equivalent to maximizing $\pi(\theta_1)$ with respect to l. Intuitively, to maximize power, one should not include irrelevant instruments. This assertion is largely documented in the literature on asymptotic theory [see, for example, Buse (1992) and Chao and Swanson (2000)]. In particular, overidentification increases bias of IV estimators in finite samples. However, proving this analytically in finite samples, appears to be difficult (if not impossible), but we can obtain evidence on this issue through Monte Carlo methods or directly from the noncentral F-distribution. Table 1 and Figure 1 show the variation of $\pi(\theta_1)$ as the number l increases for different values of λ_1. The results support the conjecture. Consequently, if we focus only on matrices of dimension $T \times k$, the optimal matrix of instruments is given in the following corollary.

Corollary 3.4 OPTIMAL INSTRUMENT MATRIX OF ORDER K WHEN DERIVATIVES ARE EXOGENOUS. *Under the assumptions of Proposition* **3.3**, *the $T \times k$ matrix of instruments that maximizes the power of the test $T(\theta_0, f)$ is given by*

$$Z = H(\bar{\theta})P$$

where P is any $k \times k$ nonsingular matrix.

To further justify restricting ourselves to matrices of dimension $T \times k$, consider again the artificial regression (2.3). Replacing Z by $[H(\bar{\theta}), C]D$, we can write :

$$f_t(y_t, x_t, \theta_0) \;=\; Z_t\gamma + \varepsilon_t = [H_t', C_t']D\gamma + \varepsilon_t$$

$$= H_t'\gamma_{1*} + C_t'\gamma_{2*} + \varepsilon_t \tag{3.6}$$

where H_t and C_t are the t-th rows of $H(\bar{\theta})$ and C respectively, and $[\gamma_{1*}', \gamma_{2*}']' = D\gamma$. The hypothesis H_0 then takes the form $H_0 : (\gamma_{1*}', \gamma_{2*}')' = 0$. If the true value of the parameter is θ, introducing the expectation operator on both sides of equation (3.2) with θ_1 replaced by θ, we get:

$$E_\theta[f_t(y_t, x_t, \theta_0)] = H_t'(\theta_0 - \theta) = H_t'\gamma_{1*}$$

where $\gamma_{1*} = \theta_0 - \theta$. Thus we can write:

$$f_t(y_t, x_t, \theta_0) = H_t\gamma_{1*} + \varepsilon_t. \tag{3.7}$$

In summary,

$$f_t(y_t, x_t, \theta_0) = \begin{cases} u_t\,, & \text{if } H_0 \text{ is true,} \\ H_t\gamma_{1*} + \varepsilon_t\,, & \text{if } H_0 \text{ is not true} \end{cases}$$

Hence, in the regression (3.6), $\gamma_{2*} = 0$ irrespective wether H_0 holds or not. Testing $H_0 : (\gamma_{1*}', \gamma_{2*}')' = 0$ is equivalent to testing $H_0 : \gamma_{1*} = 0$, and to test H_0, it is sufficient to regress $f_t(y_t, x_t, \theta_0)$ on $Z = H(\bar{\theta})$. Regressing on other non relevant variables does not provide more information and hence should not improve the quality of the test.

Using optimal instruments, the regression equation (2.3) can be written:

$$f_t(y_t, x_t, \theta_0) = \frac{\partial f_t(y_t, x_t, \bar{\theta}_t)}{\partial \theta'}\gamma + \varepsilon_t. \tag{3.8}$$

This equation is similar to a Gauss-Newton type artificial regressions (GNR) as met in nonlinear regressions [see Davidson and MacKinnon (1993)]. However the two regressions are different for several reasons. Regression (3.8) is based on an optimality criterion and is valid even for small samples. GNR regressions are based on local linear asymptotic approximations of nonlinear estimators. Moreover the regressors in (3.8) are exogenous while in the artificial regressions like those considered by Davidson and MacKinnon (1993), θ is replaced by a consistent estimator and thus depends on

the endogenous variables. This may lead to Wald-type statistics [see Davidson and MacKinnon (1993, page 230)].

Proposition **3.3** also shows that the instruments proposed by Hartley (1964) in the context of nonlinear regressions are not optimal. This author proposed as instruments exhaustive statistics obtained by decomposing in an exact or approximate way $f_t(x_t, \theta)$ as

$$f_t(x_t, \theta) = \sum_{i=1}^{m} w_i(\theta) h_i(x_t).$$

The matrix of optimal instruments depends on $\bar{\theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_T)$. Each $\bar{\theta}_t$ may be calculated by solving the equation

$$G_{1t}(x_t, \theta_0) - G_{1t}(x_t, \theta_1) = \frac{\partial G_{1t}(x_t, \bar{\theta}_t)}{\partial \theta'}(\theta_0 - \theta_1). \tag{3.9}$$

In many cases, especially if θ is scalar, it is possible to determine $\bar{\theta}_t$ explicitly, but in general, this equation is nonlinear and its solution may require one to use numerical methods to get the root of an equation. If θ is not scalar, equation (3.9) may have several roots, but will result in the same matrix of optimal instruments. However, it is worth noting that what one needs is not $\bar{\theta}$ but the matrix $H(\bar{\theta}) = \partial f(X, \bar{\theta})/\partial \theta'$. When θ is scalar, this matrix can be obtained without calculating $\bar{\theta}$.[1] From

$$f(Y, X, \theta_0) = f(Y, X, \theta_1) + H(\bar{\theta})(\theta_0 - \theta_1),$$

we deduce the optimal instrument:

$$Z^* = H(\bar{\theta}) = \frac{f(Y, X, \theta_0) - f(Y, X, \theta_1)}{\theta_0 - \theta_1}. \tag{3.10}$$

When θ is a vector of dimension $k > 1$, we can fix its first $k - 1$ components at the mean of θ_0 and θ_1 $[\bar{\theta}^i = (\theta_0^i + \theta_1^i)/2,\ i = 1, \ldots, k-1]$ and solve the equation for the last component.

[1] Even when θ is not scalar, $H(\bar{\theta})$ may be obtained in a simple way (see the example of Subsection 3.2).

The optimality of the instruments $Z^* = H(\bar{\theta})$ is established when we test $H_0 : \theta = \theta_0$ against a simple alternative $H_1 : \theta = \theta_1$, so it is natural that the optimal instruments depend on the alternative. The instruments Z^* are optimal among the class given by $T(\theta_0, f)$ only at the point $\theta = \theta_1$ of the alternative. The test with $Z = Z^*$ is a **point-optimal test** [King (1988), Dufour and King (1991)]. The basic idea behind such tests is to maximize power against a given alternative, which is selected either because it is particularly plausible or important or because it provides good global properties. Further, point-optimal tests may be used to calculate the **power envelope** [see King (1988), Dufour and King (1991), Elliott, Rothenberg, and Stock (1996), and Perron and Rodriguez (1998)] of the class of tests considered. For each point in the alternative, we calculate the maximum power that can be achieved, providing a benchmark for evaluating any other test in this class.

In general, the alternative hypothesis is not simple, being of the form $H_1 : \theta \neq \theta_0$. In this case, a natural choice consists in evaluating $\partial f(X, \theta)/\partial\theta'$ at θ_0. Obviously this choice will induce power loss against each alternative in H_1. The loss will be small if the value of the alternative is near θ_0. However the power loss will be significant if $\partial f(X, \theta)/\partial\theta'$ has a high slope (in absolute value) around θ_0 such that it is near 0 at θ_0 and is largely different from 0 around θ_0. In this case the instrument Z evaluated at θ_0 induces weak power given that it is close to 0.

When θ is scalar, one can also build two-sided tests by choosing θ_1 and θ_2 such that $\theta_1 < \theta_0 < \theta_2$ and rejecting H_0 against $H_1 : \theta \neq \theta_0$ if H_0 is rejected against $H_1' : \theta = \theta_1$ or against $H_1'' : \theta = \theta_2$. If α_1 and α_2 are the respective sizes of the later tests with $\alpha_1 + \alpha_2 = \alpha$, using the Bonferroni inequality, we have:

$$\begin{aligned} P_{\theta_0}[\text{reject } H_0 \text{ against } H_1] &= P_{\theta_0}[(\text{reject } H_0 \text{ against } H_1') \text{ or } (\text{reject } H_0 \text{ against } H_1'')] \\ &\leq P_{\theta_0}[\text{reject } H_0 \text{ against } H_1'] + P_{\theta_0}[\text{reject } H_0 \text{ against } H_1''] \\ &= \alpha_1 + \alpha_2 = \alpha \end{aligned}$$

Hence the test has a size smaller than or equal α. This approach can be generalized easily to the case where θ is a vector. For further discussion, see Dufour and King (1991).

3.2. The derivative of f_t is a linear function of the endogenous variables

In this section, we consider the case where the derivative of f_t with respect to θ is given by

$$\frac{\partial f_t(y_t, x_t, \theta)}{\partial \theta} = Ay_t + g_{1t}(x_t, \theta)$$

where A is a $k \times n$ fixed matrix independent of θ (n is the number of endogenous variables) and g_{1t}, $t = 1, \ldots, T$ are functions in $\mathbb{R}^k$ and do not depend on the endogenous variables. The function f_t can then be written

$$f_t(y_t, x_t, \theta) = y_t' A'\theta + G_{1t}(x_t, \theta) + G_{2t}(y_t, x_t). \tag{3.11}$$

This form of the model is nonlinear in parameters as well as in variables. It covers as a special case linear simultaneous equations models (2.9)-(2.10):

$$f_t(y_t, x_t, \theta) = y_{1t} - y_{2t}'\theta_1 - x_{1t}'\theta_2$$

with,

$$y_t = (y_{1t}, y_{2t}')', \; A = \begin{pmatrix} 0 & -I_{n-1} \\ 0 & 0 \end{pmatrix}, \; G_{1t}(x_t, \theta) = -x_{1t}'\theta_2, \; G_{2t}(y_t, x_t) = y_{1t},$$

and $\theta = (\theta_1', \theta_2')'$. This form also generalizes (3.3), which corresponds to $A = \mathbf{0}$. Further, when $G_{1t}(x_t, \theta)$ is linear in θ, i.e.

$$G_{1t}(x_t, \theta) = G_{11t}'(x_t)\theta + G_{12t}(x_t),$$

the test $T(\theta_0, f)$ induces a confidence set of the form (2.15).

Applying the mean-value theorem, (3.11) yields:

$$f_t(y_t, x_t, \theta_0) = f_t(y_t, x_t, \theta_1) + [Ay_t + g_{1t}(x_t, \bar{\theta}_t)]'(\theta_0 - \theta_1). \tag{3.12}$$

The functions g_{1t} $t = 1, \ldots, T$ are evaluated at $\bar{\theta}_t$ such that $\bar{\theta}_t = p_t\theta_0 + (1 - p_t)\theta_1$, $p_t \in [0, 1]$. To derive the conditional distribution of $T(\theta_0, f)$, we assume the following hypothesis [instead of (2.2)]:

$$(u_t, y_t')' \overset{i.i.d.}{\sim} N([0, E(y_t)']', \Sigma), \; t = 1, \ldots, T \tag{3.13}$$

where $\Sigma = \begin{pmatrix} \sigma^2 & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}$. In the case of linear simultaneous equations models (2.9)-(2.10), this hypothesis is equivalent to the usual assumption that the disturbances $(u_t, V_t')'$ are normal i.i.d. It is also important in this case to notice that $\bar{\theta}_t$ does not depend on the endogenous variables. On replacing f_t by its expression in equation (3.12), we obtain:

$$\begin{aligned} y_t'A'\theta_0 + G_{1t}(x_t, \theta_0) + G_{2t}(y_t, x_t) &= y_t'A'\theta_1 + G_{1t}(x_t, \theta_1) + G_{2t}(y_t, x_t) \\ &\quad + [Ay_t + g_{1t}(x_t, \bar{\theta}_t)]'(\theta_0 - \theta_1). \end{aligned}$$

Hence

$$G_{1t}(x_t, \theta_0) - G_{1t}(x_t, \theta_1) = g_{1t}(x_t, \bar{\theta}_t)'(\theta_0 - \theta_1). \tag{3.14}$$

It is clear that $\bar{\theta}_t$ does not depend on the endogenous variables.

Under the assumptions considered above, we can derive the conditional distribution of $T(\theta_0, f)$ under the alternative $H_1 : \theta = \theta_1$. It is given by the following proposition.

Proposition 3.5 DISTRIBUTION OF GHAR STATISTICS WHEN DERIVATIVES ARE LINEAR IN THE ENDOGEOUS VARIABLES. *Under the Assumptions* (2.1), (2.7), (3.1), (3.11), *and* (3.13), *suppose* $\theta = \theta_1$. *Then, conditional on* X, *the statistic* $T(\theta_0, f)$ *defined in* (2.6) *follows a doubly noncentral Fisher distribution with degrees of freedom*

$(l, T-l)$ *and noncentrality parameters* (λ_1, λ_2), *where*

$$\lambda_1 = \frac{1}{\Delta}(\theta_0 - \theta_1)'E[H(\bar{\theta})]'P(Z)E[H(\bar{\theta})](\theta_0 - \theta_1),$$

$$\lambda_2 = \frac{1}{\Delta}(\theta_0 - \theta_1)'E[H(\bar{\theta})]'M(Z)E[H(\bar{\theta})](\theta_0 - \theta_1),$$

$$H(\bar{\theta}) = [\frac{\partial f(y_t, x_t, \bar{\theta}_t)}{\partial \theta^i}]_{t=1,\ldots,T,\ i=1,\ldots,k},\ \Delta = C'\Sigma C,\ C = [1, (\theta_0 - \theta_1)'A]',$$

and $\bar{\theta}_t$ *solves the equation* (3.14).

For the linear simultaneous equations models (2.9)-(2.10), we have

$$\frac{\partial f_t}{\partial \theta}(y_t, x_t, \theta) = -[y_{2t}', x_{1t}']'$$

and

$$E[H(\bar{\theta})] = E[\frac{\partial f(Y, X, \bar{\theta})}{\partial \theta'}] = -Z\Pi$$

where $Z = [z_1, z_2, \ldots, z_T]' = [X_2, X_1]$, $\Pi = \begin{pmatrix} \pi & 0 \\ \phi & I_{p_1} \end{pmatrix}$, and p_1 is the number of columns of x_{1t}. The second noncentrality parameter reduces to 0 ($\lambda_2 = 0$). We obtain here as a special case the result of Revankar and Mallela (1972) who derived the power function of the Anderson and Rubin (1949) test. Consequently Proposition **3.5** is a generalization of this result.

As in the previous section, we can derive the $T \times k$ matrix of optimal instruments that maximizes the power of the test $T(\theta_0, f)$. This matrix is given by the following proposition.

Proposition 3.6 OPTIMAL INSTRUMENT MATRIX OF ORDER K WHEN DERIVATIVES ARE LINEAR IN THE ENDOGENOUS VARIABLES. *Under the assumptions of Proposition* **3.5**, *among the matrices* Z *of dimension* $T \times k$, *those that maximize the power of the test* $T(\theta_0, f)$ *have the form*

$$Z^* = E[H(\bar{\theta})]P$$

where P is any $k \times k$ nonsingular matrix.

The derivative of f_t is evaluated at x_t, $\bar{\theta}_t$ and $E(y_t)$:

$$E\left[\frac{\partial f_t(y_t, x_t, \bar{\theta})}{\partial \theta'}\right] = \frac{\partial f_t(E(y_t), x_t, \bar{\theta})}{\partial \theta'} .$$

The matrix of optimal instruments does not depend on the endogenous variables. $E(y_t)$ is constant but is unknown in general. When the value of the alternative θ_1 is not far from θ_0, Z^* is near $E[\partial f(Y, X, \theta_0)/\partial \theta']$, the optimal instrument well known in the literature [Amemiya (1977)] and which is obtained by minimizing the asymptotic variance of NL3SLS estimator.

In the same way as before, when θ is scalar, it is easy to derive the explicit expression of the matrix of the optimal instruments without calculating the value of $\bar{\theta}_t$. This yields

$$Z^* = \frac{f(E(Y), X, \theta_0) - f(E(Y), X, \theta_1)}{\theta_0 - \theta_1} .$$

For linear simultaneous equations models (2.9)-(2.10), the matrix Z^* is given by

$$Z^* = Z\Pi,$$

which is unknown. This proves that the Anderson-Rubin test is not optimal in the class given by $T(\theta_0, f)$ (except when Π is square and nonsingular, which corresponds to the exact identification case), but constitutes an alternative solution since Z^* is unknown. It is worth noting that in this case, the optimal instruments do not depend on the alternative and hence the test $T(\theta_0, f)$ using Z^* is **uniformly most powerful** test among the class given by $T(\theta_0, f)$ with matrices of instruments of dimension $T \times k$.

Another alternative solution consists in estimating the matrix Π from the reduced form in (2.9)-(2.10) ($\hat{\Pi} = (Z'Z)^{-1}Z'[Y_2, X_1]$) and replacing the optimal instrument by $Z^* = P(Z)[Y_2, X_1]$. The regression equation (2.3) becomes:

$$y - Y_2\theta_1 - X_1\theta_2 = P(Z)[Y_2, X_1]\gamma + \varepsilon.$$

This regression presents two drawbacks. First, the test $T(\theta_0, f)$ is valid under the assumption that the matrix Z is exogenous or fixed which is not the case of $Z^* = P(Z)[Y_2, X_1]$, which is correlated with u_t. Second, even if we consider the test on the basis of its asymptotic validity, it will result (as noticed in the previous section) in a Wald-type test with all its drawbacks in the case of underidentification.

For practical implementation of the optimal instrument, we can use the split-sample approach [Dufour and Jasiak (1993, 2001), Angrist and Krueger (1995)]. This approach consists in dividing the sample into two parts: one is used to estimate the matrix Π and the other one to calculate the test after replacing $Z^* = Z\Pi$ by $\hat{Z}^* = Z\hat{\Pi}$ where $\hat{\Pi}$ is an estimator obtained from the first part of the sample. If the sample is *i.i.d.*, $\hat{Z}^*$ will be independent of the disturbances of the second part of the sample. This technique will be considered further in the next section where we study the general case in which f may be any continuously differentiable function.

The results of this section may be applied to a larger class than (3.11). If we replace (3.11) by

$$f_t(y_t, x_t, \theta) = S_t(y_t, x_t)'\theta + G_{1t}(x_t, \theta) + G_{2t}(y_t, x_t) \tag{3.15}$$

where $S_t(y_t, x_t)$ is any function in $\mathbb{R}^k$. The only condition for our results to hold is to change assumption (3.13) by the assumption

$$[u_t, S(y_t, x_t)']' \overset{i.i.d.}{\sim} N\big([0, E(S(y_t, x_t))']', \Sigma\big) \tag{3.16}$$

On taking $S(y_t, x_t) = Ay_t$, we obtain (3.11).

3.3. General case

In this section, we allow the functions f_t to be any continuously differentiable functions of θ, and we consider again the problem of choosing the optimal instruments for the GHAR test. In the previous sections, we showed that the expectation of the derivative of f_t with respect to θ constitutes an optimal instrument. Even if this was established

for special forms of f_t, we can guess that, even when f_t does not have these special forms, an optimal instrument may depend on $E[\partial f/\partial\theta]$.

It is easy to see that the exact results established before will still be valid for any function f_t under some assumptions. By repeating the same argument of the mean-value theorem for any function f_t that is continuously differentiable in θ, we get:

$$f_t(y_t, x_t, \theta_0) = f_t(y_t, x_t, \theta_1) + \frac{\partial f_t(y_t, x_t, \bar{\theta}_t)}{\partial \theta'}(\theta_0 - \theta_1).$$

If we assume that (2.1) holds with $\theta = \theta_1$, then

$$\begin{aligned} f_t(y_t, x_t, \theta_0) &= u_t + \left(H_t(\bar{\theta}) - E[H_t(\bar{\theta})]\right)'(\theta_0 - \theta_1) + E[H_t(\bar{\theta})]'(\theta_0 - \theta_1) \\ &= [1, (\theta_0 - \theta_1)'] \begin{pmatrix} u_t \\ \varepsilon_t \end{pmatrix} + (\theta_0 - \theta_1)' E[H_t(\bar{\theta})] \end{aligned}$$

where

$$H_t(\bar{\theta}) = \frac{\partial f_t(y_t, x_t, \bar{\theta}_t)}{\partial \theta},$$

$$\varepsilon_t = H_t(\bar{\theta}) - E[H_t(\bar{\theta})].$$

Under the assumption

$$(u_t, \varepsilon_t')' \overset{i.i.d.}{\sim} N[0, \Omega], \quad t = 1, \ldots T, \tag{3.17}$$

conditional on X, $f(Y, X, \theta_0)$ follows a normal distribution

$$f(Y, X, \theta_0) \sim N\left[E[H(\bar{\theta})](\theta_0 - \theta_1), \Delta I_T\right]$$

where $\Delta = [1, (\theta_0 - \theta_1)']\Omega[1, (\theta_0 - \theta_1)']'$ and as before, under $H_1 : \theta = \theta_1$, $T(\theta_0, f)$ follows a noncentral Fisher distribution and the optimal $T \times k$ matrix of instruments is given by $Z^* = E[H(\bar{\theta})]P$, where P is any nonsingular square matrix.

The basic idea underlying the rest of this section is to use as an optimal matrix of instruments an estimator of $H(\bar{\theta}) = E[\partial f(Y, X, \bar{\theta})/\partial\theta']$. This estimator should be exogenous to ensure the validity of exact inference. We use the split-sample technique

to achieve our objective. This technique was used in Dufour and Jasiak (1993, 2001) to improve the power of the Anderson-Rubin test in linear simultaneous equations models and in Angrist and Krueger (1995) to improve finite-sample characteristics of 2SLS estimation [for related methods see also Dufour and Torrès (1998, 2000)].

Assume we have a set Z of l instruments, where Z may include the variables X or may be a part of it. To obtain an estimation of the matrix $Z^* = E[H(\bar{\theta})] = E[\partial f(Y, X, \bar{\theta})/\partial\theta]$, we regress $H(\bar{\theta})$ on Z:[2]

$$H(\bar{\theta}) = Z\delta + \varepsilon, \tag{3.18}$$

which yields $\hat{Z}^* = Z\hat{\delta} = Z(Z'Z)^{-1}Z'H(\bar{\theta}) = P(Z)H(\bar{\theta})$. To test $H_0 : \theta = \theta_0$, we consider the regression

$$f_t(y_t, x_t, \theta_0) = \hat{z}_t^{*\prime}\gamma + u_t, \tag{3.19}$$

and we test $H_0' : \gamma = 0$. Since $\hat{Z}^* = P(Z)H(\bar{\theta})$ depends on the endogenous variables and is generally correlated with the left hand side of (3.19), the statistic $T(\theta_0, f)$ with $\hat{Z}^*$ instead of Z no longer follows an F-distribution. To circumvent this problem, we divide the sample into two parts $(Y_{(1)}, X_{(1)}, Z_{(1)})$ and $(Y_{(2)}, X_{(2)}, Z_{(2)})$ with respective sizes T_1 and T_2 $(T_1 + T_2 = T)$. We use the first part of the sample in the regression (3.18) to obtain an estimator of δ :

$$\hat{\delta} = (Z_{(1)}'Z_{(1)})^{-1}Z_{(1)}'H_{(1)}(\bar{\theta}_{(1)})$$

where $H_{(1)}(\bar{\theta}_{(1)}) = \frac{\partial f_{(1)}}{\partial\theta'}(Y_{(1)}, X_{(1)}, \bar{\theta}_{(1)})$, and $f_{(1)} = (f_1, \ldots, f_{T_1})'$. Then we use the second part in the regression (3.19) after replacing $\hat{Z}^*$ by

$$\hat{Z}^*_{(2)} = Z_{(2)}\hat{\delta}\ ,$$

which is an estimator of $E[\frac{\partial f_{(2)}}{\partial\theta'}(Y_{(2)}, X_{(2)}, \bar{\theta}_{(2)})]$. If the observations are i.i.d., this estimator is independent of the endogenous variables of the second part of the sample.

[2] $H(\bar{\theta})$ is known since it depends on x_t, y_t, θ_0 and θ_1.

The regression

$$f_t(y_t, x_t, \theta_0) = \hat{z}^{*\prime}_{(2)t}\gamma + u_t, \ t = T_1 + 1, \ldots, T$$

yields an exact F-test based on the statistic:

$$T_{SS}(\theta_0, f) = \frac{f(Y_{(2)}, X_{(2)}, \theta_0)' P[\hat{Z}^*_{(2)}] f(Y_{(2)}, X_{(2)}, \theta_0)/l}{f(Y_{(2)}, X_{(2)}, \theta_0)' M[\hat{Z}^*_{(2)}] f(Y_{(2)}, X_{(2)}, \theta_0)/(T_2 - l)}. \tag{3.20}$$

In Angrist and Krueger (1995), the sample was divided into two parts of equal numbers of observations ($T_{(1)} = T_{(2)} = \frac{T}{2}$). This is possibly not the optimal choice. Dufour and Jasiak (1993, 2001) used Monte Carlo simulations considering different partitions of the sample and they found that it is preferable (at least in their context) to use a small part of the sample for the first regression and allocate the big part of the sample to the second regression.

4. Monte Carlo study

To evaluate the performance of the test proposed above, we will now study through Monte Carlo methods two examples corresponding to the two first cases considered above.

The first example considers the following nonlinear regression:

$$y_t = \theta^2 x_t^\theta + u_t \tag{4.1}$$

where θ is an unknown parameter, the disturbances are i.i.d. $N(0, \sigma^2)$, where $\sigma^2 = 1$, and x_t is generated from a χ^2 distribution. The function f_t is thus given by

$$f_t(y_t, x_t, \theta) = y_t - \theta^2 x_t^\theta = u_t$$

and its derivative by

$$\frac{\partial f_t(x_t, \theta)}{\partial \theta'} = -2\theta x_t^\theta - \theta^2 x_t^\theta \ln(x_t).$$

Suppose we wish to test $H_0 : \theta = \theta_0$. If we consider as instrument $Z = \partial f(X, \theta_0)/\partial \theta'$,

the test $T(\theta_0, f)$ will have weak power when θ_0 is near 0 given that Z approaches 0. For each alternative $H_1 : \theta = \theta_1$, the optimal instrument is

$$z_t^* = \frac{\theta_1^2 x_t^{\theta_1} - \theta_0^2 x_t^{\theta_0}}{\theta_0 - \theta_1}, \; t = 1, \ldots, T.$$

Even if θ_0 is 0, Z^* will be different from 0. Table 2 and Figure 2 give the results of a Monte Carlo simulation of the power of the test $T(\theta_0, f)$ using the instrument

$$Z = \frac{\partial f(X, \theta_0)}{\partial \theta'} = [-2\theta_0 x_t^{\theta_0} - \theta_0^2 x_t^{\theta_0} ln(x_t)]_{t=1,\ldots,T},$$

as in Gauss Newton Regressions (GNR) [see Davidson and MacKinnon (1993, chapter 6)] and Z^* (third column of Table 2).[3] We test the hypothesis $H_0 : \theta = \theta_0 = 0.001$. The results show that the difference in power is substantial when we use point-optimal instruments. We should however recall that what we present is not a power function of one test but at each point, we give the power of the optimal test at this point. This is not the case of the GNR test. On the other hand, since the procedures are exact the size of both tests is well controlled.

The second example considers the following equation:

$$f_t(y_t, x_t, \theta) = \lambda y_t - \delta^2 x_{1t}^{\delta} - x_{2t} = u_t \tag{4.2}$$

where $\theta = (\lambda, \delta)'$ is the vector of unknown parameters, x_1 and x_2 are exogenous variables generated from independent χ^2 distributions, and the disturbances u_t are i.i.d. $N(0, 1)$ disturbances. We want to test $H_0 : \theta = \theta_0 = (\lambda_0, \delta_0)' = (1; 0.001)'$ against the alternative $H_1 : \theta = \theta_1 = (\lambda_1, \delta_1)' = (1, \delta_1)$. The extension of the Gauss-Newton (GNR) regressions to this equation [see Davidson and MacKinnon (1998, equation (35))] gives:

$$f_t(y_t, x_t, \theta_0) = z_t'\gamma + \varepsilon_t$$

[3] In both examples, the size of the sample is $T = 50$ and the number of replications is $n = 10.000$.

where $Z = P(X)\frac{\partial f}{\partial \theta}(Y, X, \theta_0)$, $P(X) = X(X'X)^{-1}X'$, $X = [x_1, x_2]$ and:

$$\frac{\partial f_t}{\partial \theta'}(y_t, x_t, \theta) = [y_t, -2\delta x_{1t}^{\delta} - \delta^2 x_{1t}^{\delta} \ln(x_{1t})].$$

On the other hand, the matrix of optimal instruments (with $P = I$) is given by

$$z_t^* = \left[\frac{\delta^2 x_{1t}^{\delta} + x_{2t}}{\lambda}, \frac{\delta_1^2 x_{1t}^{\delta_1} - \delta_0^2 x_{1t}^{\delta_0}}{\delta_0 - \delta_1}\right], \ t = 1, \ldots, T.$$

The first column is $E(y)$, which is unknown since it depends on the true value of θ. Under H_0, $E(y) = (\delta_0^2 x_{1t}^{\delta_0} + x_{2t})/\lambda_0$ and under H_1, $E(y) = (\delta_1^2 x_{1t}^{\delta 1} + x_{2t})/\lambda_1$.

The results of a Monte Carlo simulation of the power of $T(\theta_0, f)$ and its split-sample version are given in Table 3 (see also Figure 3). This table gives the power of the test $T_{SS}(\theta_0)$ for several partitions of the sample [s is the part of the sample used in the regression (3.18)]. The table shows clearly that we should use a small part of the sample in the first regression. In this case, the split-sample (with $s = 0.2$) performs better than the GNR test.

Finally, another approach for estimating the optimal instrument would be the Jackknife technique [see Angrist, Imbens, and Krueger (1999)]. This technique avoids the loss of observations in the first regression but induces a test that is valid only asymptotically.

5. Asymptotic validity

In Section 3, we derived the matrix of optimal instruments under the assumption that the disturbances in equation (2.1) are i.i.d. $N(0, \sigma^2)$. In this section we relax this assumption and consider the case where the disturbances verify a weaker assumption. We will show that the test $T(\theta_0, f)$ remains valid asymptotically under standard assumptions. Under the null, $T(\theta_0, f)$ follows asymptotically a χ_l^2 distribution divided by its degree of freedom, l. Under the non local alternative $H_1 : \theta = \theta_1$, $T(\theta_0, f)$ gets large as T increases and the test is consistent. Under a local alternative, $H_1 : \theta = \theta_0 + C/\sqrt{T}$, the distribution tends to a noncentral χ_l^2 divided by its degree of freedom,

l.

5.1. Asymptotic null distribution

Suppose Z is the instrument used in the artificial regression (2.3), and that the following assumptions hold

$$(\frac{Z'Z}{T}, \frac{u'u}{T}) \underset{T\longrightarrow\infty}{\overset{p}{\longrightarrow}} (Q_{ZZ}, \sigma^2) \tag{5.1}$$

where Q_{ZZ} is a full rank matrix, and

$$\frac{u'Z}{\sqrt{T}} \underset{T\longrightarrow\infty}{\overset{d}{\longrightarrow}} N(0, \sigma^2 Q_{ZZ}). \tag{5.2}$$

Under $H_0 : \theta = \theta_0$,

$$T(\theta_0, f) = \frac{u'P(Z)u/l}{[u'u - u'P(Z)u]/(T-l)}$$

and

$$\begin{aligned} u'P(Z)u &= u'Z(Z'Z)^{-1}Z'u \\ &= \frac{u'Z}{\sqrt{T}}(\frac{Z'Z}{T})^{-1}\frac{Z'u}{\sqrt{T}} \underset{T\longrightarrow\infty}{\overset{d}{\longrightarrow}} S'Q_{ZZ}^{-1}S = \sigma^2\chi_l^2 \end{aligned}$$

where $S = N(0, \sigma^2 Q_{ZZ})$. Further,

$$\frac{u'u}{T-l} = \frac{T}{T-l}\frac{u'u}{T} \underset{T\longrightarrow\infty}{\overset{p}{\longrightarrow}} \sigma^2,$$

$$\frac{u'P(Z)u}{T-l} \underset{T\longrightarrow\infty}{\overset{d}{\longrightarrow}} \frac{\sigma^2\chi_l^2}{T-l} \underset{T\longrightarrow\infty}{\overset{d}{\longrightarrow}} 0.$$

Hence, under H_0

$$T(\theta_0, f) \underset{T\longrightarrow\infty}{\overset{d}{\longrightarrow}} \frac{\chi_l^2}{l}. \tag{5.3}$$

5.2. Asymptotic distribution under a fixed alternative

Under the alternative hypothesis $H_1 : \theta = \theta_1$, we can write

$$f(Y, X, \theta_0) = f(Y, X, \theta_1) + H(\theta_0 - \theta_1)$$

where

$$H = H(\bar{\theta}) = \frac{\partial f}{\partial \theta'}(Y, X, \bar{\theta}) = [\partial f / \partial \theta_i(y_t, x_t, \bar{\theta}_t)]_{i=1,\ldots,k,\ t=1,\ldots,T}$$

is the $T \times k$ matrix of partial derivatives .

Consider first the case where H does not depend on the endogenous variables and assume further that in addition to (5.1) and (5.2), the following hypothesis hold:

$$(\frac{Z'H}{T}, \frac{H'H}{T}) \underset{T \longrightarrow \infty}{\xrightarrow{p}} (Q_{ZH}, Q_{HH}) \tag{5.4}$$

where Q_{ZH} and Q_{HH} are full rank matrices of dimensions $l \times k$ and $k \times k$. If Z is chosen equal to H, then (5.4) is implied by (5.1). The numerator of $T(\theta_0, f)$ may be written (To simplify notations, we suppress the arguments y and x.)

$$\begin{aligned} \frac{1}{l} f(\theta_0)' P(Z) f(\theta_0) &= \frac{1}{l}\big[u + H(\theta_0 - \theta_1)\big]' P(Z) \big[u + H(\theta_0 - \theta_1)\big] \\ &= \frac{1}{l}\big[u' P(Z) u + 2(\theta_0 - \theta_1)' H' P(Z) u + (\theta_0 - \theta_1)' H' P(Z) H(\theta_0 - \theta_1)\big]. \end{aligned}$$

We also have:

$$u' P(Z) u = \frac{u'Z}{\sqrt{T}} (\frac{Z'Z}{T})^{-1} \frac{Z'u}{\sqrt{T}} \underset{T \longrightarrow \infty}{\xrightarrow{d}} \sigma^2 \chi^2_l,$$

$$\begin{aligned} 2(\theta_0 - \theta_1)' H' P(Z) u &= 2(\theta_0 - \theta_1)' \sqrt{T} \frac{H'Z}{T} (\frac{Z'Z}{T})^{-1} \frac{Z'u}{\sqrt{T}} \\ &\underset{T \longrightarrow \infty}{\xrightarrow{d}} 2(\theta_0 - \theta_1)' \sqrt{T} Q'_{ZH} Q_{ZZ}^{-1} N(0, \sigma^2 Q_{ZZ}), \end{aligned}$$

$$\begin{aligned} (\theta_0 - \theta_1)' H' P(Z) H(\theta_0 - \theta_1) &= T(\theta_0 - \theta_1)' \frac{H'Z}{T} (\frac{Z'Z}{T})^{-1} \frac{Z'H}{T} (\theta_0 - \theta_1) \\ &\underset{T \longrightarrow \infty}{\xrightarrow{p}} T(\theta_0 - \theta_1)' Q'_{ZH} Q_{ZZ}^{-1} Q_{ZH} (\theta_0 - \theta_1). \end{aligned}$$

hence

$$\frac{1}{T-l}f(\theta_0)'P(Z)f(\theta_0) \underset{T\longrightarrow\infty}{\xrightarrow{p}} (\theta_0-\theta_1)'Q'_{ZH}Q_{ZZ}^{-1}Q_{ZH}(\theta_0-\theta_1).$$

The denominator can be written :

$$\frac{1}{T-l}f(\theta_0)'M(Z)f(\theta_0) = \frac{1}{T-l}f(\theta_0)'f(\theta_0) - \frac{1}{T-l}f(\theta_0)'P(Z)f(\theta_0)$$

$$\begin{aligned}\frac{1}{T-l}f(\theta_0)'f(\theta_0) &= \frac{1}{T-l}[u+H(\theta_0-\theta_1)]'[u+H(\theta_0-\theta_1)] \\ &= \frac{1}{T-l}[u'u+2u'H(\theta_0-\theta_1)+(\theta_0-\theta_1)'H'H(\theta_0-\theta_1)] \\ &\underset{T\longrightarrow\infty}{\xrightarrow{p}} \sigma^2+(\theta_0-\theta_1)'Q_{HH}(\theta_0-\theta_1).\end{aligned}$$

Hence the denominator converges to a constant given by

$$\sigma^2+(\theta_0-\theta_1)'[Q_{HH}-Q'_{ZH}Q_{ZZ}^{-1}Q_{ZH}](\theta_0-\theta_1).$$

We deduce that when T gets large, under $H_1 : \theta = \theta_1 \neq \theta_0$, $T(\theta_0, f)$ may be approximated by

$$T(\theta_0,f) \approx \frac{1}{l}\frac{\sigma^2\chi_l^2+T(\theta_0-\theta_1)'Q'_{ZH}Q_{ZZ}^{-1}Q_{ZH}(\theta_0-\theta_1)}{\sigma^2+(\theta_0-\theta_1)'[Q_{HH}-Q'_{ZH}Q_{ZZ}^{-1}Q_{ZH}](\theta_0-\theta_1)}. \tag{5.5}$$

It gets infinitely large and:

$$P[T(\theta_0,f)\geq c(\alpha)] \underset{T\longrightarrow\infty}{\longrightarrow} 1 \tag{5.6}$$

where $c(\alpha)$ is the $1-\alpha$ quantile of χ_l^2/l. The test $T(\theta_0, f)$ is thus consistent. Notice that approximation (5.5) of $T(\theta_0, f)$ increases as Z becomes close to H, the matrix of optimal instrument.

When f is given as in (3.11), we get under H_1:

$$f(Y,X,\theta_0) = f(Y,X,\theta_1)+\frac{\partial f}{\partial\theta'}(Y,X,\bar{\theta})(\theta_0-\theta_1)$$

$$= u + YA'(\theta_0 - \theta_1) + g(X, \bar{\theta})(\theta_0 - \theta_1)$$
$$= u + (Y - E(Y))A'(\theta_0 - \theta_1) + [g(X, \bar{\theta}) + E(Y)A'](\theta_0 - \theta_1).$$

Let $\Delta = [1, (\theta_0 - \theta_1)'A]'$, $\varepsilon = [u, Y - E(Y)]$ and $H = \partial f(Y, X, \bar{\theta})/\partial\theta' = [g(X, \bar{\theta}) + YA']$, we get:

$$f(Y, X, \theta_0) = \varepsilon\Delta + E[H](\theta_0 - \theta_1).$$

If Assumptions (5.1), (5.2) and (5.4) hold with u replaced by $\varepsilon\Delta$, and H by $E(H)$, we obtain the same result as before : under H_1, $P[T(\theta_0, f) \geq c(\alpha)] \underset{T \longrightarrow \infty}{\longrightarrow} 1$.

Finally in the more general case of any continuous and differentiable function, the same results hold if Assumptions (5.1), (5.2) and (5.4) hold with u replaced by $\eta = u + [\partial f/\partial\theta'(Y, X, \bar{\theta}) - E(\partial f/\partial\theta'(Y, X, \bar{\theta}))](\theta_0 - \theta_1)$ and H is replaced by $E(\partial f/\partial\theta'(Y, X, \bar{\theta}))$.

5.3. Asymptotic distribution under a local alternative

Consider now a local alternative:

$$H_1 : \theta = \theta_0 + C/\sqrt{T}$$

where C is a $k \times 1$ vector of fixed values. As T gets large the value of the alternative approaches that of the null.

We consider first the case where $\partial f/\partial\theta$ does not depend on the endogenous variables, and assume as before that Assumptions (5.1) and (5.2) hold. Applying once again the mean-value theorem, we get:

$$f_t(y_t, x_t, \theta_0) = f_t(y_t, x_t, \theta_0 + \frac{C}{\sqrt{T}}) - H_t'\frac{C}{\sqrt{T}}$$

where $H_t = \partial f_t/\partial\theta(y_t, x_t, \theta_0 + p_t C/\sqrt{T})$ and $p_t \in [0, 1]$. Under H_1,

$$f(Y, X, \theta_0) = u - \frac{HC}{\sqrt{T}}.$$

If we assume that Assumption (5.4) holds, we get:

$$\frac{1}{\sqrt{T}}Z'f(Y,X,\theta_0) = \frac{1}{\sqrt{T}}Z'(u - \frac{HC}{\sqrt{T}}) = \frac{1}{\sqrt{T}}Z'u - \frac{Z'HC}{T} \underset{T\longrightarrow\infty}{\overset{d}{\longrightarrow}} N(-Q_{ZH}C, \sigma^2 Q_{ZZ}).$$

$$\begin{aligned} f(Y,X,\theta_0)Z(Z'Z)^{-1}Z'f(Y,X,\theta_0) &= \tfrac{f(Y,X,\theta_0)'Z}{\sqrt{T}}\left(\tfrac{Z'Z}{T}\right)^{-1}\tfrac{Z'f(Y,X,\theta_0)}{\sqrt{T}} \\ &\underset{T\longrightarrow\infty}{\overset{d}{\longrightarrow}} N(-Q_{ZH}C, \sigma^2 Q_{ZZ})Q_{ZZ}^{-1}N(-Q'_{ZH}C, \sigma^2 Q_{ZZ}) \\ &\sim \sigma^2\chi^2_{(\lambda,l)} \end{aligned}$$

where λ is the noncentrality parameter and is given by

$$\lambda = C'Q'_{ZH}Q_{ZZ}^{-1}Q_{ZH}C\ .$$

Finally,

$$\frac{1}{T-l}f(Y,X,\theta_0)M(Z)f(Y,X,\theta_0) = \frac{f(Y,X,\theta_0)'f(Y,X,\theta_0)}{T-l} - \frac{f(Y,X,\theta_0)P(Z)f(Y,X,\theta_0)}{T-l},$$

$$\begin{aligned} \frac{f(Y,X,\theta_0)'f(Y,X,\theta_0)}{T-l} &= \frac{1}{T-l}(u - \frac{HC}{\sqrt{T}})'(u - \frac{HC}{\sqrt{T}}) \\ &= \frac{1}{T-l}[u'u - 2C'\frac{H'u}{\sqrt{T}} + \frac{C'H'HC}{T}] \underset{T\longrightarrow\infty}{\overset{p}{\longrightarrow}} \sigma^2. \end{aligned}$$

So under $H_1 : \theta = \theta_0 + C/\sqrt{T}$, we see that

$$T(\theta_0, f) \underset{T\longrightarrow\infty}{\overset{d}{\longrightarrow}} \frac{\chi^2_{(\lambda,l)}}{l}\ .$$

When f_t depends on the endogenous variables, the same arguments as in the case of a fixed alternative hold and the same result can be obtained.

6. Conclusion

In this paper we proposed an exact and simple test in nonlinear structural models. The **GHAR** test, which is based on an artificial regression, generalizes the approach

of Hartley (1964) for hypothesis testing and confidence sets construction in nonlinear regressions and the Anderson and Rubin (1949) procedure in linear structural models. The first generalization is done in the sense of allowing for endogenous explanatory variables, and the second generalization is done in the sense of allowing for nonlinear structural models and more general type of instruments.

Like the majority of tests robust to identification problems, the proposed test requires the specification of the entire vector of parameters. However, this can be circumvent by using the projection technique as it is used in Dufour (1990, 1997).

We derived the power function of the proposed test and generalized a result by Revankar and Mallela (1972) who gave the power function of the Anderson-Rubin test in a more restricted framework. For a given alternative hypothesis we derived the optimal matrix of instruments for testing the simple hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. The optimal instruments are those that maximize the power of the proposed test at θ_1, and thus are point-optimal instruments. The matrix of optimal instruments cannot be used directly in practice since it is unknown in general, as a solution to this problem we used the split-sample technique [Dufour and Jasiak (2001)].

The exact distribution of the test is derived under the assumptions of independence and normality of the disturbances, but we showed that the test remains valid asymptotically under weaker assumptions similar to those usually assumed to derive asymptotic distributions.

For future work, it will be interesting to extend the approach proposed here to statistical inference on the parameters of several equations of the model (a full information approach).

7. Appendix: Proofs

Proof of Proposition 3.1 Under $H_1 : \theta = \theta_1$, we have $f_t(y_t, x_t, \theta_1) \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Since $\partial f_t(y_t, x_t, \theta)/\partial\theta'$ does not depend on the endogenous variables, conditional on X, we have:

$$f_t(y_t, x_t, \theta_0) = f_t(y_t, x_t, \theta_1) + H_t(\bar{\theta})'(\theta_0 - \theta_1) \sim N[H_t(\bar{\theta}_t)'(\theta_0 - \theta_1), \sigma^2]$$

where

$$H_t(\bar{\theta}) = \frac{\partial f_t(y_t, x_t, \bar{\theta}_t)}{\partial\theta}.$$

and $\bar{\theta}_t$ depends on θ_0, θ_1, and x_t, and is given by (3.4). If $f(Y, X, \theta)$ is the $T \times 1$ vector whose components are $f_t(y_t, x_t, \theta)$, $t = 1, \ldots, T$, and $H(\bar{\theta}) = [H_1(\bar{\theta}), \ldots, H_T(\bar{\theta})]'$, a $T \times k$ matrix, then

$$f(Y, X, \theta_1) \sim N(0, \sigma^2 I_T)$$

and

$$f(Y, X, \theta_0) \sim N[H(\bar{\theta})(\theta_0 - \theta_1), \sigma^2 I_T].$$

$P(Z)$ and $M(Z) = I - P(Z)$ are the projection matrices on two orthogonal and complementary subspaces of $\mathbb{R}^T$. By Cochran's theorem [see Arnold (1981, page 50)], the variables

$$\left\| \frac{P(Z)f(Y, X, \theta_0)}{\sigma} \right\|^2 = \frac{f(Y, X, \theta_0)'P(Z)f(Y, X, \theta_0)}{\sigma^2}$$

and

$$\left\| \frac{M(Z)f(Y, X, \theta_0)}{\sigma} \right\|^2 = \frac{f(Y, X, \theta_0)'M(Z)f(Y, X, \theta_0)}{\sigma^2}$$

are independent and follow noncentral χ^2 distributions with degrees of freedom $rank[P(Z)] = l$, $rank[M(Z)] = T - l$, and noncentrality parameters:

$$\lambda_1 = \frac{1}{\sigma^2}(\theta_0 - \theta_1)'H(\bar{\theta})'P(Z)H(\bar{\theta})(\theta_0 - \theta_1),$$

$$\lambda_2 = \frac{1}{\sigma^2}(\theta_0 - \theta_1)'H(\bar{\theta})'M(Z)H(\bar{\theta})(\theta_0 - \theta_1).$$

This entails that

$$T(\theta_0, f) = \frac{f(Y, X, \theta_0)'P(Z)f(Y, X, \theta_0)/l}{f(Y, X, \theta_0)'M(Z)f(Y, X, \theta_0)/(T-l)}$$

follows a doubly noncentral F-distribution with degrees of freedom l and $T-l$ and noncentrality parameters λ_1 and λ_2 [see Johnson, Kotz, and Balakrishnan (1994, page 443)]. □

Proof of Lemma 3.2 A doubly noncentral F-variable is the ratio of two independent noncentral χ^2-variables divided by their degrees of freedom:

$$F_{k_1,k_2}(\lambda_1, \lambda_2) = \frac{\chi^2_{k_1}(\lambda_1)/k_1}{\chi^2_{k_2}(\lambda_2)/k_2}.$$

Let $F(z|\ k_1, k_2, \lambda_1, \lambda_2)$ and $F(z|\ k_i, \lambda_i)$ be respectively the distribution functions of $F_{k_1,k_2}(\lambda_1, \lambda_2)$ and $\chi^2_{k_i}(\lambda_i)$ evaluated at $z \geq 0$, and $f_{\chi^2_{k_i}(\lambda_i)}(y)$ the density function of $\chi^2_{k_i}(\lambda_i)$ at $y \geq 0$. We have:

$$\begin{aligned}
F(z|\ k_1, k_2, \lambda_1, \lambda_2) &= P\left[\frac{\chi^2_{k_1}(\lambda_1)/k_1}{\chi^2_{k_2}(\lambda_2)/k_2} \leq z\right] \\
&= \int_0^{+\infty}\int_{x/z'}^{+\infty} f_{\chi^2_{k_1}(\lambda_1)}(x) f_{\chi^2_{k_2}(\lambda_2)}(y)dydx \\
&= \int_0^{+\infty}\left[\int_{x/z'}^{+\infty} f_{\chi^2_{k_2}(\lambda_2)}(y)dy\right] f_{\chi^2_{k_1}(\lambda_1)}(x)dx \\
&= \int_0^{+\infty}\left[1 - F(\frac{x}{z'}|\ k_2, \lambda_2)\right] f_{\chi^2_{k_1}(\lambda_1)}(x)dx \\
&= E\left[1 - F(\frac{k_2}{k_1}\frac{\chi^2_{k_1}(\lambda_1)}{z}|\ k_2, \lambda_2)\right]
\end{aligned}$$

where $z' = (k_1/k_2)z$. We know that $F(x|\ k_2, \lambda_2)$ is a decreasing function of λ_2 [see Johnson, Kotz, and Balakrishnan (1994, page 443)], hence $F(x|\ k_1, k_2, \lambda_1, \lambda_2)$ is increasing in λ_2.

Using the same arguments as before,

$$\begin{aligned} F(z|\ k_1,k_2,\lambda_1,\lambda_2) &= P\Big[\frac{\chi^2_{k_1}(\lambda_1)/k_1}{\chi^2_{k_2}(\lambda_2)/k_2} \le z\Big] \\ &= \int_0^{+\infty}\int_0^{yz'} f_{\chi^2_{k_1}(\lambda_1)}(x) f_{\chi^2_{k_2}(\lambda_2)}(y)dxdy \\ &= \int_0^{+\infty}\left[\int_0^{yz'} f_{\chi^2_{k_1}(\lambda_1)}(x)dx\right] f_{\chi^2_{k_2}(\lambda_2)}(y)dy \\ &= \int_0^{+\infty}[F(yz'|k_1,\lambda_1)]\, f_{\chi^2_{k_2}(\lambda_2)}(y)dy \\ &= E\Big[F(\frac{k_1}{k_2}z\chi^2_{k_2}(\lambda_2)|\ k_1,\lambda_1)\Big] \end{aligned}$$

Since $F(x|\ k_1,\lambda_1)$ is decreasing in λ_1, also is $F(z|\ k_1,k_2,\lambda_1,\lambda_2)$. □

Proof of Proposition 3.3 Since λ_2 depends on Z only through λ_1, and since it is a decreasing function of λ_1, we need to find the maximal value of $\lambda_1 = x'H'P(Z)Hx$, where $x = \frac{1}{\sigma}(\theta_0 - \theta_1)$ and $H = H(\bar{\theta}) = \partial f(Y,X,\bar{\theta})/\partial\theta$. If $Z = H$, λ_1 reduces to $\lambda_1 = x'H'Hx$. This solution maximizes λ_1 if $x'H'Hx \ge x'H'P(Z)Hx$ for any $T \times l$ matrix Z. To see that it does, let $C = H - P(Z)H$, we then have :

$$\begin{aligned} C'C &= (H - P(Z)H)'(H - P(Z)H) \\ &= H'H - H'P(Z)H - H'P(Z)H + H'P(Z)H \\ &= H'H - H'P(Z)H. \end{aligned}$$

Hence $H'H - H'P(Z)H$ is a positive semi-definite matrix and

$$x'H'Hx \ge x'H'P(Z)Hx, \forall x.$$

Thus Z is an optimal matrix of instruments if and only if

$$H'H = H'P(Z)H. \tag{7.1}$$

Finally, on noting the equivalence

$$\begin{aligned} H'H = H'P(Z)H &\iff H'M(Z)H = 0 \\ &\iff M(Z)H = 0 \\ &\iff P(Z)H = H, \end{aligned}$$

we see that the optimality condition means that the space spanned by the columns of Z contains the space spanned by the columns of H. □

Proof of Proposition 3.5 When $\theta = \theta_1$, we have $f_t(y_t, x_t, \theta_1) = u_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$. Using the mean-value theorem and (3.11), we can write:

$$\begin{aligned} f_t(y_t, x_t, \theta_0) &= f_t(y_t, x_t, \theta_1) + \frac{\partial f_t(y_t, x_t, \bar{\theta}_t)}{\partial \theta'}(\theta_0 - \theta_1) \\ &= u_t + [Ay_t + g_{1t}(x_t, \bar{\theta}_t)]'(\theta_0 - \theta_1) \\ &= [1, (\theta_0 - \theta_1)'A] \begin{pmatrix} u_t \\ y_t \end{pmatrix} + (\theta_0 - \theta_1)' g_{1t}(x_t, \bar{\theta}_t). \end{aligned}$$

where $g_{1t}(x_t, \theta) = \partial G_{1t}(x_t, \theta)/\partial\theta$. Conditional on X, $(\theta_0 - \theta_1)' g_{1t}(x_t, \bar{\theta}_t)$ is fixed, so that using (3.13),

$$f_t(y_t, x_t, \theta_0) \sim N[(\theta_0 - \theta_1)'AE(y_t) + (\theta_0 - \theta_1)' g_{1t}(x_t, \bar{\theta}_t), \Delta]$$

where $\Delta = [1, (\theta_0 - \theta_1)'A]\Sigma[1, (\theta_0 - \theta_1)'A]'$ is a fixed scalar independent of t.

Let $f(Y, X, \theta_1)$ be the $T \times 1$ vector of components $f_t(y_t, x_t, \theta_1)$, $E(Y)$ the $T \times n$ matrix with i-th row $E(y_t')$ and $g_{1t}(X, \bar{\theta})$ the $T \times k$ matrix whose i-th row is $g_{1t}(x_t, \bar{\theta}_t)'$ $t = 1, \ldots, T$.

We have:

$$f(Y, X, \theta_0) \sim N[E\{yA' + g_1(X, \bar{\theta})\}(\theta_0 - \theta_1), \Delta I_T]$$

and

$$f(Y, X, \theta_0) \sim N[E[H(\bar{\theta})](\theta_0 - \theta_1), \Delta I_T].$$

Since $P(Z)$ and $M(Z) = I - P(Z)$ are the projection matrices on two orthogonal and complementary subspaces of $\mathbb{R}^T$, it follows from Cochran's theorem [see Arnold (1981, page 50)] that

$$\left\| \frac{P(Z)f(Y,X,\theta_0)}{\sqrt{\Delta}} \right\|^2 = \frac{f(Y,X,\theta_0)'P(Z)f(Y,X,\theta_0)}{\Delta}$$

and

$$\left\| \frac{M(Z)f(Y,X,\theta_0)}{\sqrt{\Delta}} \right\|^2 = \frac{f(Y,X,\theta_0)'M(Z)f(Y,X,\theta_0)}{\Delta}$$

are independent and follow noncentral χ^2 distributions with degrees of freedom $rank[P(Z)] = l$ and $rank[M(Z)] = T - l$ and noncentrality parameters

$$\lambda_1 = \frac{1}{\Delta}(\theta_0 - \theta_1)'E[H(\bar{\theta})]'P(Z)E[H(\bar{\theta})](\theta_0 - \theta_1),$$

$$\lambda_2 = \frac{1}{\Delta}(\theta_0 - \theta_1)'E[H(\bar{\theta})]'M(Z)E[H(\bar{\theta})](\theta_0 - \theta_1),$$

respectively. $T(\theta_0, f)$ is thus the ratio of two independent noncentral χ^2 distributions divided by their respective degrees of freedom and hence follows a doubly noncentral F-distribution with degrees of freedom $(l, T-l)$ and noncentrality parameters (λ_1, λ_2).

□

Proof of Proposition 3.6 The proof is similar to the one of Proposition **3.3**, replacing H by $E[\partial f(Y,X,\bar{\theta})/\partial\theta']$. □

Table 1: Power of the GHAR test as a function of the number of instruments

(a) : $T = 50$

l	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$	$\lambda = 10$
1	0.050	0.051	0.060	0.166	0.166	0.500	0.998	1.000
2	0.050	0.057	0.072	0.127	0.127	0.214	0.478	0.789
3	0.050	0.055	0.067	0.110	0.110	0.180	0.408	0.722
4	0.050	0.055	0.064	0.099	0.099	0.158	0.359	0.665
5	0.050	0.054	0.062	0.092	0.092	0.143	0.323	0.615
6	0.050	0.054	0.060	0.088	0.088	0.132	0.294	0.573
7	0.050	0.053	0.059	0.083	0.083	0.124	0.270	0.535
8	0.050	0.053	0.059	0.081	0.081	0.117	0.250	0.499
9	0.050	0.053	0.058	0.077	0.077	0.111	0.234	0.468
10	0.050	0.052	0.057	0.076	0.076	0.106	0.218	0.441
11	0.050	0.053	0.057	0.074	0.074	0.102	0.205	0.415
12	0.050	0.052	0.056	0.072	0.072	0.097	0.194	0.391
13	0.050	0.052	0.056	0.071	0.071	0.095	0.184	0.370
14	0.050	0.052	0.056	0.069	0.069	0.091	0.175	0.350
15	0.050	0.052	0.055	0.068	0.068	0.089	0.167	0.330
16	0.050	0.052	0.055	0.067	0.067	0.086	0.159	0.314
17	0.050	0.052	0.055	0.066	0.066	0.085	0.153	0.298
18	0.050	0.051	0.054	0.065	0.065	0.082	0.146	0.284
19	0.050	0.051	0.054	0.064	0.064	0.081	0.141	0.269
20	0.050	0.051	0.054	0.064	0.064	0.079	0.135	0.256
21	0.050	0.051	0.054	0.063	0.063	0.078	0.131	0.244
22	0.051	0.051	0.053	0.063	0.063	0.076	0.125	0.232
23	0.050	0.051	0.053	0.062	0.062	0.074	0.121	0.222
24	0.050	0.051	0.053	0.061	0.061	0.074	0.117	0.212
25	0.050	0.051	0.053	0.061	0.061	0.072	0.114	0.202
26	0.050	0.051	0.053	0.060	0.060	0.071	0.110	0.193
27	0.050	0.051	0.052	0.060	0.060	0.069	0.106	0.184
28	0.050	0.051	0.053	0.059	0.059	0.069	0.103	0.176
29	0.050	0.051	0.052	0.059	0.059	0.068	0.100	0.169
30	0.050	0.051	0.052	0.058	0.058	0.067	0.097	0.160
31	0.050	0.051	0.052	0.058	0.058	0.066	0.094	0.153
32	0.050	0.050	0.052	0.057	0.057	0.065	0.092	0.146
33	0.050	0.051	0.052	0.057	0.057	0.064	0.089	0.140
34	0.050	0.050	0.052	0.057	0.057	0.063	0.087	0.134
35	0.050	0.051	0.052	0.056	0.056	0.063	0.084	0.128
36	0.050	0.051	0.051	0.055	0.055	0.062	0.082	0.122
37	0.050	0.051	0.051	0.055	0.055	0.061	0.079	0.116
38	0.050	0.051	0.051	0.055	0.055	0.060	0.077	0.111
39	0.050	0.050	0.052	0.054	0.054	0.060	0.075	0.105
40	0.050	0.050	0.052	0.054	0.054	0.058	0.073	0.100

The table gives $P[F_{l,T-l}(\lambda) > F_{l,T-l}(a)]$, *where* $F_{l,T-l}(\lambda)$ *is a non central F distribution with degrees of freedom* $(l, T-l)$, *and noncentrality parameter* λ, *and* $F_{l,T-l}(\alpha)$ *is the quantile of order* α *of a central F distribution with degrees of freedom* $(l, T-l)$

Table 1 (cont'd): Power of the GHAR test as a function of the number of instruments
(b) : $T = 100$

l	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$	$\lambda = 10$
1	0.054	0.050	0.058	0.082	0.173	0.508	0.999	1.000
2	0.051	0.059	0.071	0.091	0.132	0.216	0.486	0.799
3	0.047	0.055	0.070	0.079	0.114	0.182	0.423	0.744
4	0.053	0.054	0.064	0.073	0.106	0.170	0.384	0.687
5	0.047	0.056	0.063	0.068	0.100	0.145	0.338	0.648
6	0.050	0.051	0.058	0.069	0.089	0.137	0.318	0.617
7	0.048	0.048	0.060	0.061	0.089	0.130	0.291	0.574
8	0.049	0.055	0.057	0.063	0.092	0.129	0.268	0.542
9	0.052	0.055	0.060	0.067	0.077	0.117	0.254	0.516
10	0.049	0.053	0.056	0.063	0.082	0.119	0.253	0.497
11	0.051	0.053	0.058	0.062	0.086	0.110	0.227	0.470
12	0.050	0.053	0.058	0.064	0.078	0.106	0.219	0.454
13	0.051	0.053	0.056	0.062	0.075	0.103	0.212	0.432
14	0.053	0.050	0.053	0.059	0.072	0.094	0.204	0.412
15	0.051	0.051	0.056	0.064	0.068	0.099	0.192	0.389
16	0.051	0.050	0.056	0.060	0.071	0.096	0.187	0.388
17	0.051	0.056	0.052	0.055	0.069	0.089	0.175	0.364
18	0.048	0.053	0.055	0.058	0.066	0.088	0.169	0.358
19	0.048	0.048	0.055	0.055	0.065	0.085	0.170	0.344
20	0.054	0.051	0.057	0.056	0.071	0.088	0.161	0.334
21	0.049	0.050	0.056	0.054	0.067	0.087	0.162	0.318
22	0.050	0.047	0.056	0.053	0.066	0.085	0.153	0.305
23	0.049	0.047	0.055	0.057	0.059	0.091	0.150	0.298
24	0.050	0.055	0.055	0.061	0.066	0.086	0.149	0.291
25	0.051	0.051	0.051	0.057	0.062	0.077	0.139	0.276
26	0.049	0.054	0.056	0.053	0.066	0.086	0.142	0.278
27	0.050	0.055	0.053	0.057	0.062	0.080	0.142	0.271
28	0.047	0.051	0.055	0.051	0.065	0.078	0.131	0.269
29	0.053	0.053	0.055	0.054	0.062	0.079	0.136	0.244
30	0.051	0.048	0.054	0.054	0.061	0.077	0.125	0.247
31	0.051	0.051	0.054	0.058	0.062	0.076	0.131	0.243
32	0.050	0.057	0.052	0.058	0.062	0.077	0.123	0.227
33	0.051	0.051	0.054	0.051	0.060	0.067	0.124	0.215
34	0.050	0.054	0.057	0.056	0.065	0.073	0.124	0.223
35	0.051	0.050	0.051	0.056	0.060	0.073	0.123	0.217
36	0.053	0.054	0.051	0.053	0.057	0.073	0.114	0.212
37	0.050	0.049	0.049	0.053	0.058	0.068	0.114	0.208
38	0.048	0.051	0.056	0.052	0.061	0.073	0.111	0.201
39	0.050	0.050	0.055	0.056	0.059	0.074	0.111	0.205
40	0.051	0.053	0.054	0.057	0.057	0.069	0.114	0.194

The table gives $P[F_{l,T-l}(\lambda) > F_{l,T-l}(a)]$, *where* $F_{l,T-l}(\lambda)$ *is a non central F distribution with degrees of freedom* $(l, T-l)$, *and noncentrality parameter* λ, *and* $F_{l,T-l}(\alpha)$ *is the quantile of order* α *of a central F distribution with degrees of freedom* $(l, T-l)$

Table 1 (cont'd): Power of the GHAR test as a function of the number of instruments
(c) : $T = 200$

l	$\lambda = 0$	$\lambda = 0.1$	$\lambda = 0.3$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 5$	$\lambda = 10$
1	0.051	0.053	0.051	0.078	0.172	0.516	0.998	1.000
2	0.050	0.057	0.072	0.084	0.127	0.227	0.497	0.806
3	0.049	0.054	0.068	0.081	0.113	0.192	0.425	0.757
4	0.045	0.054	0.068	0.074	0.102	0.164	0.382	0.705
5	0.049	0.055	0.061	0.070	0.105	0.155	0.348	0.661
6	0.049	0.058	0.063	0.071	0.091	0.143	0.330	0.630
7	0.050	0.050	0.060	0.069	0.089	0.138	0.298	0.596
8	0.045	0.052	0.060	0.066	0.088	0.125	0.294	0.559
9	0.052	0.055	0.058	0.064	0.086	0.119	0.265	0.548
10	0.050	0.050	0.057	0.066	0.076	0.112	0.253	0.511
11	0.052	0.050	0.058	0.063	0.087	0.115	0.239	0.497
12	0.047	0.052	0.059	0.062	0.074	0.116	0.236	0.474
13	0.050	0.058	0.058	0.060	0.076	0.103	0.219	0.462
14	0.050	0.053	0.053	0.064	0.074	0.105	0.217	0.458
15	0.049	0.048	0.054	0.056	0.071	0.108	0.212	0.434
16	0.050	0.052	0.053	0.055	0.070	0.098	0.203	0.415
17	0.052	0.048	0.052	0.059	0.075	0.090	0.197	0.408
18	0.053	0.053	0.057	0.060	0.069	0.095	0.187	0.395
19	0.050	0.052	0.054	0.060	0.073	0.093	0.181	0.384
20	0.049	0.052	0.053	0.059	0.067	0.091	0.184	0.368
21	0.052	0.049	0.053	0.058	0.067	0.094	0.172	0.358
22	0.057	0.051	0.057	0.061	0.066	0.089	0.174	0.349
23	0.049	0.053	0.061	0.058	0.068	0.087	0.171	0.334
24	0.051	0.050	0.053	0.056	0.066	0.089	0.163	0.331
25	0.048	0.054	0.060	0.060	0.064	0.085	0.160	0.324
26	0.057	0.050	0.053	0.063	0.071	0.082	0.159	0.313
27	0.049	0.051	0.057	0.059	0.068	0.082	0.155	0.301
28	0.051	0.050	0.047	0.059	0.068	0.080	0.153	0.300
29	0.052	0.051	0.053	0.056	0.067	0.079	0.148	0.294
30	0.047	0.051	0.056	0.053	0.065	0.081	0.150	0.289
31	0.049	0.052	0.054	0.057	0.070	0.078	0.144	0.277
32	0.047	0.048	0.054	0.054	0.063	0.080	0.134	0.280
33	0.049	0.051	0.054	0.055	0.067	0.082	0.134	0.267
34	0.051	0.049	0.054	0.057	0.062	0.079	0.141	0.264
35	0.051	0.051	0.051	0.059	0.063	0.075	0.136	0.256
36	0.054	0.050	0.058	0.055	0.062	0.078	0.129	0.252
37	0.051	0.051	0.053	0.058	0.066	0.077	0.132	0.254
38	0.045	0.052	0.053	0.062	0.060	0.076	0.126	0.245
39	0.055	0.053	0.053	0.055	0.062	0.080	0.127	0.234
40	0.049	0.052	0.054	0.055	0.061	0.077	0.126	0.236

The table gives $P[F_{l,T-l}(\lambda) > F_{l,T-l}(a)]$, *where* $F_{l,T-l}(\lambda)$ *is a non central F distribution with degrees of freedom* $(l, T-l)$, *and noncentrality parameter* λ, *and* $F_{l,T-l}(\alpha)$ *is the quantile of order* α *of a central F distribution with degrees of freedom* $(l,\ T-l)$

Table 2 : Power of the *GHAR* test when derivatives are exogenous

θ_1	*GNR test*	*Optimal test*
-0.60	0.91	1.00
-0.55	0.84	1.00
-0.50	0.69	1.00
-0.45	0.49	0.94
-0.40	0.31	0.61
-0.35	0.18	0.30
-0.30	0.12	0.15
-0.20	0.06	0.06
-0.10	0.05	0.05
0.001	***0.05***	***0.05***
0.10	0.05	0.05
0.20	0.06	0.06
0.30	0.10	0.11
0.35	0.15	0.17
0.40	0.23	0.28
0.45	0.36	0.46
0.50	0.53	0.69
0.55	0.71	0.88
0.60	0.87	0.98

Table 3 : Power of the *GHAR* test when derivatives are endogenous

β_1	*GNR*	Non feasible optimal test	Optimal test with split sample			
			$s = 2T/10$	$s = 3T/10$	$s = 4T/10$	$s = 5T/10$
-0.60	1.00	1.00	1.00	0.97	0.89	0.98
-0.55	0.98	1.00	0.99	0.93	0.83	0.93
-0.50	0.90	1.00	0.95	0.81	0.71	0.77
-0.45	0.72	1.00	0.80	0.58	0.51	0.53
-0.40	0.46	0.96	0.52	0.35	0.31	0.29
-0.35	0.24	0.54	0.26	0.18	0.17	0.15
-0.30	0.12	0.22	0.12	0.10	0.09	0.09
-0.25	0.08	0.10	0.08	0.07	0.07	0.06
-0.20	0.06	0.06	0.06	0.06	0.05	0.06
0.001	***0.05***	***0.05***	***0.05***	***0.05***	***0.05***	***0.05***
0.20	0.05	0.05	0.05	0.05	0.05	0.05
0.25	0.07	0.07	0.06	0.06	0.06	0.06
0.30	0.08	0.09	0.08	0.07	0.07	0.07
0.35	0.11	0.13	0.10	0.10	0.09	0.08
0.40	0.17	0.21	0.17	0.16	0.15	0.13
0.45	0.26	0.37	0.28	0.25	0.24	0.21
0.50	0.40	0.58	0.46	0.42	0.39	0.32
0.55	0.60	0.82	0.70	0.66	0.63	0.53
0.60	0.81	0.97	0.91	0.88	0.86	0.76

Figure 1.a: Power of the GHAR test as a function of the number of instruments: T = 50

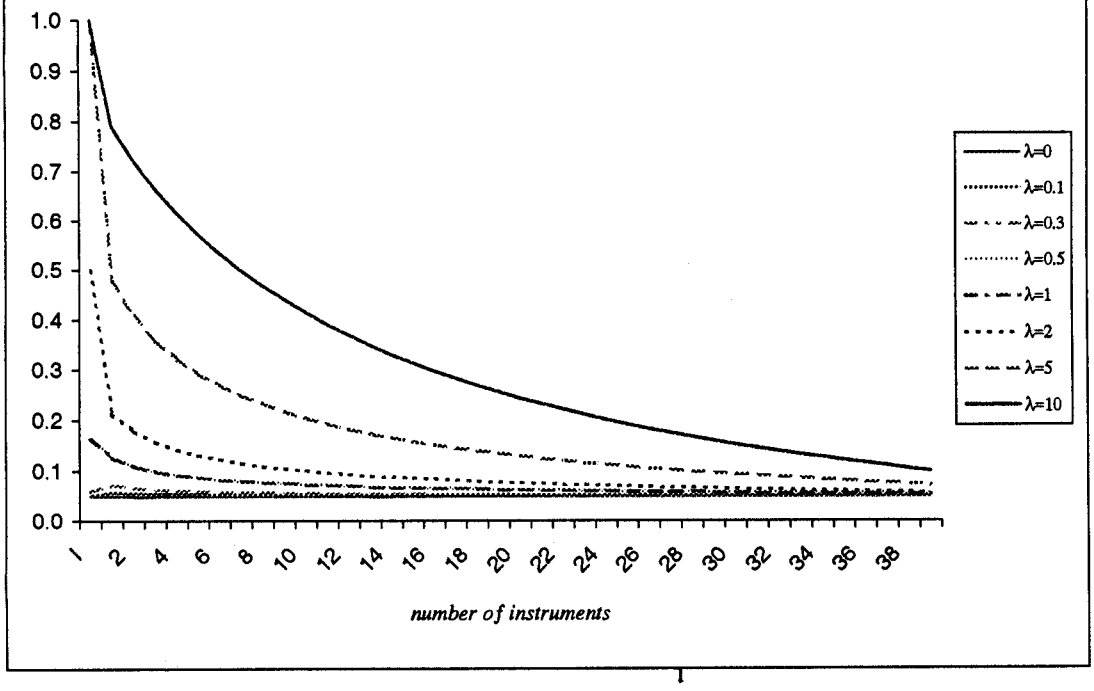


Figure 1.b: Power of the GHAR test as a function of the number of instruments: T = 100

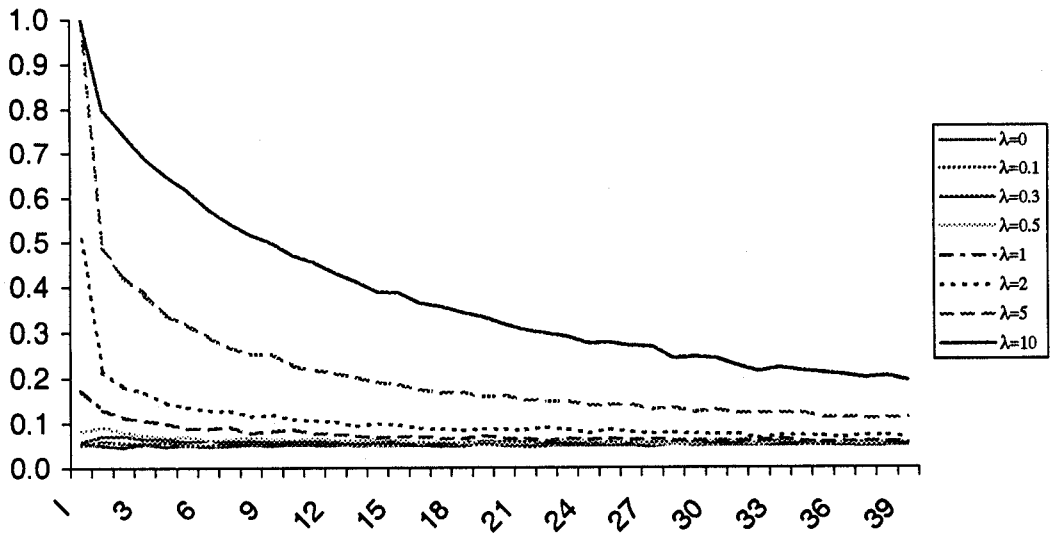


Figure 1.c: Power of the GHAR test as a function of the number of instruments: T = 200

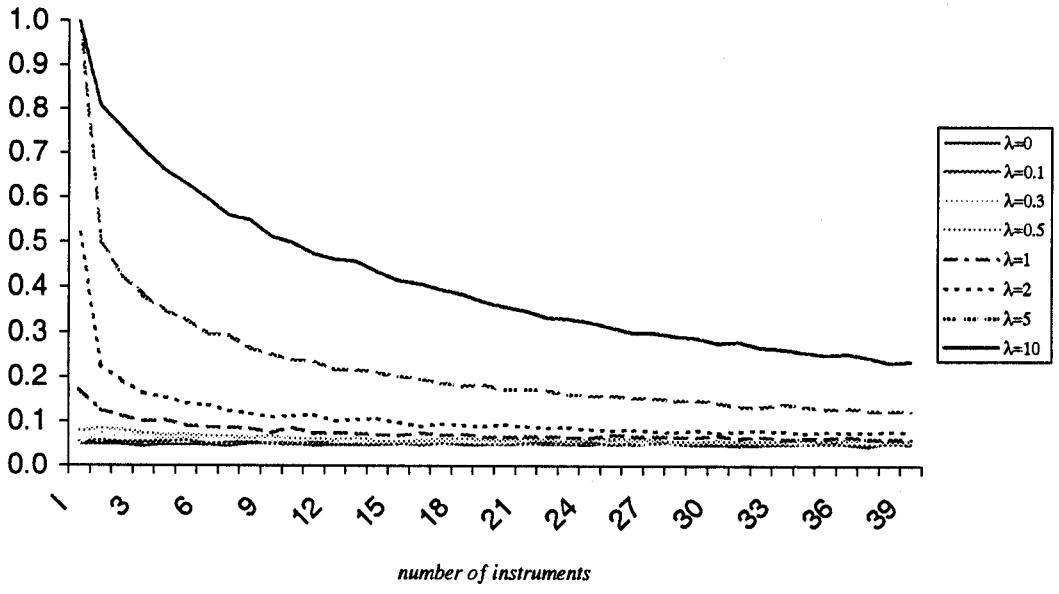


Figure 2: Power of the GHAR test when derivatives are exogenous

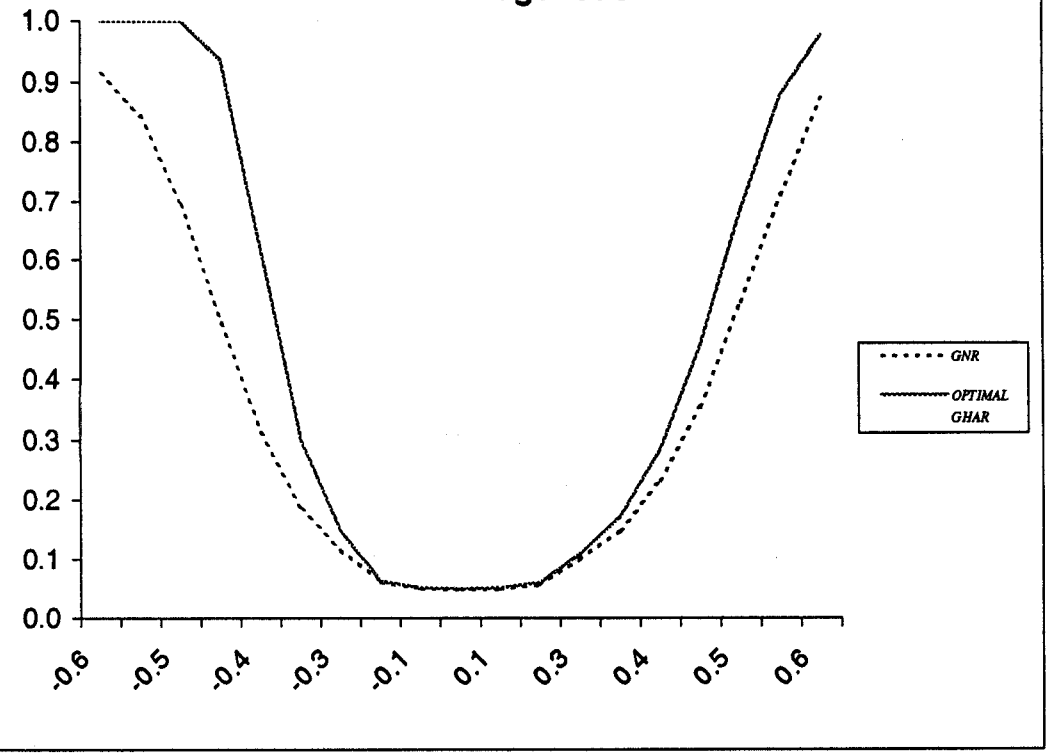
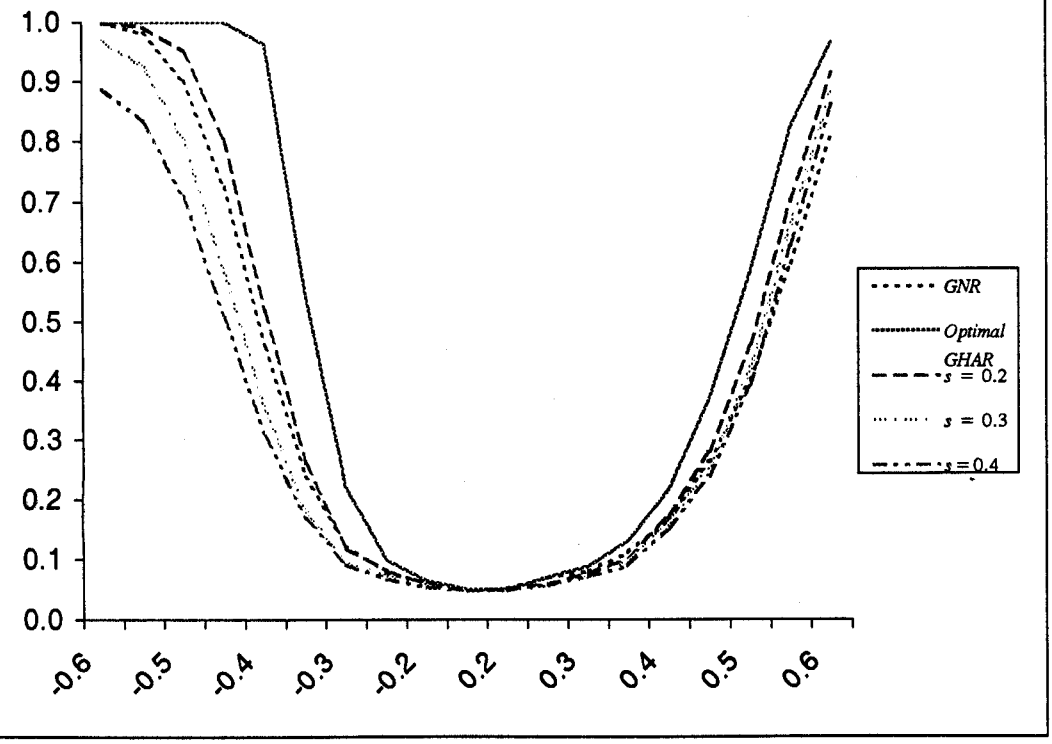


Figure 3: Power of the GHAR test when derivatives are endogenous



Chapter 3

A sequential approach for instrument selection in linear structural models

1. Introduction

The instrumental variables (IV) method is a widely used approach in econometrics designed to handle the problem of endogeneity in econometric models. To apply the method one needs instruments that should verify two main conditions: **exogeneity** (they should be orthogonal to the error term in the structural equation to be estimated) and **relevance** (they should be correlated with the endogenous explanatory variables). For many years, all what empirical researchers cared about was the exogeneity condition; relevance and the number of instruments to use came to forefront only recently. The first paper to draw attention to the consequences of using irrelevant instruments (instruments that are weakly correlated with the endogenous explanatory variables) is the Nelson and Startz (1990a) paper. A decade later the literature on weak instruments is large and it continues to be a topic of interest in econometrics; see for example, Nelson and Startz (1990b), Buse (1992), Maddala and Jeong (1992), Bound, Jaeger, and Baker (1993, 1995), Angrist and Krueger (1995), Hall, Rudebusch, and Wilcox (1996), Dufour (1997), Shea (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz, and Nelson (1998), Startz, Nelson, and Zivot (1999), Dufour and Jasiak (2001), Kleibergen (2001a, 2001b), Perron (1999), Dufour and Taamouti (2000b) and Stock and Wright (2000).

The most important consequence of weak instruments is the unreliability of standard asymptotic theory. The fact that standard asymptotic theory can be arbitrarily inaccurate in finite samples (of any size) is shown rigorously in Dufour (1997), where it is observed that valid confidence intervals in a standard linear structural equations model must be unbounded with positive probability and Wald-type statistics have distributions which can deviate arbitrarily from their large-sample distribution (even when identification holds). Further, even the asymptotic distributions of many usual estimators are far from normality and depend on nuisance parameters. Staiger and Stock (1997) derived (under the local to nonidentification assumption) the asymptotic distributions of IV estimators and limited information maximum likelihood (LIML) estimator in simultaneous equations models. They showed that the usual asymptotic

distributions of these estimators are inaccurate. Stock and Wright (2000) considered the same problem in the more general context of models estimated by GMM, and found the same result. The other features emphasized in the literature are the significant bias of 2SLS in the same direction as OLS [Buse (1992) and Staiger and Stock (1997)], fat tails far from normality and concentration of the distributions away from the true value with small misleading standard errors [Nelson and Startz (1990a)] and oversized t-tests [Hall, Rudebusch, and Wilcox (1996), and Dufour and Jasiak (2001)]. Finally, In finite-sample distributional theory, Choi and Phillips (1992) shows that when the model is partially identified, finite-sample distributions of usual IV estimators exhibit strong dependence on nuisance parameters and usual test statistics are not pivotal.

All these problems raised the interest of many researchers in the selection of instruments. The instruments should be relevant as well as exogenous. The natural way to proceed is to choose instruments that optimize an objective, for example to minimize the variance or the mean squared error (MSE) of an estimator [Amemiya (1977), Newey (1993), Donald and Newey (Forthcoming)] or to maximize the power of a test [Dufour and Taamouti (2000a)]. A survey of a related literature in the context of models estimated by GMM is provided by Mátyás (1999). Unfortunately this usually suggests an optimal matrix of instruments that is unknown (because it depends on unknown parameters). The other way to proceed is to define formally the notion of relevance and then choose the instruments that are most relevant. This approach has been particularly followed by Hall and Peixe (2000) who proposed to use the canonical correlation between instruments and endogenous explanatory variables as a measure of relevance and select the set of instruments that minimizes a canonical correlation criterion. Earlier, Shea (1997) suggested another criterion for instruments relevance which is the partial correlation between each endogenous explanatory variable and the set of instruments. The problem with this type of approach is that it is cumbersome to apply in practice, because of the potentially very large number of instruments subsets to consider (2^l subsets for l instruments).

To insure identification, one may wish to select a large number of instruments.

However, this may be harmful because, even though adding irrelevant instruments does no harm asymptotically [see White (1984)], the consequences may be quite different in finite samples. In particular, this will depend on the relevance of the instruments used. If an instrument is weak, it will increase bias [see Phillips (1980) and Buse (1992)]. This raises another issue: choosing the right number of instruments. Donald and Newey (Forthcoming) proposed an approximate mean squared error criterion to choose the number of instruments to use in the LIML, 2SLS, and a jackknife version of 2SLS [see Angrist, Imbens, and Krueger (1999)]. They take as given an ordered vector of potential instruments and they look for the number k such that the first k instruments optimize the MSE criterion. This approach has an important drawback: if the irrelevant instruments are put first in the matrix of potential instruments, the set of instruments that will be chosen will include all the first irrelevant instruments, and the best choice may not be a good one, as shown in their simulations (Table 3 of their paper).

In this paper, we examine the problem of selecting instruments in the case of a linear structural equation for estimating or testing. We concentrate on the case of 2SLS method for estimation and Anderson and Rubin (1949) (hereafter AR) test for inference. We first analyze what determines the performance of these methods and how the matrix of instruments affects the quality of estimation and inference. For this purpose, we use some early results on finite-sample properties of 2SLS. We then consider the problem of choosing a subset from a large set of potential instruments on the basis of these results. The problem considered here is different from the issue addressed in Andrews (1999) who propose selection procedures for valid instruments (those verifying the exogeneity condition) [see also Wu (1973), Hausman (1978), Dufour (1987), and Magdalinos (1994) for some related literature].

We consider the case of one endogenous explanatory variable. This case is interesting because it is often met in empirical work and at the same time the exact distributions of the usual estimators and test statistics are relatively tractable. Further many recent papers have considered the same framework to provide interesting results on IV [Zivot, Startz, and Nelson (1998), Donald and Newey (Forthcoming),

Chao and Swanson (2000), Hahn and Hausman (Forthcoming)].

The method we propose is based on the sequential maximization of the concentration parameter. It is an upward procedure since it begins by selecting one instrument and then adding others depending on their relevance. It is simple to apply and is reliable even in small samples as is shown by Monte Carlo simulations. In large samples, the method is shown to be consistent in the sense that it selects with probability 1 only instruments with non zero coefficient in the true DGP of the endogenous explanatory variable.

We also investigate whether selection of instruments and relevance diagnostic procedures are useful in practice as screening tests, or as suggested in Hall, Rudebusch, and Wilcox (1996) paper they are useless. We use the split-sample technique [Dufour and Jasiak (1993), Angrist and Krueger (1995)] to circumvent the problem of control of the tests levels. This technique consists in dividing the sample into two parts, one is used for selecting instruments, and the other for statistical inference. Finally, we evaluate and compare the main selection methods proposed in the literature on the basis of a Monte Carlo simulation.

The main framework of the paper is presented in Section 2. In Section 3 we review some results on finite sample characteristics of 2SLS and AR test, as well as the recent literature on relevance diagnostic and instrument selection. Section 4 presents some results on the relation between the number of instruments, the concentration parameter and estimation and inference quality. Section 5 presents a new selection method based on finite-sample properties of 2SLS and AR test. The consistency of the method is proven is Section 6. A Monte Carlo study is presented in Section 7 which compares the main methods suggested in this literature. Section 8 presents an empirical application and Section 9 concludes.

2. Framework

We study here a linear structural model of the form:

$$y = Y\beta + X_1\gamma + u \ , \tag{2.1}$$

$$Y = X_2\Pi_2 + X_1\Pi_1 + U_2 \ , \tag{2.2}$$

where y is a $T \times 1$ vector containing the endogenous variable to be explained, Y is a $T \times G$ matrix of endogenous explanatory variables, X_1 is a $T \times k_1$ matrix of included exogenous variables, X_2 is a $T \times k_2$ matrix of excluded exogenous instruments, u is a $T \times 1$ vector of structural disturbances of mean 0, and U_2 is a $T \times G$ matrix of reduced form disturbances. β and γ are vectors of unknown parameters. Π_2 and Π_1 are respectively $k_2 \times G$ and $k_1 \times G$ matrices of reduced form coefficients. From (2.1)-(2.2) we deduce easily the reduced form equation for the endogenous variable y :

$$y = X_2\pi_2 + X_1\pi_1 + v_1 \tag{2.3}$$

where π_2 and π_1 are respectively $k_2 \times 1$ and $k_1 \times 1$ vectors of reduced form coefficients and v_1 is a $T \times 1$ vector of disturbances of mean 0.

We assume that the columns of X_2 are selected from a possibly large matrix of potential instruments X of dimension $T \times l$. An admissible choice of X_2 must insure identifiability of equation (2.1): X_2 should contain at least G instruments and have full column rank. To get exact distributional theory and inference, we assume that

$$V = [v_1, U_2] \text{ is independent of } X, \tag{2.4}$$

and that the rows of V verify the following assumption:

$$V_t \overset{i.i.d.}{\sim} N(0, \Sigma), \ t = 1, \ldots T. \tag{2.5}$$

The parameter of interest is β : our purpose is to estimate or produce tests and

confidence sets for it.

We focus on two methods, namely 2SLS for estimation and the AR test for inference. The 2SLS estimator of the vector $\theta = (\beta', \gamma')'$ is given by

$$\tilde{\theta} = [Z'P(\bar{X})Z]^{-1}Z'P(\bar{X})y \tag{2.6}$$

where $Z = [Y, X_1]$, $\bar{X} = [X_1, X_2]$ and $P(\bar{X})$ is the projection matrix on the space spanned by the columns of $\bar{X}$, $P(\bar{X}) = \bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}$. For the parameter of interest β, the 2SLS estimator is given by

$$\tilde{\beta} = \left[Y'P[M(X_1)X_2]Y\right]^{-1}Y'P[M(X_1)X_2]y \tag{2.7}$$

where $M(X_1) = I - P(X_1)$. The asymptotic properties of this estimator are well known under assumptions weaker than (2.5) [see, for example, Johnston (1972)]. Assumption (2.5) is however necessary for the derivation of exact finite-sample properties of this estimator.

For testing and confidence sets on β, several test statistics (Wald, LR, ...) can be used. However, if the instruments X_2 are weak, i.e. if equation (2.1) is weakly identified many of them may have serious problems [see Nelson and Startz (1990a, 1990b), Bound, Jaeger, and Baker (1993, 1995), Hall, Rudebusch, and Wilcox (1996), Dufour (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz, and Nelson (1998), Startz, Nelson, and Zivot (1999), Kleibergen (2001a, 2001b)]. The AR test has the important feature of remaining valid despite such problems. To test

$$H_0 : \beta = \beta_0 \ , \tag{2.8}$$

the AR statistic is given by

$$T(\beta_0) = \frac{(y - Y\beta_0)'[M(X_1) - M(\bar{X})](y - Y\beta_0)/k_2}{(y - Y\beta_0)'M(\bar{X})(y - Y\beta_0)/(T - k)} \tag{2.9}$$

where $k = k_1 + k_2$. This test is asymptotically valid under assumptions weaker than

(2.5), but in finite-samples this assumption guarantees an exact F-distribution with degrees of freedom k_2 and $T-k$ [see Dufour and Jasiak (2001)]. This test may be inverted to construct confidence sets and test specifications for transformations of β [see Dufour and Taamouti (2000b)]. Note also that the test may be computed for the vector θ in the same way.

3. Literature review

This section presents a review of the relevant econometric and statistical literature on two topics related to this paper. The first topic concerns the finite sample properties of 2SLS estimator and the AR test. The results of this literature provide a basis for the selection method we propose in Section 5. The second topic is the problem of instrument relevance, diagnostic and selection.

3.1. Finite-sample properties of 2SLS estimator and AR test

3.1.1. Bias and MSE of 2SLS estimator

The finite sample characteristics of IV estimators have been extensively studied in the seventies and early eighties. A survey of the main results on this topic is provided by Mariano (1982), Phillips (1983), and Taylor (1983). Despite the assumptions adopted, the expressions for the density functions of these estimators remain difficult to handle and to deduce from the main features of the estimators. In the case of G endogenous variables, the density function is given by Phillips (1983). It is derived from the fact that the 2SLS estimator of β can be written in the form:

$$\tilde{\beta} = \left[\left(\bar{Y}'P[M(X_1)X_2]\bar{Y}\right)_{22}\right]^{-1}\left(\bar{Y}'P[M(X_1)X_2]\bar{Y}\right)_{21} \qquad (3.1)$$

where $\bar{Y} = [y, Y]$. Since the rows of V are assumed to be *i.i.d.* $N(0,\Sigma)$ and since $\bar{Y}_t' = \bar{X}_t'\Pi + V_t'$, where $\Pi = \begin{bmatrix} \pi_1 & \Pi_1 \\ \pi_2 & \Pi_2 \end{bmatrix}$, the rows of the matrix $\bar{Y}$ are *i.i.d.*

$N(\bar{X}_i'\Pi, \Sigma)$ and the matrix $\bar{Y}'P[M(X_1)X_2]\bar{Y}$ follows a non central Wishart distribution [see Anderson (1984)] with degree of freedom k_2 and non-centrality matrix $M = E[\bar{Y}]'P[M(X_1)X_2]E[\bar{Y}]$. Thus the distribution of $\tilde{\beta}$ is a function of a Wishart distribution. In fact, beyond the expression of the density, little of interest can be said for the case of more than 2 endogenous variables.[1] This is one of the reasons we focus in this paper on the case of one endogenous explanatory variable. This case is interesting because its exact finite sample characteristics are relatively tractable and it is quite often met in empirical work [e.g. labor economics and macroeconomics]. Further, many recent papers have considered the same framework to provide interesting results on IV; see Zivot, Startz, and Nelson (1998), Donald and Newey (Forthcoming), Chao and Swanson (2000), and Hahn and Hausman (Forthcoming).

The exact finite-sample distribution of the 2SLS estimator of β when $G = 1$ has been studied in several papers. A review of the main results is provided by Mariano (1982). We focus here on the formulas for the bias and mean squared error, which are the more common quality measures of an estimator. These formulas are provided under Assumption (2.5) and the following one

$$\bar{X} \text{ is a nonstochastic matrix and } X_1'X_2 = 0 \tag{3.2}$$

where $\bar{X} = [X_1, X_2]$. Note that the condition that $X_1'X_2 = 0$ is without loss of generality. The bias of $\tilde{\beta}$ [see Richardson and Wu (1971, page 976, equation (3.1))] is given by

$$B(\beta) = E(\tilde{\beta}) - \beta = -(\beta - \rho)e^{-\frac{T\delta}{2}}{}_1F_1(\frac{k_2}{2} - 1, \frac{k_2}{2}, \frac{T\delta}{2}) \tag{3.3}$$

where $\rho = \sigma_{12}/\sigma_{22}$, δ is called the **concentration parameter** defined by

$$\delta = \frac{\Pi_2'X_2'M(X_1)X_2\Pi_2/T}{\sigma_{22}}, \tag{3.4}$$

[1] Ullah and Nagar (1974) derived the mean of the 2SLS in the case of 3 endogenous explanatory variables, however its expression is difficult to interpret.

σ_{12} and σ_{22} are components of the matrix $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$, and ${}_1F_1(p,q,x)$ is the hypergeometric confluent function [see Slater (1960)]:

$$ {}_1F_1(p,q,x) = \sum_{h=0}^{+\infty} \frac{(p)_h}{(q)_h h!} x^h, \quad (a)_h = \frac{\Gamma(p+h)}{\Gamma(p)} $$

where p, q and x are any real numbers excluding the values of $q = 0, -1, -2, \ldots$. The bias of OLS estimator is given by the same formulas as (3.3) except that k_2 is replaced by $T - k_1$.

It is important to note here that the instruments used for estimation influence the bias only through their number k_2 and the concentration parameter δ. This means that all instrument subsets with the same number of instruments and the same concentration parameter give the same bias for 2SLS. The main idea we exploit in this paper is the fact that the mean squared error of 2SLS and the power of the Anderson-Rubin test also depend upon the instruments only through these two parameters.

The mean squared error of 2SLS [see Richardson and Wu (1971, page 978, equation (4.1))] is given by

$$ \begin{aligned} \mathrm{MSE}(\beta) &= \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}^2} e^{\frac{-T\delta}{2}} \Big\{ \frac{1}{k_2 - 2}(1 + \bar{\beta}^2) {}_1F_1\Big(\frac{k_2}{2} - 1, \frac{k_2}{2}, \frac{T\delta}{2}\Big) \\ &\quad + \frac{k_2 - 3}{k_2 - 2}\bar{\beta}^2 \, {}_1F_1\Big(\frac{k_2}{2} - 2, \frac{k_2}{2}, \frac{T\delta}{2}\Big) \Big\} \end{aligned} \tag{3.5} $$

where $\bar{\beta} = (\sigma_{22}\beta - \sigma_{12})/\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$. We assume that $k_2 \geq 3$. For $k_2 < 3$ the mean squared error of 2SLS does not exist.[2] As noticed before the MSE depends on the instruments only through their number k_2 and the concentration parameter δ.

[2]When $G = 1$, the moments of 2SLS estimator exist up to $k_2 - 1$ while those of OLS estimator exist up to $T - 1$ [see, for example, Sawa (1969)].

3.1.2. Power of the AR test

The power function of the AR test (2.9) has been studied by Revankar and Mallela (1972) and Dufour and Taamouti (2000a) [see also Maddala (1974) for a Monte Carlo study]. Under the null hypothesis $H_0 : \beta = \beta_0$, the AR test follows a central F-distribution with k_2 and $T - k$ degrees of freedom. Under an alternative hypothesis $H_1 : \beta = \beta_1$, the AR test is distributed according to a non-central F-distribution with degrees of freedom k_2 and $T - k$ and non-centrality parameter given by

$$\lambda = \frac{(\beta_0 - \beta_1)^2}{\Delta} \Pi_2 X_2' M(X_1) X_2 \Pi_2 = \frac{\sigma_{22}(\beta_0 - \beta_1)^2}{\Delta} T\delta \tag{3.6}$$

where $\Delta = e' var[(u_t, U_{2t})'] e$, $e = (1, \beta_1 - \beta_0)'$, and

$$var[(u_t, U_{2t})'] = \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix} \Sigma \begin{pmatrix} 1 & 0 \\ -\beta & 1 \end{pmatrix}.$$

The power of the AR test depends on the instruments again only through their number (in the degrees of freedom) and the concentration parameter δ.

The main consequence of these remarks is the following: if we aim either at estimation using 2SLS or inference based on the AR test, the choice of the set of instruments depends on two basic parameters: the number of instruments, and the concentration parameter which represents the degree to which the endogenous explanatory variable Y is explained by the set of instruments after removing the effect of the included exogenous variables X_1. The role of δ was emphasized in many papers in the literature using asymptotic arguments and appears in the approximation of the bias of 2SLS [Buse (1992), Bound, Jaeger, and Baker (1995), Hahn and Hausman (Forthcoming)]. Here, under the additional assumption of normality the parameter is showed to be crucial for both estimation (bias and MSE) and testing.

3.2. Relevance diagnostics and instrument selection

Since the first results on the consequences of weak instruments have appeared, many researchers have tried to find methods to diagnostic situations where such problems are present and to find solutions that improve estimation and inference. Three related issues have been studied: (1) diagnostic the quality of a given set of instruments [Bound, Jaeger, and Baker (1995), Shea (1997)]; (2) tests for identifiability of the model [Cragg and Donald (1993), Hall, Rudebusch, and Wilcox (1996), and Wright (2000)]; (3) in the situation where a large set of potential instruments is available, choose the best or the most relevant ones [Hall and Peixe (2000) and Donald and Newey (Forthcoming)]. A review of the early literature on IV methods may be found in Bowden and Turkington (1984).

3.2.1. Diagnostics for relevance

3.2.1.1. Cragg and Donald (1993) and Wright (2000). In an early paper, Cragg and Donald (1993) developed a test for the rank condition of identifiability in IV models. The model considered is given by (2.1)−(2.2). The structural equation (2.1) is identified if $\rho(\Pi_2) = G$, where $\rho[.]$ indicates the rank of a matrix and G is the number of the columns of Y. The null hypothesis of nonidentifiability can be expressed as:

$$H_I : \rho(\Pi_2) < G.$$

Let $\hat{\Pi} = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y}$, the OLS estimator of Π, $\Delta = vec(\Pi)$, $\hat{\Delta} = vec(\hat{\Pi})$ and Δ^0 the true value of Δ. Suppose the following assumptions hold:

(i) $\bar{X}'\bar{X}$ is of full rank,

(ii) $\hat{\Pi}$ is a consistent estimator of Π,

(iii) $\sqrt{T}(\hat{\Delta} - \Delta^0) \overset{a}{\sim} N(0, W)$ where W is a positive-definite matrix that can be estimated by an almost surely positive-definite consistent estimator $\hat{W}$.

The test statistic for

$$H_0 : \rho(\Pi_2) = p < G$$

is given by

$$C(\hat{\Pi}) = T \min_{\Pi_2 \in \Omega(p)} (\hat{\Delta} - vec(\Pi))' \hat{W}^{-1} (\hat{\Delta} - vec(\Pi))$$

where $\Omega(p)$ is the subset of matrices Π_2 of rank p. Under H_0 the asymptotic distribution of $C(\hat{\Pi})$ is given by

$$C(\hat{\Pi}) \overset{a}{\sim} \chi_{((G-p)(k_2-p))}.$$

A drawback of the test is that it tests a specified value of the rank of Π_2, whereas the hypothesis of nonidentifiability is $\rho(\Pi_2) \leq G - 1$. Under certain additional assumptions, the authors show that $C(\hat{\Pi})$ with $p = G - 1$ provides asymptotically a conservative test for $H_0 : \rho(\Pi_2) < G$. However their simulations show that the test is very conservative even when the true rank is $G - 2$ and its conservatism increases with the rank deficiency.

An extension of this test to the context of models estimated by generalized method of moments (GMM) is provided by Wright (2000). The rank test is applied to the derivative (with respect to the unknown parameter) of the moment conditions. To deal with the presence of nuisance parameters in this case, Wright uses the Bonferroni inequality in the same way as Staiger and Stock (1997). However as, his simulation shows, this procedure makes the test very conservative.

3.2.1.2. Bound, Jaeger, and Baker (1995). Bound, Jaeger, and Baker paper emphasizes the consequences of using weak instruments. They use the data considered by Angrist and Krueger (1991) to illustrate that the results of IV estimation may be misleading by showing small standard errors while in fact small sample bias may be large. On the basis of the results of Buse (1992), they show that if a set of potential instruments is weakly correlated with the endogenous explanatory variables, then even a small correlation between the instruments and the structural error can seriously bias IV estimates relatively to OLS even with large sample sizes. They suggest to use the partial R^2 and the F statistic on the excluded instruments in the first-stage regression as a rough guide to the quality of IV estimates, and recommend that both statistics be routinely reported when IV estimates are presented.

3.2.1.3. Hall, Rudebusch and Wilcox (1996). Hall, Rudebusch and Wilcox considered the problem of assessing instrument relevance in the context of linear models with i.i.d. observations [model (2.1)]. If X_2 is a $T \times k_2$ matrix of instruments with $k_2 \geq G$, they focus on the IV estimator given by (2.6), and which can also be obtained by minimizing with respect to $\theta = (\beta', \gamma')'$ the objective function

$$Q_T(\theta) = u(\theta)' P(\bar{X}) u(\theta) \tag{3.7}$$

where $u(\theta) = y_t - Y_t'\beta - X_{1t}'\gamma$. Following Bowden and Turkington (1984, page 36),

$$\frac{Q_T(\theta)}{T} = (\theta - \theta_0)' V_T^{-1} (\theta - \theta_0) + o(1) \tag{3.8}$$

where $V_T = T(Z'P(\bar{X})Z)^{-1}$ and $Z = [Y, X_1]$. V_T can also be expressed as:

$$V_T = A\Lambda^{-2}A' \tag{3.9}$$

where A is an $(G + k_1) \times (G + k_1)$ matrix whose i-th column yields the linear combination of Z_t that gives the i-th canonical correlation r_i, $i = 1, ..., G + k_1$ with the appropriate linear combination of $\bar{X}$, and $\Lambda = diag[r_1, r_2, ..., r_{G+k_1}]$. Equations (3.8) and (3.9) imply that β is identified only if all the canonical correlations converge to nonzero limits. So we can test instrument relevance by testing whether the smallest canonical correlation between Z and $\bar{X}$ equals zero:

$$H_0 : r_{G+k_1} = 0.$$

The likelihood ratio statistic is given by

$$\text{LR} = -T \; ln(1 - r_{G+k_1}^2), \tag{3.10}$$

and is asymptotically distributed χ_1^2.

The Monte Carlo evaluation of this test shows two main features. First, for a given set of instruments, the test behaves quite well in detecting whether they are relevant,

the size is controlled and it has power against non relevance. The second feature however shows that the statistic cannot be used as a pre-estimation screening test for instrument relevance, worse, it exacerbates the poor finite-sample properties of the IV estimator. The authors suggest that this deterioration may reflect the fact that when endogeneity is serious, a high sample correlation between Z and $\bar{X}$ is associated with high sample correlation between $\bar{X}$ and u; in other words, the spuriously high relevance of the instrument is associated with a spuriously high endogeneity of the instrument. In Section 7, we use the same Monte Carlo design as Hall, Rudebusch, and Wilcox (1996) to verify this assertion.

3.2.1.4. Shea (1997). Shea considered the problem of measuring relevance in multivariate models (models with more than one endogenous explanatory variable). In the simple case of two explanatory variables, the model can be written:

$$y = Y_1\beta_1 + Y_2\beta_2 + u.$$

The main idea of Shea is the following: for a matrix of instruments X to be relevant for an explanatory variable, say Y_1, it must have components important to Y_1 that are linearly independent of those important to Y_2. To see this, consider the case where X contains two instruments $X = [x_1, x_2]$. Suppose that x_1 is highly correlated with both Y_1 and Y_2 but x_2 is uncorrelated with both Y_1 and Y_2. Then regressing Y_1 or Y_2 on X will give a high R^2 while in fact $\beta = (\beta_1, \beta_2)'$ is unidentified. Shea proposes to use the partial R^2, denoted R_p^2, to measure relevance of instruments for each of the endogenous explanatory variables.[3] The R_p^2 for Y_1 is defined to be the squared correlation between $\tilde{Y}_1 = M(Y_2)Y_1$ (the component of Y_1 orthogonal to Y_2) and $\bar{Y}_1 = M(P(X)Y_2)P(X)Y_1$ the component of Y_1's projection on X orthogonal to Y_2's projection on X. The sample partial R^2 can be written

$$R_p^2(Y_1) = \frac{(\bar{Y}_1'\tilde{Y}_1)^2}{(\bar{Y}_1'\bar{Y}_1)(\tilde{Y}_1'\tilde{Y}_1)} = \frac{(\bar{Y}_1'\tilde{Y}_1)}{(\bar{Y}_1'\bar{Y}_1)}. \tag{3.11}$$

[3]This is different from the partial R^2 considered in Bound, Jaeger, and Baker (1995).

When we have only one endogenous explanatory variable, this measure reduces to the standard R^2, and if we have one endogenous explanatory variable and more than one exogenous explanatory variable then this measure is the same as the partial R^2 reported in Bound, Jaeger, and Baker (1995). As the standard R^2, adding more instruments (even irrelevant) increases R_p^2, therefore R_p^2 may be replaced by

$$\bar{R}_p^2 = 1 - \frac{T-1}{T-k}(1 - R_p^2) \tag{3.12}$$

where k is the number of instruments.

Compared to canonical correlations measure proposed by Hall, Rudebusch, and Wilcox (1996), Shea argues that R_p^2 has the advantage of assigning a relevance measure to each Y_i, allowing the researcher to pinpoint variables that require better instruments, while canonical correlations do not map readily into particular Y variables. More importantly, canonical correlations do not distinguish problems due to instrument irrelevance from those due to poor conditioning on Y (that may result for example, from a high multicollinearity among the columns of Y). Meanwhile, R_p^2 measures relevance alone [Shea (1997, equation (7))].

3.2.2. Instrument selection

3.2.2.1. Hall-Piexe selection method. Hall and Peixe (2000) considered the problem of instrument selection in the case of models with i.i.d. observations to be estimated by generalized instrumental variables (GIV). Their method is exposed in the case of the univariate linear regression model (2.1) with $\gamma = 0$. The population based moment conditions for GMM estimation are:

$$E[X_{2t}u_t(\beta)] = 0 \tag{3.13}$$

where X_{2t} is a $k_2 \times 1$ vector of instruments assumed to be chosen from some candidate set of l instruments denoted X_t that are orthogonal to $u_t(\beta) = y_t - Y_t'\beta$, and β is the true value of the parameter. If S_{k_2} is the $k_2 \times l$ selection matrix that selects X_{2t}, we

have:

$$X_{2t} = S_{k_2} X_t.$$

The concept of relevance is defined on the basis of a canonical correlation criterion. Let $\rho_i[a_t : b_t]$ denote the i-th canonical correlation between a_t and b_t.

Definition 3.1 *An instrument x_{tj} is said to be relevant if*

$$\rho_i[Y_t : \{x_{tj}, S_{jq} X_t\}] - \rho_i[Y_t : S_{jq} X_t] \geq 0$$

for all $i = 1, 2, ..., G$, all $q = G, G+1, ..., l-1$ and all S_{jq}, and

$$\rho_i[Y_t : \{x_{tj}, S_{jq} X_t\}] - \rho_i[Y_t : S_{jq} X_t] > 0$$

for some i, all $q = G, ..., l-1$, and all $q \times l$ matrices S_{jq} whose j-th column is $\mathbf{0}$.

For a given choice of S_{k_2}, the canonical correlations information criterion (CCIC) is defined by

$$CCIC(S_{k_2}) = \sum_{i=1}^{G} \ln(1 - r_i^2) + (k_2 - G)\frac{f(T)}{T} \tag{3.14}$$

where r_i, $i = 1, \ldots, G$ are the sample canonical correlations between Y_t and $X_{2t} = S_{k_2} X_t$. Three choices were considered for the function f : $f(T) = 2$, $f(T) = ln(T)$, and $f(T) = Qln(ln(T))$ with $Q > 2$. These functions are associated respectively with the Akaike, Schawrz, and Hannan-Quinn information criteria. The main result of Hall and Piexe, can be stated as follows: if $\tilde{S}$ is the matrix that minimizes $CCIC(S_{k_2})$, under the assumptions of identifiability and normal i.i.d. observations $[(Y_t', X_t', u_t) \overset{i.i.d.}{\sim} N(0, \Sigma)]$, and some other assumptions [see Hall and Peixe (2000)], then $\tilde{S} \underset{T \longrightarrow \infty}{\xrightarrow{p}} S_0$, where S_0 is the matrix that selects the true relevant instruments (those verifying Definition **3.1**).

To evaluate the finite sample properties of their method, Hall and Piexe performed a simulation study using a model with a small set of 8 candidate instruments. Instead of considering the $2^8 - 1 = 255$ possible combinations, they consider only 4 pairs of instruments which give 15 possible combinations. The results show that the selection

procedure works well for a size of $T = 500$ observations. For $T = 100$, the method works quite well when $f(T) = ln(T)$ but not with the two other criteria, it tends to include more instruments than is needed.

When the number of potential instruments is large (which is often the case in empirical work) the method becomes cumbersome. With a number of l instruments, there is $2^l - 1$ possible choices, and for each choice we have to calculate all the canonical correlations.[4] The extension of this approach to non-normal and/or nonlinear but still i.i.d. models goes through adding some technical assumptions and replacing (for nonlinear models) Y_t by $d_t(\theta) = \frac{\partial u_t(\theta)}{\partial \theta}$ evaluated at some consistent estimator of θ.

Finally, it is worth noting that this kind of method selects instruments on the basis of a criterion that does not clearly aim to improve a specific quality of an estimator or a test statistic. For example it would be interesting to see how the estimates obtained using this selection method perform relatively to those using all the potential instruments. As is the case with Hall, Rudebusch, and Wilcox (1996) method, it is not sure that the estimates will be better. This will be the focus of Section 7 of this paper.

3.2.2.2. Donald and Newey (2000). Donald and Newey (Forthcoming) considered the problem of choosing the number of instruments. Their results are derived in the context of a model with only one endogenous explanatory variable and homoskedastic disturbances, but they allow the reduced form relationship between the endogenous explanatory variable and the exogenous variables to be nonparametric. The model is given by

$$y_t = \beta Y_t + \gamma x_{1t} + u_t$$

and

$$Y_t = f(x_t) + \varepsilon_t \quad .$$

where $x_t = [x'_{1t}, x'_{2t}]'$ is a vector of exogenous variables, y_t and Y_t are scalars, $E[u_t|x_t] =$

[4] If we have G endogenous variables, we need at least G instruments and the the number of choices is $2^l - \sum_{i=1}^{G-1} C_l^i$, where C_l^i is the number of subsets of i elements chosen from a set of l elements.

$E[\varepsilon_t|x_t] = 0$, $var[u_t|x_t] = \sigma_u^2$ and $var[\varepsilon_t|x_t] = \sigma_\varepsilon^2$.

Given a set of potential instruments $Z = [z_1, z_2, ..., z_l]$, the problem considered in the paper is how to choose a number K ($K \leq l$) such that the subset $Z_{(K)} = [z_1, z_2, ..., z_K]$ will give the best estimator of β within a given class of estimators. Three estimators were considered: 2SLS, LIML and the Jackknife IV estimator [JIVE, see Angrist, Imbens, and Krueger (1999)]. The term best means the smallest MSE. To choose K, they propose to minimize a criterion S based on the asymptotic MSE. For 2SLS, LIML, and JIVE, the criteria to minimize are given respectively by

$$\hat{S}_{\text{2SLS}}(K) = \hat{\sigma}_{u\varepsilon}^2 \frac{K^2}{N} + \hat{\sigma}_u^2 [\hat{R}(K) - \hat{\sigma}_\varepsilon^2 \frac{K}{N}],$$

$$\hat{S}_{\text{LIML}}(K) = \hat{\sigma}_u^2 [\hat{R}(K) - \frac{\hat{\sigma}_{u\varepsilon}^2}{\hat{\sigma}_u^2} \frac{K}{N}],$$

and

$$\hat{S}_{\text{JIVE}}(K) = \hat{\sigma}_u^2 [\hat{R}(K) + \frac{\hat{\sigma}_{u\varepsilon}^2}{\hat{\sigma}_u^2} \frac{K}{N}]$$

where $\hat{\sigma}_{u\varepsilon}$, $\hat{\sigma}_u^2$, and $\hat{\sigma}_\varepsilon^2$ are some estimators obtained from preliminary estimates of the residuals u and ε that do not depend on K, and are given by[5]

$$\hat{\sigma}_{u\varepsilon} = \frac{\hat{u}'\hat{\varepsilon}}{T},\ \hat{\sigma}_u^2 = \frac{\hat{u}'\hat{u}}{T},\ \hat{\sigma}_\varepsilon^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{T}\ .$$

$\hat{R}(K)$ is a goodness of fit criterion that may be either the Mallows goodness of fit criterion

$$\hat{R}(K) = \frac{\hat{\varepsilon}^{K\prime}\hat{\varepsilon}^K}{T}[1 + \frac{2K}{T}],$$

or the cross validation goodness of fit

$$\hat{R}(K) = \frac{1}{T}\sum_{t=1}^{T} \frac{(\hat{\varepsilon}_t^K)^2}{(1 - P_{tt}^K)^2}$$

where $\hat{\varepsilon}^K$ is the reduced form residual vector from regressing Y on $Z_{(K)}$ and P_{tt}^K

[5] $\hat{\varepsilon}$ may be obtained by regressing Y on the set of all instruments and $\hat{u}$ may be obtained from an IV estimation using all the potential instruments.

are the diagonal elements of the matrix $P^K = Z_{(K)}(Z'_{(K)}Z_{(K)})^{-1}Z'_{(K)}$. Choosing K to minimize the criterion S will result in the best asymptotic MSE. The optimal (asymptotically) subset of instruments contains the first $\hat{K}$ instruments of Z where:[6]

$$\hat{K} = \arg\min_{K}(\hat{S}(K)).$$

The optimality is defined in the sense that

$$\frac{S(\hat{K})}{\inf_K S(K)} \xrightarrow[T\longrightarrow\infty]{p} 1. \tag{3.15}$$

Despite the sophisticated tools used in showing the asymptotic optimality results of these criteria, the Monte Carlo simulations conducted by the authors show no convincing results even for a sample size of 400 observations. An improvement is found for the coverage rate of 2SLS-based confidence sets (the estimate plus or minus 1.96 times the asymptotic standard error), but for LIML, the results with the selection procedure are worse than those using the hole set of instruments (see Tables 1.1, 2.1 and 3.1 of the their paper). On the other hand, a serious drawback of the method is that it depends on the initial order of the instruments in $Z = [z_1, z_2, ..., z_l]$. For example if the first 4 instruments z_1, z_2, z_3, z_4 are irrelevant but z_5 is relevant, then the procedure will tend to choose $\hat{K} = 5$ and all the 4 irrelevant instruments will be included in the "optimal" matrix of instruments (see Table 2.2 of their paper).

The main problem we consider in the sequel is the following one: given a set of l potential instruments, how does one determine which subset gives the best quality of estimation and inference. If all the instruments are weak, estimation makes no real sense but we can still construct valid confidence sets using the Anderson-Rubin procedure. In such cases, valid confidence sets should have a large probability of being unbounded or uninformative [see Dufour (1997)]. We begin by examining how the number of instruments and the concentration parameter influence the bias and

[6]The optimality was proven under the additional assumption that the preliminary estimates $\hat{\sigma}^2_{u\varepsilon}$, $\hat{\sigma}^2_{\varepsilon}$, and $\hat{\sigma}^2_{u}$ are consistent and other assumptions including the existence of the 8^{th} conditional moment of ε_t.

MSE of 2SLS, as well as the power of AR tests. Then we consider the selection of instruments on the basis of these observations. We focus on the case where there is only one endogenous explanatory variable. The generalization to the case of multiple endogenous explanatory variables will be studied in future work.

4. Number of instruments, concentration parameter and inference quality

The relation between the number of instruments, k_2, the concentration parameter, δ, and the bias of 2SLS has been mentioned in many papers [see Mariano (1982) for a survey]. The absolute value of the bias increases as k_2 increases and δ remains fixed.[7] On the other hand, it is a decreasing function of δ when k_2 remains fixed. It is also worth noting that the bias depends in a simple way on the endogeneity of the explanatory variable Y. Setting $\sigma_{u2} = cov(u_t, U_{2t})$ and noting that $\sigma_{u2} = \sigma_{12} - \beta\sigma_{22}$, we see from (3.3) that the absolute value of the bias is

$$|B(\beta)| = \left|\frac{\sigma_{u2}}{\sigma_{22}}\right| e^{\frac{-T\delta}{2}} {}_1F_1(\frac{k_2}{2} - 1, \frac{k_2}{2}, \frac{T\delta}{2}).$$

Consequently it is an increasing function of $|\sigma_{u2}|$. This confirms the approximation of Hahn and Hausman (Forthcoming, page 6). For the MSE, the only result available in the literature is its monotonicity with respect to δ when k_2 is held fixed. For the AR test, monotonicity with respect to the concentration parameter is established for $G \geq 1$ in Dufour and Taamouti (2000a). For convenience, we summarize the results in the following proposition. Although this proposition can be derived from results available in the literature, it is not stated in this form elsewhere (to the best of our knowledge). A proof is provided in the Appendix.

Proposition 4.1 MONOTONOCITY. *Under Assumptions* (2.4), (2.5), *and* (3.2), *the*

[7] If δ remains constant while k_2 increases, this does not mean necessarily that all the added instruments have a zero-coefficient in the reduced form equation but may be seen as a change in the set of instruments in a way that the new set contains more instruments but with the explanatory power measured by δ remaining constant.

following results hold:

(i) if k_2 remains fixed, an increase in the concentration parameter δ decreases the absolute bias and the MSE of the 2SLS estimator and increases the power of AR test at each point in the alternative;

(ii) if δ remains fixed, an increase in the number of instruments k_2 increases the absolute bias of 2SLS.

The monotonicity of the MSE with respect to k_2 is not clear analytically. Even asymptotically the approximate analytic expression of the MSE is an indefinite function of k_2 [see Chao and Swanson (2000)]. Given that a simple analytical relation does not appear to hold, we shall present some simulation evidence, which suggests (in accordance with intuition) that MSE increases when we add totally irrelevant instruments (which means a fixed concentration parameter). Of course such simulation evidence should be taken with the usual caveats.

Consider the following data generating process :

$$y_1 = \beta y_2 + u_1,$$

$$y_2 = \sum_{i=1}^{k_2} x_i \pi_{2i} + u_2,$$

where $X_2 = [x_1, x_2, ..., x_{k_2}]$ is the matrix of instruments. We consider two cases for X_2. In case 1 only the first 2 instruments are relevant ($\pi_{21} = \pi_{22} = \frac{1}{2}$, and $\pi_{2i} = 0$ for $i > 2$), and in case 2, only the first 8 instruments are relevant ($\pi_{21} = \pi_{22} = ... = \pi_{28} = \frac{1}{2}$, and $\pi_{2i} = 0$ for $i > 8$). The total number of instruments varies from 1 to 40 in each case. The disturbances u_1 and u_2 were generated from a bivariate normal distribution, $(u_{1t}, u_{2t}) \sim N(0, \Omega)$ where

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}.$$

The parameter β is set equal to 1, the size of the sample is taken to be $T = 50, 100,$ or 200 and the model is replicated 20000 times for each experiment. The results are

presented in Table 1 and Figure 1.

When the number of relevant instruments is 8 (case 2) the MSE decreases as we add instruments but after the 8-th relevant one is added it clearly begins to increase.[8] In case 1, the MSE (which exists if the number of instruments is greater than 2) increases as we add instruments. Notice also that MSE clearly decreases when T increases but this is not surprising [see Owen (1976)]. A special case occurs when $cov(u_1, u_2) = 0$, there is no endogeneity. In this case the MSE is a decreasing function of k_2 (Table 2 and Figure 2). This is not surprising since OLS is the best linear unbiased estimator (BLUE) and a large set of instruments makes 2SLS near OLS. This remark is also mentioned on the basis of an asymptotic approximation by Donald and Newey (Forthcoming, page 18).

On the power of the AR test, an extensive simulation study is presented in Dufour and Taamouti (2000a). The results show that adding more instruments while holding the concentration parameter fixed decreases the power of the AR test. Similar evidence was also obtained in another Monte Carlo experiment conducted by Kleibergen (2001a).

The main conclusions on these issues are the following: both for estimation by 2SLS and for inference using the AR test, the set of instruments influences the quality of estimation and inference through two parameters: the number of instruments and the concentration parameter. An increase in the number of instruments holding δ fixed increases the absolute bias and the MSE and decreases the power of the AR test while an increase in δ, holding k_2 fixed, decreases the absolute bias and the MSE and increases the power of the AR test. A "good" set of instruments should not include irrelevant instruments and the associated concentration parameter δ should be as large as possible. Unfortunately, the two directions are opposite since adding an instrument increases k_2 but at the same time can never decrease δ. The following table summarizes these results.

[8]Note that in our design, the MSE exists when $k_2 \geq 3$, that's why we do not report MSE for $k_2 = 1, 2$.

	Overidentification: k_2	Concentration parameter: δ
|Bias| of 2SLS	↗	↘
MSE of 2SLS	↗	↘
Power of AR test	↘	↗

5. Ordering and selecting instruments

In this section, we propose on the basis of the previous results a method of selecting instruments that aims at improving the quality of estimation and inference. First, the concentration parameter $\delta = \Pi_2' X_2' M(X_1) X_2 \Pi_2 / (T\sigma_{22})$ is unknown in practice while σ_{22} is independent of the selected set of instruments. When comparing two sets of instruments on the basis of δ only the numerator matters, therefore we replace σ_{22} by $Y'M(X_1)Y/T$. To estimate the numerator, a natural method consists in applying OLS to the reduced form equation (2.2) to estimate Π_2 :

$$\hat{\Pi}_2 = [X_2' M(X_1) X_2]^{-1} X_2' M(X_1) Y. \tag{5.1}$$

On replacing Π_2 and σ_{22} by their estimators, we get

$$\hat{\delta} = \frac{Y'M(X_1)X_2[X_2'M(X_1)X_2]^{-1}X_2'M(X_1)Y}{Y'M(X_1)Y}$$

or

$$\hat{\delta} = \frac{Y'P[M(X_1)X_2]Y}{Y'M(X_1)Y} \tag{5.2}$$

$\hat{\delta}$ is what is called the partial R^2 in Bound, Jaeger, and Baker (1995) paper. When there is no exogenous variable included in equation (2.1), (5.2) turns to be the usual non-central R^2 from regressing Y on X_2, so maximizing the estimated concentration parameter is equivalent to maximizing the non-central R^2. It is also worth noting that maximizing $\hat{\delta}$ and minimizing k_2 is equivalent to maximizing the first stage F-statistic for $H_0 : \Pi_2 = 0$ in the regression (2.2). This statistic is given by

$$F = \frac{Y'[M(X_1) - M(\bar{X})]Y/k_2}{Y'M(\bar{X})Y/(T - k_1 - k_2)}$$

where $\bar{X} = [X_1, X_2]$, and can be written as

$$F = \frac{T - k_1 - k_2}{k_2} \frac{\hat{\delta}}{1 - \hat{\delta}}. \tag{5.3}$$

Hence F is an increasing function of $\hat{\delta}$ when k_2 is fixed and a decreasing function of k_2 when $\hat{\delta}$ is held fixed.

Second, the selected set should contain a number of instruments at least equal to the number of endogenous explanatory variables to ensure the necessary order condition for identification. Since we consider the case of $G = 1$, the set will contain at least one instrument.

One possible procedure is to choose for each number of instruments k_2 ($1 \leq k_2 \leq l$) the set with the highest estimated concentration parameter among all subsets of X with k_2 instruments. We obtain l potential candidate sets. Each one of these sets is optimal among the sets with the same number of instruments. The next step is to eliminate all sets for which there is another set with smaller number of instruments but with a higher or equal concentration parameter. The remaining candidates are all acceptable, hence the following definition.

Definition 5.1 *A subset of instruments is said to be acceptable if there is no other subset of columns of X with same (or smaller) dimension such that $\hat{\delta}$ is higher.*

A drawback of such a procedure is the computational burden. With l potential instruments, one has to calculate $2^l - 1$ concentration parameters. This is a huge number given that in empirical work the set of candidate instruments can be quite large. For example in Angrist and Krueger (1991), 180 instruments are considered, while in the empirical example of Hahn and Hausman (Forthcoming), this number is 136. Applying the above procedure in the case of Hahn-Hausman paper is equivalent to choose a subset among more than 8.7×10^{40} candidate sets, clearly a huge number. This remark is also valid for the procedure proposed by Hall and Piexe (2000).

The main idea we introduce to circumvent this drawback is to proceed sequentially. We proceed in two main steps: we first order instruments on the basis of $\hat{\delta}$ and we then

apply an upward testing procedure to select a subset of instruments. Hereafter we will refer to this method as the **upward maximization of concentration (UMC).** The ordering of instruments is obtained through the following steps:

(1) choose the instrument that gives the highest $\hat{\delta}$ if it is used alone in (2.2) with X_1. Suppose that the instrument corresponds to the column i_1 of X noted x_{i_1};

(2) choose among the remaining $l-1$ instruments the one which, when added to x_{i_1} provides the highest $\hat{\delta}$;

(3) choose among the remaining $l-2$ instruments the one which, when added to $[x_{i_1}, x_{i_2}]$ provides the highest $\hat{\delta}$;

(4) Continuing so forth gives an order or priority of including instruments, and yields a matrix of ordered instruments $\tilde{X} = [\tilde{x}_{(1)}, \tilde{x}_{(2)}, ..., \tilde{x}_{(l)}]$ where $\tilde{x}_{(i)}$ is the column selected at the i-th step.

Now, to select a subset of instruments, we suggest to use an upward testing procedure. Begin by selecting $\tilde{x}_{(1)}$. For the following instrument $\tilde{x}_{(2)}$ to be added, its coefficient should be significantly different from zero in the regression of $M(X_1)Y$ on $M(X_1)\tilde{x}_{(1)}$ and $M(X_1)\tilde{x}_{(2)}$. If $\tilde{x}_{(2)}$ is added we do the same with $\tilde{x}_{(3)}$, and so forth. More generally if a subset of i columns X_2^1 is selected, then for the column X_2^2 to be added we test the redundancy of X_2^2 given X_2^1 [see Breusch and Wyhowski (1999)]. The underlying regression for the F-test is:

$$Y = X_1\phi_1 + X_2^1\lambda + \mu X_2^2 + \varepsilon.$$

The F-test for $H_0 : \mu = 0$ is given by

$$F = \frac{Y'M(X_1)\left[M[M(X_1)X_2^1] - M[M(X_1)[X_2^1, X_2^2]]\right]M(X_1)Y}{Y'M(X_1)M[M(X_1)[X_2^1, X_2^2]]M(X_1)Y/(T-k_1-i-1)}.$$

If α is the level of the test and $c(\alpha)$ is the critical point, then X_2^2 is added if $F \geq c(\alpha)$.

This is equivalent to

$$\frac{S^{\perp}(X_1) - S\big[M(X_1)[X_2^1, X_2^2]\big]}{S^{\perp}(X_1) - S\big[M(X_1)X_2^1\big]} \leq \frac{T - k_1 - i - 1}{T - k_1 - i + (c(\alpha) - 1)} \tag{5.4}$$

where for any matrix A with T rows, $S(A) = Y'P(A)Y$ and $S^{\perp}(A) = Y'M(A)Y$. We stop the procedure as soon as the F-test does not reject the hypothesis of zero-coefficient. Note that an instrument that has been selected may get a zero-coefficient when a new instrument is added: in this case, the former should be removed if the new instrument is not selected. This may happen, for example, if the best subset of instruments is $[x_1, x_2]$ but there is another instrument x_3 which has a higher correlation with Y than both of x_1 and x_2. Note however that this situation cannot happen if the columns of X are orthogonal.

The method has many advantages, it rules out any collinearity between instruments. If two instruments are collinear only one will be included, and when the number of instruments is large relatively to the sample size, it allows one to get quite precise evaluation since it does not consider the hole set of instruments at the same time. An important issue raised by this kind of procedure is the multiple testing problem. The level of the second and so forth F-tests may be distorted. However, this is a finite sample issue, in large samples it is easy to circumvent such a drawback by appropriately choosing the critical values for the F-tests (see Section 6).

Several other intuitive stopping rules may be adopted.

(i) We can use a rule similar to the adjusted R^2. Like R^2, $\hat{\delta}$ never decreases when we add an instrument. To penalize the increase in the number of instruments, we define:

$$\bar{\delta} = 1 - \frac{T - k_1}{T - k_1 - k_2}(1 - \delta). \tag{5.5}$$

Unlike δ, $\bar{\delta}$ may decrease if the addition of an instrument does not induce a sufficient increase in δ. $\bar{\delta}$ may be estimated by

$$\tilde{\delta} = 1 - \frac{T - k_1}{T - k_1 - k_2}\left(1 - \frac{Y'P[M(X_1)X_2]Y}{Y'M(X_1)Y}\right) \tag{5.6}$$

A set X_2^2 of k_2^2 instruments will increase the value of this measure relatively to a set X_2^1 of k_2^1 instruments iff:

$$\frac{S^{\perp}(X_1) - S[M(X_1)X_2^2]}{S^{\perp}(X_1) - S[M(X_1)X_2^1]} \leq \frac{T - k_1 - k_2^2}{T - k_1 - k_2^1}. \tag{5.7}$$

Notice that this rule is a special case of (5.4), it corresponds to the case where $c(\alpha) = 1$. This is similar to the relation between the adjusted R^2 and the t-statistic in linear regression models.

We can also correct $\hat{\delta}$ by

$$\tilde{\delta} = \frac{T - k_1}{k_2}\hat{\delta} = \frac{Y'P[M(X_1)X_2]Y/k_2}{Y'M(X_1)Y/(T - k_1)}. \tag{5.8}$$

In this case the value of $\tilde{\delta}$ increases if

$$\frac{S[M(X_1)X_2^2]}{S[M(X_1)X_2^1]} \geq \frac{k_2^2}{k_2^1}. \tag{5.9}$$

This condition is similar to the asymptotic result of Buse (1992) who showed that the bias decreases if

$$\frac{S[M(X_1)X_2^2]}{S[M(X_1)X_2^1]} \geq \frac{k_2^2 - 1}{k_2^1 - 1}, \tag{5.10}$$

provided that $k_2^1 \geq 2$ to ensure the existence of the first moment of 2SLS. Notice that Buse's condition which is derived without the normality assumption is more requiring for an instrument to be added ($\frac{k_2^2-1}{k_2^1-1} \geq \frac{k_2^2}{k_2^1}$ when $k_2^2 \geq k_2^1$), so it will tend to select less instruments than (5.9).

(*ii*) Given that the number of instruments and the concentration parameter have different effects, we may select on the basis of a selection criterion (to minimize) that decreases when δ increases and that penalizes for the addition of instruments. This is what is done in Hall and Peixe (2000) (see the review in Section 3.2). In the case of one endogenous explanatory variable, the canonical correlation is simply the R^2. Note however that this procedure will be cumbersome to apply in practice since the number of candidate sets is $2^l - 1$.

(*iii*) The procedure of ordering instruments may also be used as a first step before applying Donald and Newey (Forthcoming) selection method. The combination of these two procedures may give better results than Donald-Newey alone.

When the columns of X are orthogonal, whatever the stopping rule is, it is easy to see that the selected subset will always be acceptable, i.e. there will be no subset of X with greater concentration parameter and smaller or equal number of instruments.

Proposition 5.2 ACCEPTABILITY. *Let $\tilde{X}$ be the matrix of ordered instruments. If the columns of $[X_1, X]$ are orthogonal then for any i, $(1 \leq i \leq l)$, the subset $[\tilde{x}_{(1)}, \tilde{x}_{(2)}, ..., \tilde{x}_{(i)}]$ containing the first i columns of $\tilde{X}$ is acceptable.*

In general, the instruments are not orthogonal. However, there is many ways to transform them in orthogonal ones while keeping the total explanatory power of X. Particularly two methods may be used:

- The principal components decomposition [see Anderson (1984)] which consists in replacing X by a matrix X_p whose columns are the eigenvectors of the matrix $X'X$.

- The Gram-Schmidt orthogonalization, transforming $X = [x_1, ..., x_l]$ into $X^* = [x_1, M(x_1)x_2, M[x_1, x_2]x_3, ..., M[x_1, x_2, ..., x_{l-1}]x_l]$. The number of possible orthogonalization by this method is the same as the number of permutations of the columns of X, which is $n!$.

In empirical work, the instruments are chosen on the basis of theoretical considerations. If we transform them according to one of the above methods, what we get is not necessarily random variables with theoretical meaning and as instruments they are not necessarily better in terms of relevance. The Monte Carlo simulations of Section 7 show that the Gram-Schmidt orthogonalization works as well as using the initial set of instruments while the principal decomposition seems to have a negative impact on estimation and inference.

6. Consistency of the UMC method

The finite sample performance of the proposed selection method will be studied via Monte Carlo simulations in Section 7. In this section we study its behavior when the sample size T gets large while the number of instruments remains finite. For comparison purposes, it is worth recalling the desirable asymptotic properties of a selection method considered in the literature. In Hall and Peixe (2000) paper, a selection method is consistent if it selects the instruments verifying Definition **3.1**. In Donald and Newey (Forthcoming) paper, the desirable property is higher order asymptotic optimality in the sense of (3.15). Here, we define consistency relatively to the true data generating process of the endogenous explanatory variables Y. Assume that the true DGP of Y is:

$$Y = X_2^* \Pi_2^* + X_1 \Pi_1 + v_2 \tag{6.1}$$

where Π_2^* is a vector of non-zero coefficients. Then, a selection method is consistent if asymptotically it selects X_2^* with probability 1.

In this section we show that the UMC method is (under standard assumptions) consistent. We assume that the instruments X_2 are chosen from a large set X containing the columns of X_2^* and verifying the following assumptions:

$$\text{Assumption}(i) : \frac{x_i' x_j}{T} \xrightarrow[T \longrightarrow \infty]{p} 0 \text{ for } i \neq j \tag{6.2}$$

where x_i, $i = 1, \ldots, l$, are the columns of the matrix $[X_1, X]$. This assumption means that the columns of $[X_1, X]$ are asymptotically orthogonal. This is a weaker assumption than assuming the orthogonality of the columns of $[X_1, X]$.

$$\textbf{Assumption}(ii) : \left(\frac{[X_1, X]'[X_1, X]}{T}, \frac{[X_1, X]'u}{T}, \frac{u'u}{T}\right) \xrightarrow[T \longrightarrow \infty]{p} (Q, 0, \sigma^2) \tag{6.3}$$

where Q is a diagonal positive definite matrix.

$$\text{Assumption}(iii): \frac{[X_1, X]'u}{\sqrt{T}} \xrightarrow[T\longrightarrow\infty]{d} N(0, \sigma^2 Q).$$

This is a standard assumption in asymptotic theory. Here, it will allow the F-statistic to have a χ^2 distribution asymptotically. Finally, the critical values for the F-tests in the upward testing procedure are assumed to verify the following conditions:

$$\text{Assumption}(iv): c_{i,T} \xrightarrow[T\longrightarrow\infty]{p} +\infty, \text{ and } \frac{c_{i,T}}{T} \xrightarrow[T\longrightarrow\infty]{p} 0, \; i = 2, ..., l \tag{6.4}$$

where $c_{i,T}$ is the critical value for the inclusion of the i-th instrument when $i-1$ have been selected. This assumption follows from Potscher (1983) who showed that if $\alpha_{k,T} \in]0,1[$ and $c_{k,T}$ is the corresponding upper quantile of a chi square distribution, then $\frac{c_{k,T}}{T} \xrightarrow[T\longrightarrow\infty]{p} 0$ if and only if $\frac{\ln(\alpha_{k,T})}{T} \xrightarrow[T\longrightarrow\infty]{p} 0$. In particular this will hold if we take $\alpha_{k,T} = \exp[\lambda_T T]$ where λ_T is such that $\lambda_T \xrightarrow[T\longrightarrow\infty]{p} 0$ as $T \longrightarrow +\infty$.

The following proposition shows that asymptotically the UMC procedure based on F-tests proposed above will select the instruments X_2^* with probability 1 and so is consistent.

Proposition 6.1 CONSISTENCY OF UMC METHOD. *If the data generating process of Y is (6.1), under Assumptions $(i)-(iv)$, the upward selection procedure using the F-test selects X_2^* with probability one asymptotically.*

Notice that this proposition does not require Assumption (2.5), it holds under Assumptions $(i)-(iv)$ which are standard in asymptotic theory.

7. A simulation study

In this section we present a simulation study of the performance of the selection method presented in this paper. We evaluate how much good is the method in selecting the "right" instruments (those with a non-zero coefficient in the first stage equation) then we address the main question of whether selection procedures improve

estimation and inference. We also consider the simulation design of Hall, Rudebusch, and Wilcox (1996) to verify whether the problem of relevance testing can be solved by using a split-sample technique. We consider the following DGP:

$$y_{1t} = \beta y_{2t} + u_t \tag{7.1}$$

$$y_{2t} = X_t'\pi_2 + v_{2t} \tag{7.2}$$

(u_t, v_{2t}) are generated from an *i.i.d.* $N(0, \Sigma)$ process with $\Sigma = \begin{pmatrix} 1 & .25 \\ .25 & 1 \end{pmatrix}$, and β is taken equal to 1. The matrix of instruments X contains either $l = 10$ or $l = 20$ instruments, but in both cases only the fourth and fifth instruments are relevant ($\pi_2[1] = \pi_2[2] = \pi_2[3] = 0$, $\pi_2[4] = \pi_2[5] = 1$ and $\pi_2[i] = 0$, for $i > 5$). For the sample size we consider four different sizes $T = 20$, 50, 100, and 200. Finally, the number of replications is set at 10000 for each experiment.

7.1. Performance of the UMC method

Since only two instruments are relevant, and others make no contribution to the explanation of the variation in the endogenous explanatory variable y_2, it is natural to expect that only these two instruments will be chosen. Table 3 presents the frequency of choosing each one of the instruments for $l = 10$ and $l = 20$ and for different sizes of the sample. When $T \geq 50$, the method selects the "right" instruments with probability 1. When $T = 20$, this probability is around 80% even with a number of instruments equal to the number of observations ($l = T = 20$). On the other hand, the probability of selecting any one of the irrelevant instruments is 0 when $T \geq 50$ and is often around 1% when $T = 20$.

Table 4 presents the frequencies of choosing the "right" number of instruments (which is 2). When $T \geq 50$ only the two relevant instruments are selected. When $T = 20$, the probability of selecting only 1 instrument instead of 2 is about 20% and there is about 10% chance to select more than 2 instruments but the probability that the selected number exceeds 3 instruments is nearly 0.

The method seems working well in selecting the relevant instruments, however what matters is wether it improves inference and estimation. This is the subject of Tables $5-8$. In each of these tables we consider 4 strategies:

(1) making estimation and inference with all the available instruments without any pre-selection procedure;

(2) selecting instruments among the columns of X and then doing estimation and inference;

(3) transforming the matrix of instruments X into a new matrix X_p corresponding to the principal components of X and then selecting instruments among the columns of X_p and doing estimation and inference;

(4) transforming the matrix of instruments X into a new matrix X_{GS} corresponding to the Gram-Schmidt orthogonalization of X and then selecting instruments among the columns of X_{GS} and doing estimation and inference.

For the last 3 strategies which involve a selection procedure, we consider two options, selecting instruments using all the observations, and selecting using only a part of the sample. This second option is necessary to control the size of the AR test. The fraction of the sample used for selection is either 20% or 30%. For the upward F-tests, the quantiles are chosen according to the rule suggested by Potscher (1983): $\alpha_{i,T} = \exp[-\sqrt{T}]$, $T = 50,\ 100,\ 200$.

In Table 5 we present the results for the bias of 2SLS.[9] The selection procedure seems to reduce the bias in a substantial way, whatever the sample size, and the fraction of the sample used for selecting instruments. This an interesting result since the bias is the most reported problem of estimation in the literature and many papers concentrate only on the improvement of this feature of estimation. In fact the bias is very sensitive to the number of irrelevant instruments, this is the reason why the improvement is huge when we use a small number of selected instruments even if the sample size is reduced with the split-sample technique.

[9]The results are multiplied by 100 to get them easier to read and to compare.

For the mean squared error (Table 6), there is no improvement compared to using all the potential instruments, the MSE with the selection procedure are slightly greater than those using all the instruments. This is due in part to the decrease in the sample size from T to $T-s$ where s is the fraction used for selecting instruments. When all the observations are used for both selection and estimation, the MSE is not reported for $T \geq 50$ given that in this case only two instruments are selected and the MSE is not finite. In Table 7 we report another measure of dispersion, it is the mean absolute error (MAE).[10] The selection procedures seem to reduce slightly the MAE when $s = T$, even if there is no improvement when a part of the sample is exclusively used for selection. If we compare the three strategies based on a preselection procedure, the third one based on principal components decomposition performs poorly relatively to the two others.

For the AR test, the results are presented in Table 8 and Figure 3, each of these tables presents a combination of the sample size T ($T = 100, 200$), the total number of instruments l ($l = 10, 20$) and the fraction of the sample used for selection s, ($s = .2 \times T$ or $s = .3 \times T$). This split-sample technique allows to control for the size of the test as can be clearly seen in all these tables. Again, when strategies (2) or (4) are used the improvement in the power of the AR test seems substantial when we choose a good fraction s of the sample for the selection procedure. This improvement is larger as the total number of instruments goes from 10 to 20. On the other hand while the strategy of selecting directly from the initial set of instruments (strategy (2)) and Gram-Schmidt orthogonalization work very well (they give nearly the same results), the principal components decomposition seems not to give good results that's why, for comparison with other methods in the literature we will consider only strategy (2).

7.2. Comparison with other methods

We consider now our strategy (2) which we compare to Hall and Peixe (2000) and Donald and Newey (Forthcoming) methods presented in Section 3.2. For Hall and

[10] Like the bias, the MAE exists if the degree of over-identification is greater than or equal to 1.

Peixe (2000), we choose for comparison the Schawrz-type criterion which is their best one. Tables 9 to 11 present the results for the bias, MSE and MAE of 2SLS respectively and Table 12 presents the results for the AR test.

The most important improvement is obtained for the bias, all the 3 selection methods induce a substantial reduction in the bias. Our method gives the best results, and the Hall-Piexe method seems slightly better than Donald-Newey's method especially when the total number of instruments is 10. In terms of MSE and MAE, the 3 methods give similar results but they do not seem to give a clear improvement relatively to using all the instruments.

In terms of the power of the AR test (Table 12 and Figure 4), all the three methods give a significant improvement in the power while the size of the test is hold under control. Again among the three selection methods, ours seems to give slightly better results (except in case (e)).

To see how the results change in the case of non-normal disturbances, we conduct another simulation where the disturbances were generated from a student distribution with 5 degrees of freedom. The results (with $s = T$) are presented in Table 13, and they are similar to those obtained with normal disturbances.

7.3. Hall, Rudebusch and Wilcox (1996) Monte Carlo design

To verify whether the poor performance of the t-test after selection of instruments reported in Hall, Rudebusch, and Wilcox (1996) [see equation (3.10)] is due to the fact that higher relevance is associated with serious endogeneity, we try here the split-sample technique [Dufour and Jasiak (1993), Angrist and Krueger (1995)] as a method to circumvent this problem. The Monte Carlo design in this section is the same as the one considered in Hall, Rudebusch, and Wilcox (1996).The DGP is given by:

$$y_t = x_t\beta + u_t,\ t = 1, \ldots T.$$

The disturbance, u_t, the regressor, x_t, and the instrument, z_t, are generated as follows:

$$(u_t, x_t, z_t) \sim N(0, \Omega),\ \Omega = \begin{pmatrix} 1 & . & . \\ \sigma_{xu} & 1 & . \\ 0 & \sigma_{xz} & 1 \end{pmatrix}$$

the parameter β is set at $\beta_0 = 0$. the IV estimator of β is

$$\tilde{\beta} = \frac{z'y}{z'x}.$$

The quality of inference with $\tilde{\beta}$ is examined through the empirical level of the t-statistic of the null hypothesis $H_0 : \beta = 0$. The results are based on 10000 replications and the sample size is $T = 100$. The t-statistic is computed only for samples where the LR test rejects the non relevance hypothesis. We adopt two strategies, the first consist in using all the observations for both the relevance test and the t-test, and the second consists in dividing the sample of observations into 2 parts, the first part is used for relevance testing and the second part is used to compute the t-test when the instruments are judged relevant. The results are presented in Table 14. The first and second column give $cov(x, z)$ and $cov(x, u)$, the third column gives the empirical probability of rejection of non-relevance hypothesis by the test (3.10), and the fourth column gives the empirical size of the t-test when the non-relevance hypothesis is rejected. Columns (5)-(6) and (7)-(8) give the same as columns (3)-(4) when the split-sample technique is used. The number of observations used for relevance testing is either 20 or 30. We find that level distortions are quite smaller when the split-sample technique is used but remains serious. Thus, in contrast to the conclusion of Hall, Rudebusch, and Wilcox (1996), the problem of screening tests may be due to other reasons than endogeneity when the instruments are judged relevant.

8. Application to the problem of returns to scale in US industry

One of the widely studied problems in recent macroeconomics literature is the extent of returns to scale in the U.S. industry. Recent work on this issue includes Hall (1990), Caballero and Lyons (1989, 1992), Basu and Fernald (1995, 1997) and Burnside (1996). The results of these researches have important implications on many fields of macroeconomics, such as growth and business cycle models.

Burnside (1996) presents a short survey of different specifications of the production function adopted in this literature. One of these specifications considers the following equation

$$Y_{it} = F(K_{it}, L_{it}, E_{it}, M_{it}) \tag{8.1}$$

where, for each industry i and each period t, Y_{it} is the gross output, K_{it} is the amount of capital services used, L_{it} is the amount of labor, E_{it} is energy used, and M_{it} is the quantity of materials. If we assume that F is a differentiable function and homogeneous of degree ρ, we get the following regression equation [see Burnside (1996)]:

$$\Delta y_{it} = \rho \Delta x_{it} + \Delta a_{it} \tag{8.2}$$

where Δy_{it} is the growth rate of the output, Δx_{it} is a weighted average of the inputs (The weights are the production cost shares of each input) and Δa_{it} represents technological changes. In this specification, ρ is the coefficient that measures the extent of returns to scale. Returns to scale are increasing, constant or decreasing depending on whether $\rho > 1$, $\rho = 1$ or $\rho < 1$.

To estimate this equation, Hall (1990) proposed a set of instruments that was used in most subsequent researches. These instruments include the growth rate of military purchases, the growth rate of world oil price, a dummy variable representing the political party of the president of Unites States and one lag of each of these variables. Estimation methods used include ordinary least squares, two stages least squares and three stages least squares.

The regressions are performed using panel data on two-digit SIC (Standard Industrial Classification) code level manufacturing industries. This classification includes 21 industries. The data set is described in detail by Jorgenson, Gollop, and Fraumeni (1987) and contains information on gross output, labor input, stock of capital, energy use, and materials inputs. As reported in Caballero and Lyons (1989), there is no evidence of serial correlation in this data from either the Durbin-Watson statistic or the Ljung-Box Q statistic. The instruments above have been studied in detail by Burnside (1996) who showed on the basis of calculations of R^2 and partial R^2 [Shea (1997)], that these instruments are weak. Given the small sample size (32 observations) using all the 6 instruments will increase bias and MSE if they are weakly correlated with the endogenous explanatory variable. Using only the most relevant instruments should improve estimation especially the bias of 2SLS.

For each sector we used our selection method to select the most relevant instruments. As can be expected (see Table 15) only 1 instrument is selected in general. The estimates are not very different for most industries except for industries 7 and 16 where the conclusion changes between the two estimations: the returns are decreasing if we use all instruments and are increasing if we use only selected instruments. The standard deviations are always smaller when we use all instruments. However the bias should be reduced in a substantial way and as simulations of Section 7 show, the MSE should not increase.

9. Conclusion

The literature on the selection of instruments is a very recent one and is still at the beginning. The topic seems very important to econometrics theory and practice in view of weak instruments problems. In this paper we propose a simple selection method based on finite-sample properties of 2SLS estimator and AR test. The Monte Carlo simulations show that in finite samples the method performs well and may be better than the other methods suggested in the literature. Asymptotically, under standard assumptions, the method is shown to be consistent in the sense that it selects

only the instruments having a non-zero coefficient in the true DGP of the endogenous explanatory variables.

Contrarily to the negative conclusion in Hall, Rudebusch, and Wilcox (1996) paper where it is reported that diagnostic tests are useless as screening tests, the results in this paper are very encouraging and show that selection of good instruments can improve the quality of estimation (especially in terms of the bias) and inference (at least when using the AR test). The challenge is to extend these selection procedures to more general models with more than one endogenous explanatory variable and/or to nonlinear models.

10. Appendix: Proofs

Proof of Proposition 4.1 Consider the model (2.1)-(2.2) with $G = 1$:

$$y_1 = \beta y_2 + X_1\gamma + u \ ,$$

$$y_2 = X_1\Pi_1 + X_2\Pi_2 + u_2 \ ,$$

$$y_1 = X_1\pi_1 + X_2\pi_2 + v_1 \ .$$

Following Sawa (1972, Theorem 1), under Assumptions (2.5) and (3.2), the first moment of the k-class estimator of β is given by (for $0 \leq k \leq 1$):

$$E(\tilde{\beta}_{(k)}) = \rho + (\beta - \rho)\frac{T\delta}{2}G(k, \frac{T\delta}{2}; m+1, n)$$

where $\rho = \sigma_{12}/\sigma_{22}$, $\delta = \Pi_2'(X_2'X_2)\Pi_2/(T\sigma_{22})$, $m = (T - k_1)/2$, $n = (T - k_1 - k_2)/2$, $G(k, T\delta; p, q) = \int_{-\infty}^{0} g(x; k, T\delta, p, q)dx$ and

$$g(x; k, T\delta, p, q) = \frac{2}{(1-2x)^{p-q}[1-2(1-k)x]^q}e^{(-T\delta + \frac{T\delta}{1-2x})}.$$

If we take $k = 1$, we obtain the first moment of two stage least squares estimator:

$$E(\tilde{\beta}_{2\text{SLS}}) = \rho + (\beta - \rho)\frac{T\delta}{2}G(1, \frac{T\delta}{2}, m+1, n) \tag{10.1}$$

where $G(1, T\delta; p, q) = \int_{-\infty}^{0} \frac{2}{(1-2x)^{p-q}}e^{(-T\delta + \frac{T\delta}{1-2x})}dx$, and the formula of the bias is given by

$$B_{2\text{SLS}}(\beta) = E(\tilde{\beta}_{2\text{SLS}}) - \beta = (\beta - \rho)[\frac{T\delta}{2}G(1, \frac{T\delta}{2}; m+1, n) - 1].$$

We first note that:

$$G(1, T\delta; p, q+1) = \int_{-\infty}^{0} \frac{2}{(1-2x)^{p-q}}e^{(-T\delta + \frac{T\delta}{1-2x})}(1-2x)dx$$

and since $x \in]-\infty, 0[$, we have $(1-2x) > 1$ and

$$G(1, T\delta; p, q+\varepsilon) > G(1, T\delta; p, q), \ \forall\ \varepsilon > 0.$$

G is thus an increasing function of q. Further, since $q = (T-k_1-k_2)/2$, $G(1, \frac{T\delta}{2}; \frac{T-k_1}{2}+1, \frac{T-k_1-k_2}{2})$ is a decreasing function of k_2 when δ is held fixed.

Second, Sawa (1972, Equation (4.30)) showed that:

$$0 < \frac{T\delta}{2} G(1, \frac{T\delta}{2}; m+1, n) < 1$$

so as k_2 increases (while δ remains constant), $\frac{T\delta}{2}G(1, \frac{T\delta}{2}; m+1, n)$ decreases towards 0, $(\frac{T\delta}{2}G(1, \frac{T\delta}{2}; m+1, n) - 1)$ decreases towards -1, and its absolute value increases towards 1. The absolute value of the bias is given by

$$|B_{2SLS}(\beta)| = |\beta - \rho| \left| \frac{T\delta}{2} G(1, \frac{T\delta}{2}; m+1, n) - 1 \right| ,$$

from which it follows it is an increasing function of k_2.

To see how the absolute value of the bias varies with the concentration parameter δ, we write the bias under another form [see Sawa (1972, Equation (4.12))]. We have:

$$\begin{aligned}
(T\delta)G(1, T\delta; m+1, n) &= T\delta e^{-T\delta} \frac{\Gamma(m-n)}{\Gamma(m-n+1)} \ {}_1F_1(m-n; m-n+1; T\delta) \\
&= \frac{e^{-T\delta}}{m-n} (T\delta) {}_1F_1(m-n; m-n+1; T\delta) \\
&= \frac{e^{-T\delta}}{m-n} (m-n)[{}_1F_1(m-n; m-n; T\delta) \\
&\quad - {}_1F_1(m-n-1; m-n; T\delta)] \\
&= e^{-T\delta}[e^{T\delta} - {}_1F_1(m-n-1; m-n; T\delta)] \\
&= 1 - e^{-T\delta} \sum_{h=0}^{+\infty} \frac{\Gamma(m-n-1+h)\Gamma(m-n)}{\Gamma(m-n-1)\Gamma(m-n+h)h!} (T\delta)^h \\
&= 1 - e^{-T\delta} \sum_{h=0}^{+\infty} \frac{(m-n-1)}{(m-n-1+h)h!} (T\delta)^h
\end{aligned}$$

where we used the relations [See Slater (1960)]

$$x\,{}_1F_1(p;q;x) = p[{}_1F_1(p;q-1;x) - {}_1F_1(p-1;q-1;x)]$$

and

$${}_1F_1(p;p;x) = e^x.$$

We deduce that,

$$\begin{aligned}
\frac{\partial(T\delta G(1,T\delta;m+1,n))}{\partial\delta} &= Te^{-T\delta}\sum_{h=0}^{+\infty}\frac{(m-n-1)}{(m-n-1+h)h!}(T\delta)^h \\
&\quad -Te^{-T\delta}\sum_{h=1}^{+\infty}\frac{(m-n-1)h}{(m-n-1+h)h!}(T\delta)^{h-1} \\
&= Te^{-T\delta}[\sum_{h=0}^{+\infty}\frac{(m-n-1)}{(m-n-1+h)h!}(T\delta)^h \\
&\quad -\sum_{h=0}^{+\infty}\frac{(m-n-1)}{(m-n+h)h!}(T\delta)^h] \\
&= Te^{-T\delta}(m-n-1)\sum_{h=0}^{+\infty}\left[\frac{1}{m-n-1+h}-\frac{1}{m-n+h}\right]\frac{(T\delta)^h}{h!} \\
&= Te^{-T\delta}(m-n-1)\sum_{h=0}^{+\infty}\frac{(T\delta)^h}{(m-n-1+h)(m-n+h)h!} > 0\,.
\end{aligned}$$

$(T\delta)G(1,T\delta;m+1,n)$ increases with δ towards 1, hence $|B_{2\text{SLS}}(\beta)|$ decreases as δ increases.

For the MSE, we take the formula in Richardson and Wu (1971, Equation (2.20))

$$\begin{aligned}
MSE(\beta) &= \frac{\sigma_{11}\sigma_{22}-\sigma_{12}^2}{\sigma_{22}^2}e^{\frac{-T\delta}{2}}\{\frac{1}{k_2-2}(1+\bar{\beta}^2)\,{}_1F_1(\frac{k_2}{2}-1;\frac{k_2}{2};\frac{T\delta}{2}) \\
&\quad +\frac{k_2-3}{k_2-2}\bar{\beta}_1^2F_1(\frac{k_2}{2}-2;\frac{k_2}{2};\frac{T\delta}{2})\}
\end{aligned}$$

or in terms of infinite series [see Slater (1960)]

$${}_1F_1(\frac{k_2}{2}-1;\frac{k_2}{2};\frac{T\delta}{2}) = \sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)h!}(\frac{T\delta}{2})^h$$

hence

$$ {}_1F_1(\frac{k_2}{2}-2;\frac{k_2}{2};\frac{T\delta}{2}) = \sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-2+h)(\frac{k_2}{2}-1+h)h!}(\frac{T\delta}{2})^h, $$

and

$$ \begin{aligned} a \times MSE(\beta) &= e^{\frac{-T\delta}{2}}\{(1+\bar{\beta}^2)\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)h!}(\frac{T\delta}{2})^h + \\ &\quad (k_2-3)\bar{\beta}^2\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-2+h)(\frac{k_2}{2}-1+h)h!}(\frac{T\delta}{2})^h\} \end{aligned} $$

where $a = (k_2-2)[(\sigma_{11}\sigma_{22}-\sigma_{12}^2)/\sigma_{22}^2]^{-1}$. Differentiating $a \times MSE(\beta)$ with respect to $\frac{T\delta}{2}$, we get:

$$ \begin{aligned} \frac{2}{T}\times\frac{\partial(a\times MSE(\beta))}{\partial\delta} &= -e^{\frac{-T\delta}{2}}\big[(1+\bar{\beta}^2)\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)h!}(\frac{T\delta}{2})^h + \\ &\quad (k_2-3)\bar{\beta}^2\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-2+h)(\frac{k_2}{2}-1+h)h!}(\frac{T\delta}{2})^h\big] + \\ &\quad e^{\frac{-T\delta}{2}}\big[(1+\bar{\beta}^2)\sum_{h=1}^{+\infty}\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)(h-1)!}(\frac{T\delta}{2})^{h-1} + \\ &\quad (k_2-3)\bar{\beta}^2\sum_{h=1}^{+\infty}\frac{(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-2+h)(\frac{k_2}{2}-1+h)(h-1)!}(\frac{T\delta}{2})^{h-1}\big] \\ &= -e^{\frac{-T\delta}{2}}\big[(1+\bar{\beta}^2)\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)h!}(\frac{T\delta}{2})^h + \\ &\quad (k_2-3)\bar{\beta}^2\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-2+h)(\frac{k_2}{2}-1+h)h!}(\frac{T\delta}{2})^h\big] + \\ &\quad e^{\frac{-T\delta}{2}}\big[(1+\bar{\beta}^2)\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}+h)h!}(\frac{T\delta}{2})^h + \\ &\quad (k_2-3)\bar{\beta}^2\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)(\frac{k_2}{2}+h)h!}(\frac{T\delta}{2})^h\big] \\ &= -e^{\frac{-T\delta}{2}}\big[(1+\bar{\beta}^2)\sum_{h=0}^{+\infty}[\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)}-\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}+h)}]\frac{(T\delta/2)^h}{h!} + \\ &\quad (k_2-3)\bar{\beta}^2\sum_{h=0}^{+\infty}[\frac{(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-2+h)(\frac{k_2}{2}-1+h)} - \end{aligned} $$

$$\frac{(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)(\frac{k_2}{2}+h)}]\frac{(T\delta/2)^h}{h!}]$$

$$= -e^{\frac{-T\delta}{2}}[(1+\bar{\beta}^2)\sum_{h=0}^{+\infty}\frac{(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-1+h)(\frac{k_2}{2}+h)}\frac{(T\delta/2)^h}{h!}+$$

$$(k_2-3)\bar{\beta}^2\sum_{h=0}^{+\infty}\frac{2(\frac{k_2}{2}-2)(\frac{k_2}{2}-1)}{(\frac{k_2}{2}-2+h)(\frac{k_2}{2}-1+h)(\frac{k_2}{2}+h)}\frac{(T\delta/2)^h}{h!}].$$

This last expression is the sum of two negative terms and so is negative. □

Proof of Proposition 5.2 Consider a subset a p instruments, $X_p = [x_{i_1}, ..., x_{i_p}]$. Since the columns of X are orthogonal

$$Y'P[M(X_1)X_p]Y = Y'P[X_p]Y = \sum_{j=1}^{p} Y'P(x_{i_j})Y$$

and

$$Y'P[M(X_1)\tilde{X}[1:i]]Y = Y'P(\tilde{X}[1:i])Y = \sum_{j=1}^{i} Y'P(\tilde{x}_{(j)})Y.$$

$\tilde{x}_{(1)}, ..., \tilde{x}_{(i)}$ are by definition the first i highest $Y'P(x_j)Y$ and $i \geq p$, therefore $Y'P(X_p)Y \leq Y'P(\tilde{X}[1:i])Y$. □

Proof of Proposition 6.1 First, since the instruments are asymptotically orthogonal, then for any column x_i of X :

$$\frac{Y'P[M(X_1)x_i]Y}{T} = \frac{(X_2^*\Pi_2^* + X_1\Pi_1 + v_2)'P[M(X_1)x_i](X_2^*\Pi_2^* + X_1\Pi_1 + v_2)}{T}$$

$$\underset{T\longrightarrow\infty}{\longrightarrow} \frac{(X_2^*\Pi_2^* + X_1\Pi_1 + v_2)'P(x_i)(X_2^*\Pi_2^* + X_1\Pi_1 + v_2)}{T}$$

If x_i is among the columns of X_2^* then

$$\frac{Y'P[M(X_1)x_i]Y}{T} \underset{T\longrightarrow\infty}{\xrightarrow{p}} \frac{\Pi_{2i}^{*2}x_i'x_i}{T} = \Pi_{2i}^{*2}Q_{x_i'x_i}.$$

where Π_{2i}^* is the i-th component of Π_2^* and $Q_{x_i'x_i}$ is the corresponding diagonal element

of Q. If x_i is not among the columns of X_2^* then

$$\frac{Y'P[M(X_1)x_i]Y}{T} \underset{T\longrightarrow\infty}{\xrightarrow{p}} 0\,.$$

Since $\Pi_{2i}^{*2}Q_{x_i'x_i}$ is positive, the first column of $\tilde{X}$ (the matrix of ordered instruments) is necessarily among the columns of X_2^*. Again since the instruments are asymptotically orthogonal

$$\frac{Y'P[M(X_1)[x_i,x_j]]Y}{T} \underset{T\longrightarrow\infty}{\longrightarrow} \frac{Y'P(x_i)Y}{T} + \frac{Y'P(x_j)Y}{T},$$

and the second column of $\tilde{X}$ will also be among the columns of X_2^*, and so forth until all the columns of X_2^* have been ordered. Now suppose that a subset X_2^1 of $m-1$ columns of X_2^* were selected and we test if we should add another instrument x_{2i}. The F-test is given by

$$F = (T-m)\frac{Y'[M[X_1,X_2^1] - M[X_1,X_2^1,x_{2i}]Y}{Y'M[X_1,X_2^1,x_{2i}]Y}.$$

After some manipulations and using the relation

$$\begin{aligned}\frac{Y'P[X_2^1,x_{2i}]Y}{T} &= \frac{Y'P(X_2^1)Y}{T} + \frac{Y'P[M(X_2^1)x_{2i}]Y}{T} \\ &\underset{T\longrightarrow\infty}{\longrightarrow} \frac{Y'P(X_2^1)Y}{T} + \frac{Y'P(x_{2i})Y}{T},\end{aligned}$$

we get

$$F \underset{T\longrightarrow\infty}{\longrightarrow} (T-m)[\frac{1-\frac{Y'P[X_1,X_2^1]Y}{Y'Y}}{1-\frac{Y'P[X_1,X_2^1]Y}{Y'Y}-\frac{Y'P(x_{2i})Y}{Y'Y}} - 1].$$

If x_{2i} is among the columns of X_2^*, then:

$$\frac{Y'P(x_{2i})Y}{Y'Y} = \frac{Y'P(x_{2i})Y}{T}(\frac{Y'Y}{T})^{-1} \underset{T\longrightarrow\infty}{\xrightarrow{p}} \frac{\Pi_{2i}^{*2}Q_{x_{2i}'x_{2i}}}{\Pi_2^{*\prime}Q^*_{X_2^{*\prime}X_2^*}\Pi_2 + \Pi_1'Q_{X_1'X_1}\Pi_1 + \sigma^2} > 0$$

and using Assumption (6.4), $F/c_{m,T} \underset{T\longrightarrow\infty}{\xrightarrow{p}} +\infty$, so the test rejects the hypothesis of zero-coefficient and the instrument is added.

If x_{2i} is not among the instruments in X_2^*, then:

$$\frac{Y'P(x_{2i})Y}{Y'Y} = \frac{Y'P(x_{2i})Y}{T}(\frac{Y'Y}{T})^{-1} \xrightarrow[T\longrightarrow\infty]{p} \frac{0}{\Pi_2^{*\prime}Q^*_{X_2^{*\prime}X_2^*}\Pi_2^* + \Pi_1'Q_{X_1'X_1}\Pi_1 + \sigma^2} = 0$$

and $F \xrightarrow[T\longrightarrow\infty]{p} \chi_1^2$, but since the critical value $c_{m,T} \longrightarrow +\infty$, the test does not reject the hypothesis of zero-coefficient and the instrument is not added.

By a same argument it is easy to see that an instrument in X_2^* that is included will never be removed after. Hence the proposition. □

Table 1: Bias and MSE of 2SLS
Model (2.1)-(2.2)

	T = 50				T = 100				T = 200			
	8 relevant instruments		2 relevant instruments		8 relevant instruments		2 relevant instruments		8 relevant instruments		2 relevant instruments	
k_2	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
1	-0.039		-0.056		-0.018		-0.025		-0.012		-0.012	
2	-0.002		0.000		-0.001		0.001		0.001		-0.001	
3	0.015	0.034	0.026	0.047	0.005	0.012	0.008	0.020	0.005	0.008	0.006	0.012
4	0.024	0.024	0.045	0.045	0.010	0.010	0.020	0.020	0.007	0.006	0.012	0.011
5	0.024	0.015	0.066	0.045	0.012	0.008	0.028	0.019	0.006	0.004	0.017	0.012
6	0.021	0.011	0.081	0.044	0.012	0.006	0.037	0.019	0.008	0.004	0.022	0.012
7	0.021	0.009	0.099	0.045	0.012	0.005	0.044	0.019	0.008	0.004	0.028	0.012
8	0.023	0.008	0.114	0.046	0.015	0.005	0.053	0.020	0.009	0.003	0.034	0.012
9	0.027	0.008	0.128	0.048	0.018	0.005	0.062	0.020	0.012	0.003	0.037	0.012
10	0.030	0.008	0.141	0.049	0.020	0.005	0.068	0.021	0.013	0.003	0.043	0.012
11	0.033	0.008	0.153	0.052	0.022	0.005	0.077	0.021	0.014	0.003	0.048	0.012
12	0.036	0.008	0.166	0.055	0.025	0.005	0.083	0.022	0.016	0.003	0.052	0.013
13	0.040	0.008	0.173	0.057	0.027	0.005	0.090	0.023	0.018	0.003	0.057	0.013
14	0.043	0.009	0.185	0.060	0.030	0.006	0.097	0.024	0.019	0.003	0.063	0.014
15	0.045	0.009	0.191	0.061	0.031	0.006	0.102	0.025	0.021	0.004	0.065	0.014
16	0.049	0.009	0.201	0.065	0.034	0.006	0.109	0.026	0.023	0.004	0.072	0.014
17	0.052	0.009	0.206	0.065	0.036	0.006	0.116	0.027	0.024	0.004	0.076	0.015
18	0.056	0.009	0.216	0.069	0.038	0.006	0.121	0.028	0.025	0.004	0.079	0.015
19	0.058	0.010	0.224	0.072	0.041	0.006	0.129	0.029	0.026	0.004	0.084	0.016
20	0.062	0.010	0.232	0.075	0.043	0.006	0.132	0.031	0.029	0.004	0.088	0.016
21	0.065	0.010	0.238	0.077	0.045	0.007	0.137	0.031	0.029	0.004	0.091	0.017
22	0.067	0.011	0.243	0.079	0.047	0.007	0.143	0.033	0.031	0.004	0.095	0.017
23	0.070	0.011	0.251	0.083	0.049	0.007	0.147	0.034	0.032	0.004	0.098	0.018
24	0.072	0.011	0.256	0.085	0.052	0.007	0.153	0.035	0.034	0.004	0.102	0.019
25	0.076	0.012	0.261	0.087	0.054	0.007	0.156	0.036	0.035	0.004	0.105	0.019
26	0.079	0.012	0.266	0.089	0.056	0.007	0.163	0.038	0.036	0.004	0.111	0.020
27	0.082	0.013	0.271	0.091	0.058	0.008	0.167	0.039	0.038	0.004	0.114	0.021
28	0.084	0.013	0.276	0.093	0.060	0.008	0.172	0.041	0.039	0.004	0.117	0.022
29	0.086	0.013	0.280	0.095	0.062	0.008	0.176	0.042	0.041	0.004	0.120	0.022
30	0.089	0.014	0.285	0.097	0.064	0.008	0.179	0.043	0.041	0.005	0.123	0.023
31	0.092	0.014	0.290	0.100	0.066	0.009	0.183	0.044	0.044	0.005	0.128	0.024
32	0.094	0.014	0.292	0.101	0.068	0.009	0.187	0.045	0.044	0.005	0.130	0.024
33	0.097	0.015	0.297	0.104	0.071	0.009	0.192	0.047	0.046	0.005	0.133	0.025
34	0.099	0.015	0.302	0.106	0.072	0.009	0.194	0.048	0.047	0.005	0.136	0.026
35	0,102	0.016	0.305	0.108	0.073	0.010	0.199	0.049	0.049	0.005	0.140	0.027
36	0.104	0.016	0.307	0.109	0.076	0.010	0.204	0.051	0.051	0.005	0.143	0.028
37	0.107	0.017	0.310	0.110	0.078	0.010	0.206	0.052	0.051	0.005	0.146	0.028
38	0.110	0.017	0.315	0.113	0.078	0.010	0.209	0.053	0.053	0.006	0.149	0.029
39	0.112	0.018	0.317	0.114	0.081	0.010	0.212	0.054	0.054	0.006	0.151	0.030
40	0.114	0.018	0.320	0.116	0.083	0.011	0.216	0.056	0.056	0.006	0.154	0.031

Table 2: Bias and MSE of 2SLS when the regressors are exogenous

Model (2.1)-(2.2), $cov(u,v2) = 0$

	T = 50				T = 100				T = 200			
	8 relevant instruments		2 relevant instruments		8 relevant instruments		2 relevant instruments		8 relevant instruments		2 relevant instruments	
k_2	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE	BIAS	MSE
1	-0.045		0.012		0.002		0.000		0.001		-0.001	
2	-0.002		0.000		0.000		-0.001		0.000		-0.002	
3	0.001	0.031	-0.001	0.045	0.000	0.014	0.000	0.018	0.001	0.006	-0.001	0.011
4	0.001	0.023	-0.001	0.043	0.000	0.011	0.001	0.017	0.000	0.005	0.000	0.011
5	0.001	0.022	0.000	0.041	0.000	0.009	0.001	0.017	0.000	0.004	-0.001	0.011
6	-0.002	0.018	-0.003	0.039	0.001	0.006	0.001	0.017	0.000	0.003	0.000	0.011
7	-0.001	0.017	-0.002	0.038	0.000	0.006	0.001	0.017	0.000	0.003	0.001	0.011
8	0.001	0.015	0.000	0.036	0.000	0.005	-0.001	0.016	0.000	0.002	-0.001	0.010
9	0.001	0.015	0.002	0.035	0.000	0.005	0.002	0.016	0.000	0.002	0.001	0.010
10	0.000	0.015	-0.002	0.034	0.000	0.005	0.001	0.016	0.000	0.002	0.000	0.010
11	-0.001	0.015	0.000	0.032	0.000	0.005	-0.001	0.016	0.000	0.002	-0.001	0.010
12	-0.001	0.014	-0.002	0.031	0.000	0.005	0.000	0.015	0.000	0.002	0.000	0.010
13	0.000	0.014	0.001	0.030	0.000	0.005	0.000	0.015	0.001	0.002	-0.001	0.010
14	0.000	0.014	0.000	0.030	0.000	0.005	0.001	0.015	0.000	0.002	-0.001	0.010
15	0.001	0.014	0.001	0.029	0.000	0.005	0.002	0.015	0.001	0.002	0.000	0.010
16	0.000	0.014	0.002	0.028	0.000	0.005	-0.001	0.014	0.000	0.002	0.000	0.010
17	-0.001	0.013	0.000	0.027	0.000	0.005	-0.001	0.014	0.000	0.002	0.001	0.009
18	0.000	0.013	0.002	0.026	0.000	0.005	0.000	0.014	0.000	0.002	-0.001	0.009
19	0.000	0.013	0.002	0.026	0.000	0.005	0.001	0.014	0.000	0.002	-0.001	0.009
20	0.000	0.013	0.002	0.025	0.000	0.005	0.001	0.014	0.000	0.002	0.000	0.009
21	0.001	0.013	-0.001	0.025	0.000	0.005	0.000	0.014	0.000	0.002	-0.001	0.009
22	0.000	0.012	-0.002	0.024	0.000	0.005	0.001	0.013	0.000	0.002	0.000	0.009
23	-0.001	0.012	0.000	0.023	-0.001	0.005	0.001	0.013	0.000	0.002	0.000	0.009
24	0.001	0.012	0.001	0.022	0.001	0.005	0.000	0.013	0.000	0.002	0.001	0.009
25	-0.001	0.012	0.000	0.022	0.001	0.005	0.000	0.013	0.000	0.002	-0.001	0.009
26	0.000	0.012	0.000	0.022	0.000	0.005	0.001	0.013	-0.001	0.002	0.001	0.009
27	0.000	0.012	0.000	0.021	0.001	0.005	0.001	0.012	0.000	0.002	0.000	0.009
28	-0.001	0.012	0.000	0.021	0.000	0.005	0.000	0.012	-0.001	0.002	0.000	0.009
29	-0.001	0.011	0.001	0.021	-0.001	0.005	0.000	0.012	-0.001	0.002	0.000	0.008
30	0.002	0.011	0.001	0.020	0.000	0.005	0.000	0.012	0.000	0.002	0.001	0.009
31	0.001	0.011	0.002	0.020	0.000	0.005	0.000	0.012	0.000	0.002	0.000	0.008
32	0.000	0.011	0.000	0.019	0.000	0.004	-0.001	0.012	0.000	0.002	0.000	0.008
33	0.000	0.011	0.000	0.019	0.001	0.005	0.000	0.012	0.000	0.002	0.000	0.008
34	0.000	0.011	0.000	0.018	0.000	0.004	0.000	0.012	0.000	0.002	-0.001	0.008
35	0.000	0.010	0.000	0.018	0.000	0.005	0.000	0.011	0.000	0.002	0.000	0.008
36	0.002	0.011	-0.001	0.018	0.000	0.005	0.000	0.011	0.000	0.002	0.000	0.008
37	0.000	0.010	0.000	0.017	0.000	0.005	-0.002	0.011	0.000	0.002	-0.001	0.008
38	-0.001	0.010	0.000	0.017	0.001	0.004	0.000	0.011	0.001	0.002	0.000	0.008
39	0.001	0.010	-0.001	0.017	0.000	0.004	0.002	0.011	0.000	0.002	0.000	0.008
40	0.000	0.010	0.001	0.017	-0.001	0.004	-0.001	0.011	0.000	0.002	0.001	0.008

Table 3 : Selection frequency for each Instrument

Instrument	$T = 20$	$T = 50$	$T = 100$	$T = 200$
		$l = 10$		
X_1	0.01	0.00	0.00	0.00
X_2	0.02	0.00	0.00	0.00
X_3	0.01	0.00	0.00	0.00
X_4	*0.91*	*1.00*	*1.00*	*1.00*
X_5	*0.86*	*0.99*	*1.00*	*1.00*
X_6	0.02	0.00	0.00	0.00
X_7	0.01	0.00	0.00	0.00
X_8	0.01	0.00	0.00	0.00
X_9	0.01	0.00	0.00	0.00
X_{10}	0.01	0.00	0.00	0.00
		$l = 20$		
X_1	0.01	0.00	0.00	0.00
X_2	0.01	0.00	0.00	0.00
X_3	0.01	0.00	0.00	0.00
X_4	*0.75*	*0.99*	*1.00*	*1.00*
X_5	*0.83*	*1.00*	*1.00*	*1.00*
X_6	0.01	0.00	0.00	0.00
X_7	0.01	0.00	0.00	0.00
X_8	0.01	0.00	0.00	0.00
X_9	0.05	0.00	0.00	0.00
X_{10}	0.02	0.00	0.00	0.00
X_{11}	0.01	0.00	0.00	0.00
X_{12}	0.01	0.00	0.00	0.00
X_{13}	0.01	0.00	0.00	0.00
X_{14}	0.03	0.00	0.00	0.00
X_{15}	0.01	0.00	0.00	0.00
X_{16}	0.01	0.00	0.00	0.00
X_{17}	0.04	0.00	0.00	0.00
X_{18}	0.01	0.00	0.00	0.00
X_{19}	0.12	0.00	0.00	0.00
X_{20}	0.01	0.00	0.00	0.00

Table 4 : Selection frequency for each number of instruments

Number of instruments	$T = 20$	$T = 50$	$T = 100$	$T = 200$
		$l = 10$		
1	0.18	0.01	0.00	0.00
2	*0.76*	*0.98*	*1.00*	*1.00*
3	0.06	0.01	0.00	0.00
4	*0.00*	*0.00*	*0.00*	*0.00*
5	*0.00*	*0.00*	*0.00*	*0.00*
6	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
		$l = 20$		
1	0.19	0.01	0.00	0.00
2	*0.68*	*0.98*	*1.00*	*1.00*
3	0.12	0.02	0.00	0.00
4	*0.02*	*0.00*	*0.00*	*0.00*
5	*0.00*	*0.00*	*0.00*	*0.00*
6	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00
13	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00
18	0.00	0.00	0.00	0.00
19	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00

Table 5 : Bias of 2SLS with selected instruments*

T	Fraction of the sample used for selection	All instruments	Selected instruments	Principal comp. orthogonalization +selection	Gram-Schmidt orthogonalization +selection
			$l = 10$		
50	20%	1.87	-0.41	0.80	-0.41
	30%	1.87	-0.09	0.81	-0.09
	100%	1.87	0.25	1.86	0.01
100	20%	1.05	0.15	0.13	0.16
	30%	1.05	0.10	0.16	0.10
	100%	1.05	0.11	0.95	0.00
200	20%	0.51	-0.01	0.16	-0.01
	30%	0.51	-0.02	0.06	-0.02
	100%	0.51	-0.03	0.24	-0.09
			$l = 20$		
50	100%	3.55	0.22	3.31	-0.01
100	20%	2.25	0.53	0.44	0.53
	30%	2.25	0.11	0.20	0.11
	100%	2.25	0.04	1.90	-0.10
200	20%	1.13	-0.17	0.22	-0.17
	30%	1.13	-0.13	-0.16	-0.13
	100%	1.13	0.09	3.00	0.02

*The numbers are multiplied by 100 to make them easier to read

Table 6 : MSE of 2SLS with selected instruments*

T	Fraction of the sample used for selection	All instruments	Selected instruments	Principal comp. orthogonalization +selection	Gram-Schmidt orthogonalization +selection
			$l = 10$		
50	20%	1.00	1.48	2.94	1.48
	30%	1.00	1.47	3.16	1.47
	100%	1.00	1.06	1.06	1.10
100	20%	0.50	0.53	1.38	0.53
	30%	0.50	0.65	0.90	0.65
	100%	0.50	-	-	-
200	20%	0.26	0.39	0.44	0.39
	30%	0.26	0.51	0.49	0.51
	100%	0.26	-	-	-
			$l = 20$		
50	100%	0.92	0.93	1.20	0.79
100	20%	0.55	1.27	3.24	1.27
	30%	0.55	0.69	3.09	0.69
	100%	0.55	-	-	-
200	20%	0.27	0.34	0.80	0.34
	30%	0.27	0.41	0.68	0.41
	100%	0.27	-	-	-

*The numbers are multiplied by 100 to make them easier to read

Table 7 : MAE of 2SLS with selected instruments*

T	Fraction of the sample used for selection	All instruments	Selected instruments	Principal comp. orthogonalization +selection	Gram-Schmidt orthogonalization +selection
			$l = 10$		
50	20%	7.99	10.02	14.34	10.03
	30%	7.99	9.45	15.39	9.45
	100%	7.99	8.17	8.99	8.20
100	20%	5.47	6.26	12.93	6.26
	30%	5.47	6.43	8.52	6.43
	100%	5.47	5.37	6.27	5.38
200	20%	4.08	4.68	5.75	4.68
	30%	4.08	5.00	5.73	5.00
	100%	4.08	3.93	4.16	3.94
			$l = 20$		
50	100%	7.70	7.66	9.85	7.70
100	20%	5.93	7.21	15.28	7.21
	30%	5.93	6.57	16.35	6.57
	100%	5.93	5.96	7.16	5.98
200	20%	4.17	4.78	8.13	4.78
	30%	4.17	4.93	7.39	4.93
	100%	4.17	4.13	6.91	4.14

*The numbers are multiplied by 100 to make them easier to read

Table 8 : Power of the *AR* test with selected instruments

H_0 $\beta = 1$ in model (7.1) – (7.2)

(*a*) $T = 100, l = 10, s = 0.2$

Alternative	All	sel-Z	Pr-comp	GS
0.50	1.00	0.99	0.96	0.99
0.55	0.99	0.98	0.93	0.98
0.60	0.97	0.96	0.88	0.96
0.65	0.91	0.92	0.80	0.92
0.70	0.80	0.84	0.69	0.84
0.75	0.59	0.70	0.55	0.70
0.80	0.39	0.52	0.39	0.52
0.85	0.21	0.32	0.24	0.32
0.90	0.11	0.16	0.13	0.16
0.95	0.06	0.08	0.07	0.08
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.06	0.07	0.07	0.07
1.10	0.10	0.15	0.12	0.15
1.15	0.18	0.27	0.21	0.27
1.20	0.32	0.44	0.33	0.44
1.25	0.48	0.61	0.46	0.61
1.30	0.65	0.74	0.58	0.74
1.35	0.78	0.84	0.69	0.84
1.40	0.88	0.90	0.78	0.90
1.45	0.94	0.94	0.84	0.94
1.50	0.98	0.96	0.88	0.96

(*b*) $T = 100, l = 10, s = 0.3$

Alternative	All	sel-Z	Pr-comp	GS
0.50	1.00	0.98	0.97	0.98
0.55	0.99	0.97	0.95	0.97
0.60	0.97	0.95	0.92	0.95
0.65	0.91	0.91	0.85	0.91
0.70	0.80	0.82	0.75	0.82
0.75	0.59	0.66	0.58	0.66
0.80	0.39	0.49	0.41	0.49
0.85	0.21	0.28	0.24	0.28
0.90	0.11	0.14	0.12	0.14
0.95	0.06	0.07	0.07	0.07
1.00	***0.05***	***0.05***	***0.06***	***0.05***
1.05	0.06	0.07	0.07	0.07
1.10	0.10	0.14	0.12	0.14
1.15	0.18	0.25	0.21	0.25
1.20	0.32	0.41	0.34	0.41
1.25	0.48	0.57	0.48	0.57
1.30	0.65	0.70	0.62	0.70
1.35	0.78	0.81	0.73	0.81
1.40	0.88	0.89	0.83	0.89
1.45	0.94	0.93	0.89	0.93
1.50	0.98	0.96	0.92	0.96

(*c*) $T = 200, l = 10, s = 0.2$

Alternative	All	sel-Z	Pr-comp	GS
0.50	1.00	1.00	1.00	1.00
0.55	1.00	1.00	1.00	1.00
0.60	1.00	1.00	1.00	1.00
0.65	1.00	1.00	1.00	1.00
0.70	1.00	1.00	0.98	1.00
0.75	0.96	0.99	0.91	0.99
0.80	0.80	0.93	0.75	0.93
0.85	0.49	0.72	0.50	0.72
0.90	0.21	0.37	0.25	0.37
0.95	0.08	0.12	0.10	0.12
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.08	0.12	0.09	0.12
1.10	0.20	0.34	0.24	0.34
1.15	0.43	0.66	0.45	0.66
1.20	0.71	0.88	0.67	0.88
1.25	0.90	0.97	0.84	0.97
1.30	0.98	0.99	0.93	0.99
1.35	1.00	1.00	0.97	1.00
1.40	1.00	1.00	0.99	1.00
1.45	1.00	1.00	1.00	1.00
1.50	1.00	1.00	1.00	1.00

(*d*) $T = 200, l = 10, s = 0.3$

Alternative	All	sel-Z	Pr-comp	GS
0.50	1.00	1.00	0.98	1.00
0.55	1.00	1.00	0.98	1.00
0.60	1.00	1.00	0.97	1.00
0.65	1.00	1.00	0.96	1.00
0.70	1.00	1.00	0.94	1.00
0.75	0.96	0.98	0.86	0.98
0.80	0.80	0.90	0.71	0.90
0.85	0.49	0.65	0.45	0.65
0.90	0.21	0.32	0.21	0.32
0.95	0.08	0.11	0.09	0.11
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.08	0.11	0.08	0.11
1.10	0.20	0.30	0.20	0.30
1.15	0.42	0.58	0.40	0.58
1.20	0.69	0.83	0.62	0.83
1.25	0.89	0.95	0.79	0.95
1.30	0.97	0.99	0.89	0.99
1.35	1.00	1.00	0.94	1.00
1.40	1.00	1.00	0.96	1.00
1.45	1.00	1.00	0.97	1.00
1.50	1.00	1.00	0.97	1.00

All: AR test using all the potential instruments

Z-sel: AR test with selected instruments among the original ones

Pr-comp.: AR test with selected instruments after principal components decomposition of the original set of instruments

GS: AR test with selected instruments after Gram-Schmidt decomposition of the original set of instruments

Table 8 (cont'd): Power of the AR test with selected instruments

H_0 $\beta = 1$ in model (7.1) – (7.2)

(e) $T = 100, l = 20, s = 0.2$

Alternative	All	sel-Z	Pr-comp	GS
0.50	0.99	0.98	0.27	0.98
0.55	0.97	0.97	0.23	0.97
0.60	0.92	0.94	0.21	0.94
0.65	0.81	0.90	0.19	0.90
0.70	0.64	0.83	0.15	0.83
0.75	0.44	0.69	0.12	0.69
0.80	0.28	0.52	0.10	0.52
0.85	0.15	0.31	0.08	0.31
0.90	0.09	0.17	0.06	0.17
0.95	0.06	0.08	0.05	0.08
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.06	0.07	0.05	0.07
1.10	0.08	0.16	0.06	0.16
1.15	0.15	0.28	0.07	0.28
1.20	0.22	0.44	0.09	0.44
1.25	0.35	0.60	0.11	0.60
1.30	0.49	0.73	0.13	0.73
1.35	0.63	0.82	0.15	0.82
1.40	0.77	0.89	0.18	0.89
1.45	0.86	0.92	0.19	0.92
1.50	0.92	0.95	0.21	0.95

(f) $T = 100, l = 20, s = 0.3$

Alternative	All	sel-Z	Pr-comp	GS
0.50	0.99	1.00	0.39	1.00
0.55	0.97	1.00	0.34	1.00
0.60	0.92	0.99	0.30	0.99
0.65	0.81	0.96	0.25	0.96
0.70	0.64	0.89	0.21	0.89
0.75	0.44	0.74	0.17	0.74
0.80	0.28	0.54	0.12	0.54
0.85	0.15	0.32	0.09	0.32
0.90	0.09	0.16	0.07	0.16
0.95	0.06	0.08	0.06	0.08
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.06	0.08	0.05	0.08
1.10	0.08	0.16	0.06	0.16
1.15	0.15	0.29	0.08	0.29
1.20	0.22	0.46	0.11	0.46
1.25	0.35	0.63	0.15	0.63
1.30	0.49	0.78	0.17	0.78
1.35	0.63	0.88	0.20	0.88
1.40	0.77	0.94	0.24	0.94
1.45	0.86	0.97	0.28	0.97
1.50	0.92	0.99	0.30	0.99

(g) $T = 200, l = 20, s = 0.2$

Alternative	All	sel-Z	Pr-comp	GS
0.50	1.00	1.00	0.72	1.00
0.55	1.00	1.00	0.68	1.00
0.60	1.00	1.00	0.61	1.00
0.65	1.00	1.00	0.53	1.00
0.70	0.97	1.00	0.43	1.00
0.75	0.86	0.98	0.34	0.98
0.80	0.63	0.90	0.24	0.90
0.85	0.33	0.65	0.16	0.65
0.90	0.14	0.32	0.09	0.32
0.95	0.06	0.11	0.06	0.11
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.07	0.11	0.06	0.11
1.10	0.14	0.30	0.09	0.30
1.15	0.29	0.59	0.14	0.59
1.20	0.50	0.83	0.20	0.83
1.25	0.74	0.95	0.29	0.95
1.30	0.89	0.99	0.36	0.99
1.35	0.97	1.00	0.45	1.00
1.40	0.99	1.00	0.49	1.00
1.45	1.00	1.00	0.56	1.00
1.50	1.00	1.00	0.61	1.00

(h) $T = 200, l = 20, s = 0.3$

Alternative	All	sel-Z	Pr-comp	GS
0.50	1.00	1.00	0.84	1.00
0.55	1.00	1.00	0.79	1.00
0.60	1.00	1.00	0.72	1.00
0.65	1.00	1.00	0.62	1.00
0.70	0.98	1.00	0.51	1.00
0.75	0.89	0.98	0.40	0.98
0.80	0.65	0.91	0.28	0.91
0.85	0.36	0.67	0.19	0.67
0.90	0.16	0.33	0.11	0.33
0.95	0.07	0.11	0.06	0.11
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.07	0.10	0.06	0.10
1.10	0.14	0.31	0.10	0.31
1.15	0.30	0.60	0.16	0.60
1.20	0.54	0.84	0.24	0.84
1.25	0.77	0.95	0.33	0.95
1.30	0.92	0.99	0.43	0.99
1.35	0.98	1.00	0.51	1.00
1.40	0.99	1.00	0.58	1.00
1.45	1.00	1.00	0.66	1.00
1.50	1.00	1.00	0.72	1.00

All: AR test using all the potential instruments
Z-sel: AR test with selected instruments among the original ones
Pr-comp.: AR test with selected instruments after principal components decomposition of the original set of instruments
GS: AR test with selected instruments after Gram-Schmidt decomposition of the original set of instruments

Table 9 : Comparison of different selection methods
Bias of 2SLS*

T	fraction of the sample used for selection	All instruments	Our methdod	Hall-Piexe	Donald-Newey
			l = 10		
50	30%	2.03	0.30	1.31	2.03
	100%	2.03	0.08	0.67	1.17
100	20%	1.04	0.16	0.40	0.63
	30%	1.04	0.20	0.31	0.71
	100%	1.04	0.01	0.26	0.70
200	20%	0.52	-0.05	0.02	0.27
	30%	0.52	0.00	0.05	0.36
	100%	0.52	0.03	0.12	0.38
			l = 20		
50	100%	3.62	0.09	1.68	1.37
100	20%	2.24	0.05	2.55	0.82
	30%	2.24	0.08	0.97	0.81
	100%	2.24	0.00	0.65	0.88
200	20%	1.12	0.07	0.34	0.43
	30%	1.12	0.07	0.24	0.51
	100%	1.12	0.00	0.21	0.39

*The numbers are multiplied by 100 to make them easier to read

Table 10 : Comparison of different selection methods
MSE of 2SLS*

T	fraction of the sample used for selection	All instruments	Our methdod	Hall-Plexe	Donald-Newey
			l = 10		
50	30%	1.10	1.54	1.59	1.69
	100%	0.90	0.87	0.89	0.93
100	20%	0.45	0.60	0.58	0.57
	30%	0.45	0.61	0.61	0.60
	100%	0.53	-	0.51	0.54
200	20%	0.27	0.40	0.35	0.35
	30%	0.27	0.29	0.43	0.41
	100%	0.28	-	0.28	0.28
			l = 20		
50	100%	0.90	0.94	0.89	0.90
100	20%	0.51	1.24	0.61	0.61
	30%	0.51	0.69	0.69	0.67
	100%	0.57	-	0.56	0.58
200	20%	0.26	0.38	0.36	0.35
	30%	0.26	0.34	0.41	0.40
	100%	0.25	-	0.25	0.26

*The numbers are multiplied by 100 to make them easier to read

Table 11 : Comparison of different selection methods
MAE of 2SLS*

T	fraction of the sample used for selection	All instruments	Our methdod	Hall-Piexe	Donald-Newey
			l = 10		
50	30%	8.39	10.17	9.98	11.60
	100%	7.57	7.72	7.66	7.67
100	20%	5.33	6.00	6.13	6.00
	30%	5.33	6.23	6.25	6.18
	100%	5.78	5.84	5.82	5.84
200	20%	4.15	4.78	4.77	4.77
	30%	4.15	5.13	5.13	5.10
	100%	4.21	4.24	4.24	4.24
			l = 20		
50	100%	7.61	7.61	7.54	7.55
100	20%	5.72	6.62	6.25	6.22
	30%	5.72	6.63	6.65	6.53
	100%	6.02	6.02	5.97	6.05
200	20%	4.08	4.74	4.74	4.71
	30%	4.08	5.08	5.07	5.06
	100%	4.04	4.04	4.02	4.07

*The numbers are multiplied by 100 to make them easier to read

Table 12 : Comparison of different selection methods
Power of the *AR* test for H_0: $\beta = 1$ in model (7.1) – (7.2)

(*a*) $T = 100, l = 10, s = .2$

Altrnative	All	Our	HP	DN
0.50	1.00	0.99	1.00	1.00
0.55	1.00	0.99	1.00	0.99
0.60	0.99	0.99	0.99	0.98
0.65	0.95	0.98	0.97	0.95
0.70	0.86	0.94	0.91	0.87
0.75	0.68	0.84	0.78	0.70
0.80	0.45	0.63	0.56	0.48
0.85	0.25	0.40	0.34	0.27
0.90	0.12	0.19	0.16	0.13
0.95	0.07	0.08	0.08	0.06
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.06	0.08	0.07	0.07
1.10	0.12	0.19	0.16	0.13
1.15	0.22	0.36	0.30	0.24
1.20	0.37	0.55	0.48	0.39
1.25	0.56	0.74	0.67	0.58
1.30	0.72	0.86	0.81	0.74
1.35	0.85	0.93	0.91	0.86
1.40	0.93	0.97	0.96	0.93
1.45	0.97	0.98	0.98	0.97
1.50	0.99	0.99	0.99	0.98

(*b*) $T = 100, l = 10, s = .3$

Altrnative	All	Our	HP	DN
0.50	1.00	0.99	1.00	1.00
0.55	1.00	0.99	1.00	1.00
0.60	0.99	0.98	1.00	0.99
0.65	0.95	0.96	0.98	0.97
0.70	0.85	0.92	0.93	0.93
0.75	0.67	0.81	0.80	0.79
0.80	0.44	0.62	0.59	0.59
0.85	0.24	0.39	0.35	0.36
0.90	0.13	0.19	0.18	0.18
0.95	0.06	0.08	0.07	0.07
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.06	0.09	0.08	0.08
1.10	0.11	0.18	0.16	0.16
1.15	0.21	0.35	0.32	0.32
1.20	0.35	0.54	0.51	0.50
1.25	0.54	0.71	0.69	0.70
1.30	0.70	0.84	0.83	0.83
1.35	0.84	0.91	0.92	0.92
1.40	0.92	0.95	0.96	0.96
1.45	0.97	0.97	0.98	0.98
1.50	0.99	0.98	0.99	0.99

(*c*) $T = 200, l = 10, s = .2$

Altrnative	All	Our	HP	DN
0.50	1.00	1.00	1.00	1.00
0.55	1.00	1.00	1.00	1.00
0.60	1.00	1.00	1.00	1.00
0.65	1.00	1.00	1.00	1.00
0.70	1.00	1.00	1.00	1.00
0.75	0.95	0.98	0.98	0.97
0.80	0.77	0.90	0.88	0.86
0.85	0.47	0.66	0.63	0.61
0.90	0.21	0.34	0.31	0.30
0.95	0.08	0.11	0.11	0.10
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.07	0.10	0.10	0.09
1.10	0.19	0.30	0.27	0.26
1.15	0.40	0.59	0.56	0.54
1.20	0.68	0.84	0.81	0.79
1.25	0.87	0.95	0.94	0.92
1.30	0.97	0.99	0.99	0.98
1.35	0.99	1.00	1.00	1.00
1.40	1.00	1.00	1.00	1.00
1.45	1.00	1.00	1.00	1.00
1.50	1.00	1.00	1.00	1.00

(*d*) $T = 200, l = 10, s = .3$

Altrnative	All	Our	HP	DN
0.50	1.00	1.00	1.00	1.00
0.55	1.00	1.00	1.00	1.00
0.60	1.00	1.00	1.00	1.00
0.65	1.00	1.00	1.00	1.00
0.70	1.00	1.00	1.00	1.00
0.75	0.97	0.98	0.98	0.97
0.80	0.83	0.90	0.88	0.85
0.85	0.53	0.66	0.63	0.59
0.90	0.22	0.33	0.30	0.28
0.95	0.08	0.11	0.11	0.10
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.08	0.10	0.10	0.10
1.10	0.21	0.31	0.29	0.27
1.15	0.45	0.59	0.56	0.53
1.20	0.74	0.84	0.81	0.78
1.25	0.91	0.95	0.94	0.92
1.30	0.98	0.99	0.99	0.98
1.35	1.00	1.00	1.00	0.99
1.40	1.00	1.00	1.00	1.00
1.45	1.00	1.00	1.00	1.00
1.50	1.00	1.00	1.00	1.00

All : Using all instruments
Our : using instruments selected by our method
HP : using instruments selected by Hall-Piexe method
DN : using instruments selected by Donal-Newey method

Table 12 (cont'd): Comparison of different selection methods
Power of the AR test for H_0: $\beta = 1$ in model (7.1) − (7.2)

(e) $T = 100, l = 20, s = .2$

Altrnative	All	Our	HP	DN
0.50	0.97	0.87	0.92	0.99
0.55	0.93	0.86	0.85	0.98
0.60	0.85	0.84	0.73	0.96
0.65	0.71	0.79	0.58	0.92
0.70	0.55	0.73	0.44	0.83
0.75	0.38	0.60	0.30	0.68
0.80	0.23	0.43	0.19	0.48
0.85	0.13	0.26	0.11	0.28
0.90	0.09	0.13	0.07	0.14
0.95	0.06	0.07	0.06	0.07
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.05	0.06	0.05	0.07
1.10	0.09	0.13	0.07	0.13
1.15	0.12	0.23	0.11	0.25
1.20	0.19	0.36	0.15	0.40
1.25	0.29	0.51	0.23	0.56
1.30	0.41	0.63	0.33	0.72
1.35	0.54	0.72	0.42	0.82
1.40	0.66	0.78	0.53	0.89
1.45	0.76	0.81	0.64	0.94
1.50	0.85	0.85	0.73	0.96

(f) $T = 100, l = 20, s = .3$

Altrnative	All	Our	HP	DN
0.50	1.00	1.00	1.00	1.00
0.55	1.00	1.00	0.99	1.00
0.60	0.98	1.00	0.97	0.99
0.65	0.91	0.98	0.90	0.97
0.70	0.79	0.95	0.80	0.91
0.75	0.58	0.84	0.63	0.79
0.80	0.36	0.64	0.42	0.58
0.85	0.20	0.39	0.24	0.35
0.90	0.10	0.19	0.12	0.17
0.95	0.06	0.08	0.07	0.07
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.06	0.08	0.07	0.07
1.10	0.09	0.18	0.12	0.16
1.15	0.18	0.35	0.21	0.32
1.20	0.30	0.55	0.35	0.50
1.25	0.46	0.74	0.51	0.68
1.30	0.63	0.87	0.67	0.82
1.35	0.78	0.94	0.80	0.90
1.40	0.88	0.98	0.89	0.96
1.45	0.95	0.99	0.94	0.98
1.50	0.98	1.00	0.97	0.99

(g) $T = 200, l = 20, s = .2$

Altrnative	All	Our	HP	DN
0.50	1.00	1.00	1.00	1.00
0.55	1.00	1.00	1.00	1.00
0.60	1.00	1.00	1.00	1.00
0.65	1.00	1.00	1.00	1.00
0.70	0.99	1.00	1.00	1.00
0.75	0.91	0.99	0.97	0.98
0.80	0.70	0.94	0.86	0.90
0.85	0.39	0.74	0.59	0.67
0.90	0.17	0.39	0.27	0.34
0.95	0.07	0.13	0.10	0.11
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.07	0.12	0.09	0.11
1.10	0.15	0.36	0.26	0.31
1.15	0.33	0.68	0.52	0.61
1.20	0.59	0.89	0.78	0.83
1.25	0.82	0.98	0.93	0.95
1.30	0.95	1.00	0.99	0.99
1.35	0.99	1.00	1.00	1.00
1.40	1.00	1.00	1.00	1.00
1.45	1.00	1.00	1.00	1.00
1.50	1.00	1.00	1.00	1.00

(h) $T = 200, l = 20, s = .3$

Altrnative	All	Our	HP	DN
0.50	1.00	1.00	1.00	1.00
0.55	1.00	1.00	1.00	1.00
0.60	1.00	1.00	1.00	1.00
0.65	1.00	1.00	1.00	1.00
0.70	0.98	1.00	1.00	0.99
0.75	0.90	0.99	0.97	0.97
0.80	0.67	0.92	0.86	0.87
0.85	0.37	0.70	0.60	0.61
0.90	0.16	0.36	0.28	0.30
0.95	0.08	0.11	0.09	0.10
1.00	***0.05***	***0.05***	***0.05***	***0.05***
1.05	0.07	0.11	0.09	0.10
1.10	0.15	0.33	0.26	0.28
1.15	0.32	0.63	0.54	0.55
1.20	0.56	0.86	0.78	0.79
1.25	0.80	0.96	0.93	0.93
1.30	0.93	0.99	0.98	0.98
1.35	0.98	1.00	1.00	0.99
1.40	1.00	1.00	1.00	1.00
1.45	1.00	1.00	1.00	1.00
1.50	1.00	1.00	1.00	1.00

All : Using all instruments
Our : using instruments selected by our method
HP : using instruments selected by Hall-Piexe method
DN : using instruments selected by Donal-Newey method

Table 13 : Comparison of different selection methods
2SLS with non normal distubances

	T	All instruments	Our methdod	Hall-Piexe	Donald-Newey
			$l = 10$		
Bias	50	1.304	-0.021	0.437	0.787
	100	0.691	0.080	0.222	0.472
	200	0.354	-0.006	0.056	0.220
MSE	50	0.655	0.636	0.659	0.670
	100	0.304	-	0.290	0.308
	200	0.184	-	0.179	0.186
MAE	50	6.465	6.586	6.525	6.541
	100	4.402	4.419	4.408	4.419
	200	3.423	3.434	3.433	3.435
			$l = 20$		
Bias	50	2.268	-0.013	0.971	0.780
	100	1.569	-0.058	0.379	0.511
	200	0.750	0.050	0.175	0.290
MSE	50	0.559	0.606	0.547	0.561
	100	0.370	-	0.373	0.375
	200	0.161	-	0.164	0.162
MAE	50	6.007	5.987	5.929	5.994
	100	4.889	4.899	4.882	4.906
	200	3.202	3.187	3.185	3.199

Table 14: Empirical level of the t-stat in high-LR instrument samples
Nominal size is 10%, $T = 100$, 10000 replications

		Without split-sample		Fraction of the sample used for relevance testing			
				20%		30%	
Relevance $cov(x,z)$	Endogeneity $cov(x,u)$	LR test for relevance	level of t-test	LR test for relevance	Level of t-test	LR test for relevance	Level of t-test
0	0	0.10	0.01	0.11	0.00	0.12	0.00
0	0.1	0.10	0.02	0.11	0.00	0.11	0.00
0	0.2	0.10	0.04	0.12	0.01	0.11	0.01
0	0.3	0.10	0.10	0.11	0.01	0.11	0.01
0	0.4	0.10	0.17	0.12	0.02	0.11	0.02
0	0.5	0.10	0.29	0.11	0.04	0.10	0.04
0	0.6	0.10	0.45	0.11	0.07	0.11	0.08
0	0.7	0.11	0.63	0.11	0.11	0.11	0.11
0	0.8	0.11	0.82	0.12	0.19	0.11	0.20
0	0.9	0.10	0.95	0.11	0.35	0.10	0.31
0.1	0	0.27	0.02	0.15	0.01	0.16	0.00
0.1	0.1	0.27	0.02	0.14	0.01	0.16	0.00
0.1	0.2	0.27	0.04	0.14	0.01	0.17	0.01
0.1	0.3	0.26	0.08	0.15	0.02	0.17	0.02
0.1	0.4	0.27	0.13	0.15	0.03	0.16	0.03
0.1	0.5	0.26	0.18	0.14	0.05	0.16	0.06
0.1	0.6	0.27	0.26	0.15	0.08	0.16	0.08
0.1	0.7	0.27	0.33	0.15	0.11	0.17	0.12
0.1	0.8	0.26	0.44	0.15	0.17	0.16	0.16
0.1	0.9	0.27	0.56	0.15	0.21	0.16	0.21
0.2	0	0.65	0.03	0.24	0.02	0.30	0.02
0.2	0.1	0.65	0.04	0.25	0.02	0.31	0.02
0.2	0.2	0.64	0.05	0.25	0.02	0.30	0.02
0.2	0.3	0.65	0.06	0.25	0.03	0.30	0.03
0.2	0.4	0.66	0.09	0.25	0.06	0.31	0.05
0.2	0.5	0.65	0.11	0.25	0.06	0.31	0.07
0.2	0.6	0.65	0.14	0.24	0.09	0.30	0.09
0.2	0.7	0.64	0.17	0.25	0.12	0.31	0.12
0.2	0.8	0.65	0.18	0.25	0.13	0.31	0.14
0.2	0.9	0.64	0.20	0.25	0.13	0.30	0.15
0.3	0	0.93	0.05	0.42	0.04	0.54	0.04
0.3	0.1	0.93	0.05	0.41	0.04	0.53	0.04
0.3	0.2	0.93	0.06	0.42	0.06	0.53	0.04
0.3	0.3	0.93	0.07	0.41	0.05	0.52	0.06
0.3	0.4	0.92	0.07	0.41	0.07	0.53	0.06
0.3	0.5	0.93	0.09	0.41	0.07	0.53	0.08
0.3	0.6	0.92	0.09	0.41	0.09	0.53	0.09
0.3	0.7	0.92	0.10	0.41	0.11	0.53	0.10
0.3	0.8	0.93	0.11	0.42	0.11	0.52	0.11
0.3	0.9	0.92	0.12	0.41	0.11	0.52	0.12
0.4	0	1.00	0.07	0.61	0.07	0.75	0.06
0.4	0.1	1.00	0.07	0.61	0.06	0.75	0.06
0.4	0.2	0.99	0.08	0.61	0.07	0.75	0.06
0.4	0.3	1.00	0.08	0.61	0.07	0.75	0.08
0.4	0.4	0.99	0.08	0.61	0.08	0.75	0.07
0.4	0.5	0.99	0.09	0.62	0.08	0.75	0.08
0.4	0.6	1.00	0.08	0.60	0.08	0.75	0.09
0.4	0.7	1.00	0.09	0.60	0.09	0.75	0.10
0.4	0.8	1.00	0.10	0.60	0.10	0.75	0.10
0.4	0.9	0.99	0.10	0.60	0.10	0.75	0.12

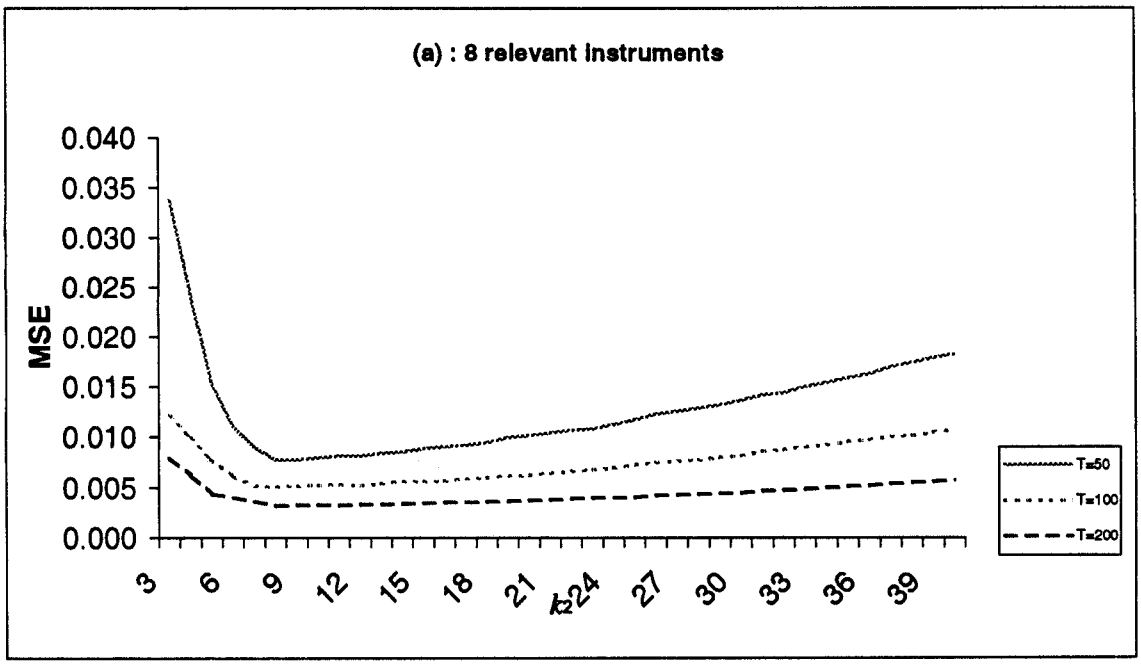
Table 15. 2SLS estimation of the coefficient of the returns to scale with selected instruments

	2SLS using all instruments		2SLS using selected instruments		
Sector *	Estimate	s.d.	Selected instruments **	estimate	s.d.
7	0.90	0.22	1	1.21	0.27
8	0.79	0.21	6	0.77	0.25
9	0.81	0.81	6	0.5	0.26
10	1.05	0.10	3 and 4	1.03	0.1
11	0.80	0.14	6	0.38	0.36
12	1.12	0.07	6 and 4	1.05	0.08
13	1.08	0.12	4	1.15	0.16
14	1.15	0.12	4	1.15	0.14
15	1.07	0.18	4	1.12	0.25
16	0.62	0.20	1	1.06	0.37
17	1.08	0.09	1	1.22	0.16
18	0.40	0.23	3	.35	0.3
19	1.19	0.07	6	1.15	0.11
20	0.93	0.05	6	0.96	0.06
21	1.04	0.08	6 and 4	1.05	0.08
22	1.17	0.07	4	1.17	0.1
23	1.21	0.08	1	1.36	0.12
24	1.20	0.05	6	1.09	0.11
25	1.11	0.06	4	1.14	0.09
26	1.32	0.11	4	1.38	1.38
27	1.06	0.16	4	1.21	0.32

* The sectors considered are: 7:Food and kindred products; 8:Tobacco; 9:Textile mill products; 10:Apparel; 11:Lumber and wood; 12:Furniture and fixtures; 13:Paper and allied; 14:Printing; publishing; and allied; 15:Chemicals; 16:Petroleum and coal products; 17:Rubber and misc. plastics; 18:Leather; 19:Stone, clay, glass; 20:Primary metal; 21:Fabricated metal; 22:Machinery, non-electrical; 23:Electrical machinery; 24:Motor vehicles; 25:Transportation equipment & ordnance; 26:Instruments; 27:Misc. manuf.

** The instruments are: 1: growth rate of military purchases; 2: the growth rate of world oil price; 3: a dummy variable representing the political party of the president of Unites States; 4: One lag of the growth rate of military purchases; 5: one lag of the growth rate of world oil price; 6: one lag of a dummy variable representing the political party of the president of Unites States

Figure 1 : MSE of 2SLS



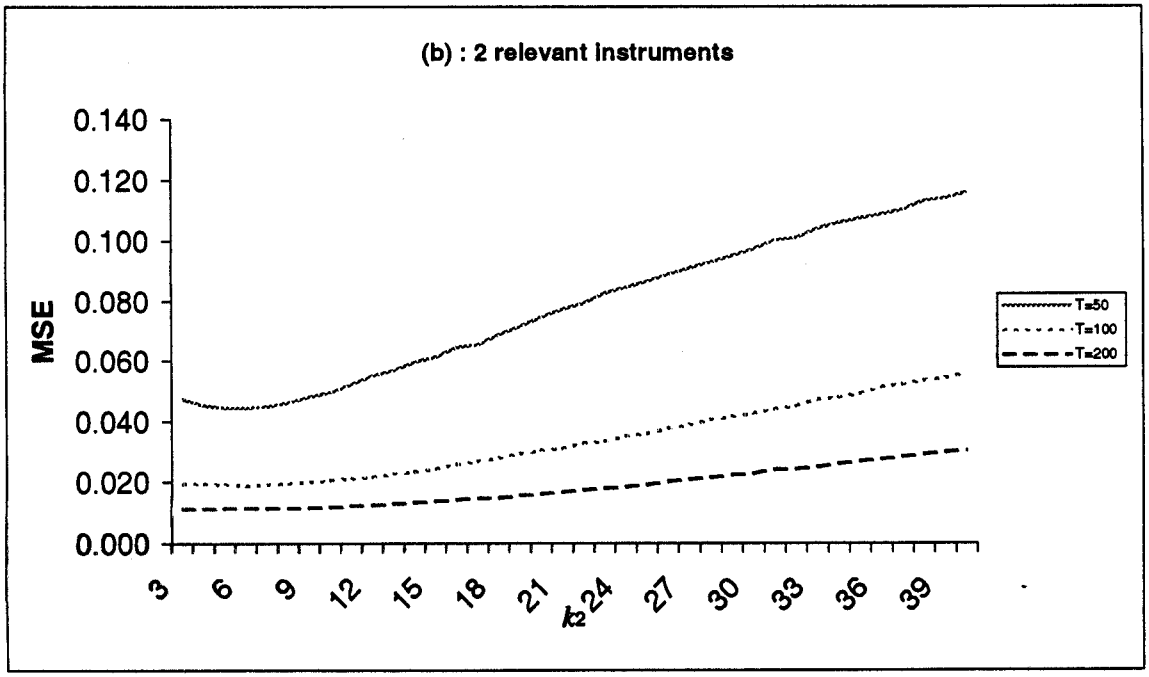
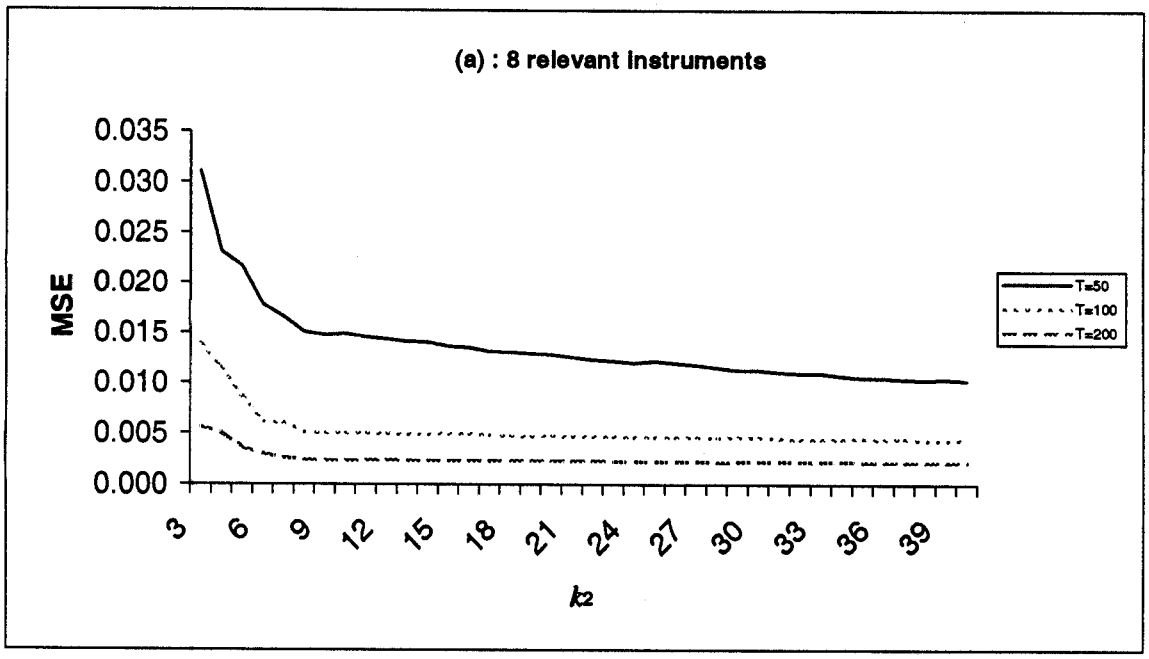


Figure 2 : MSE of 2SLS when the regressors are exogenous



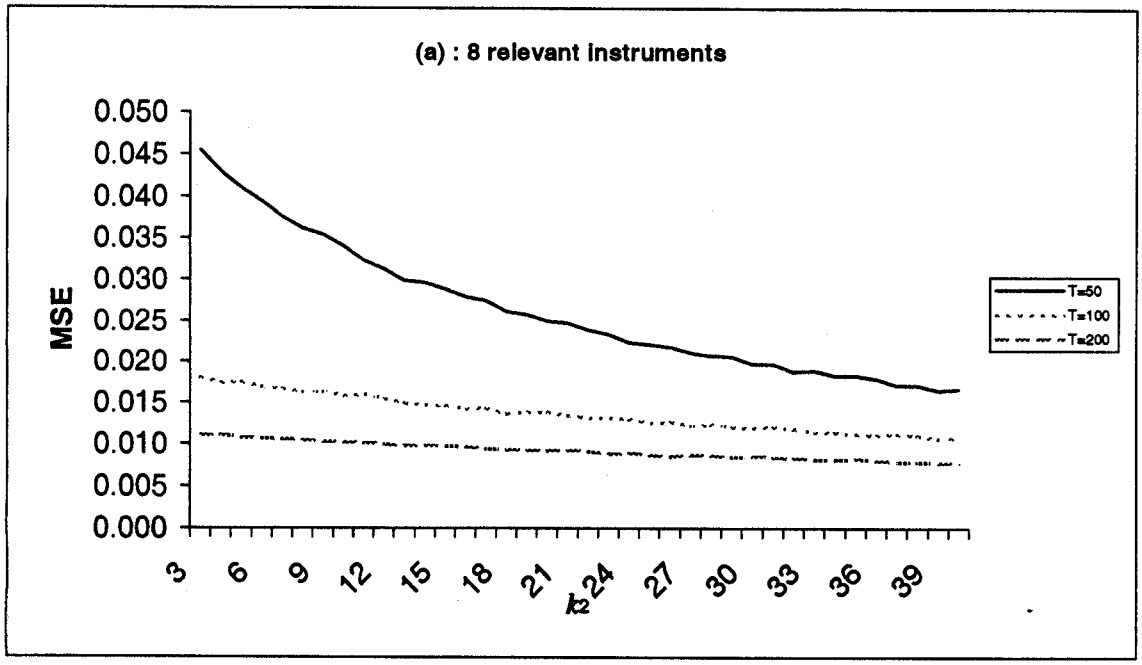
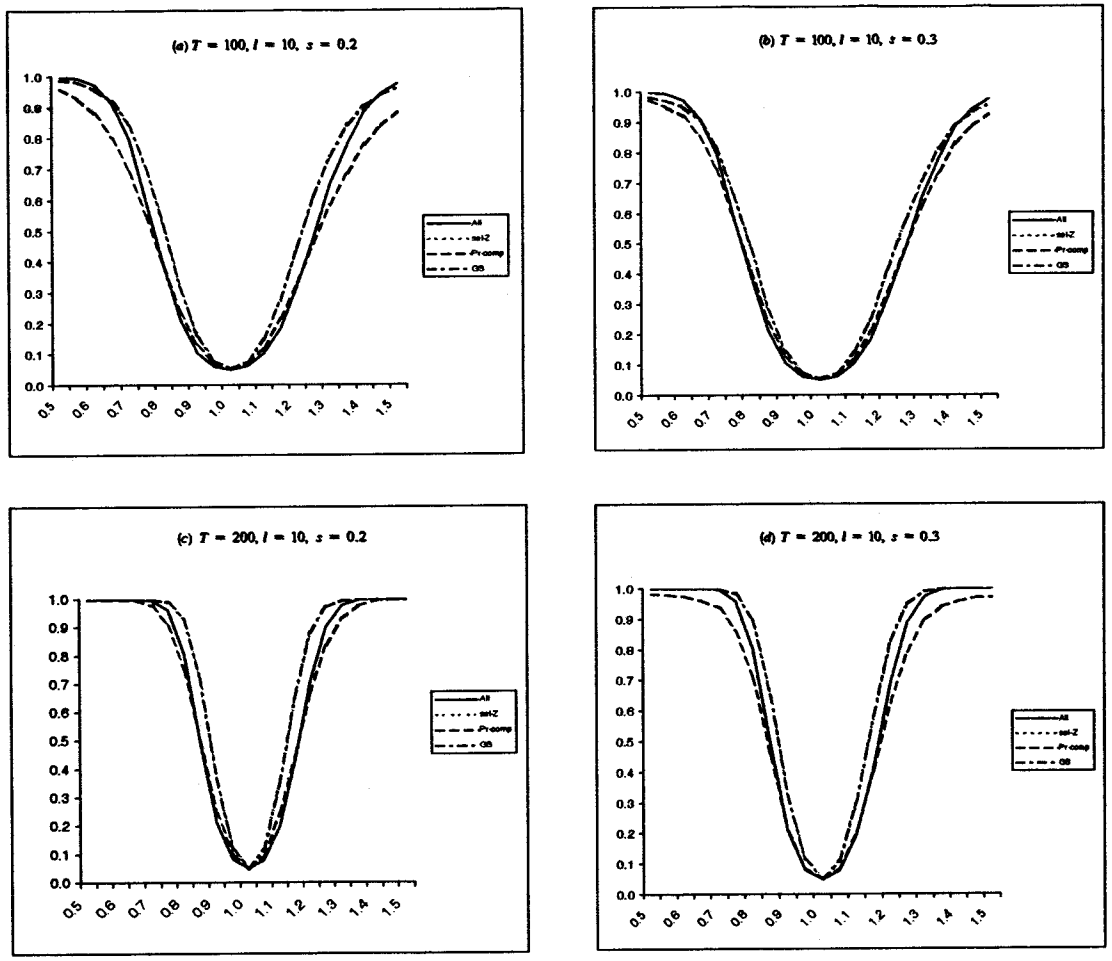
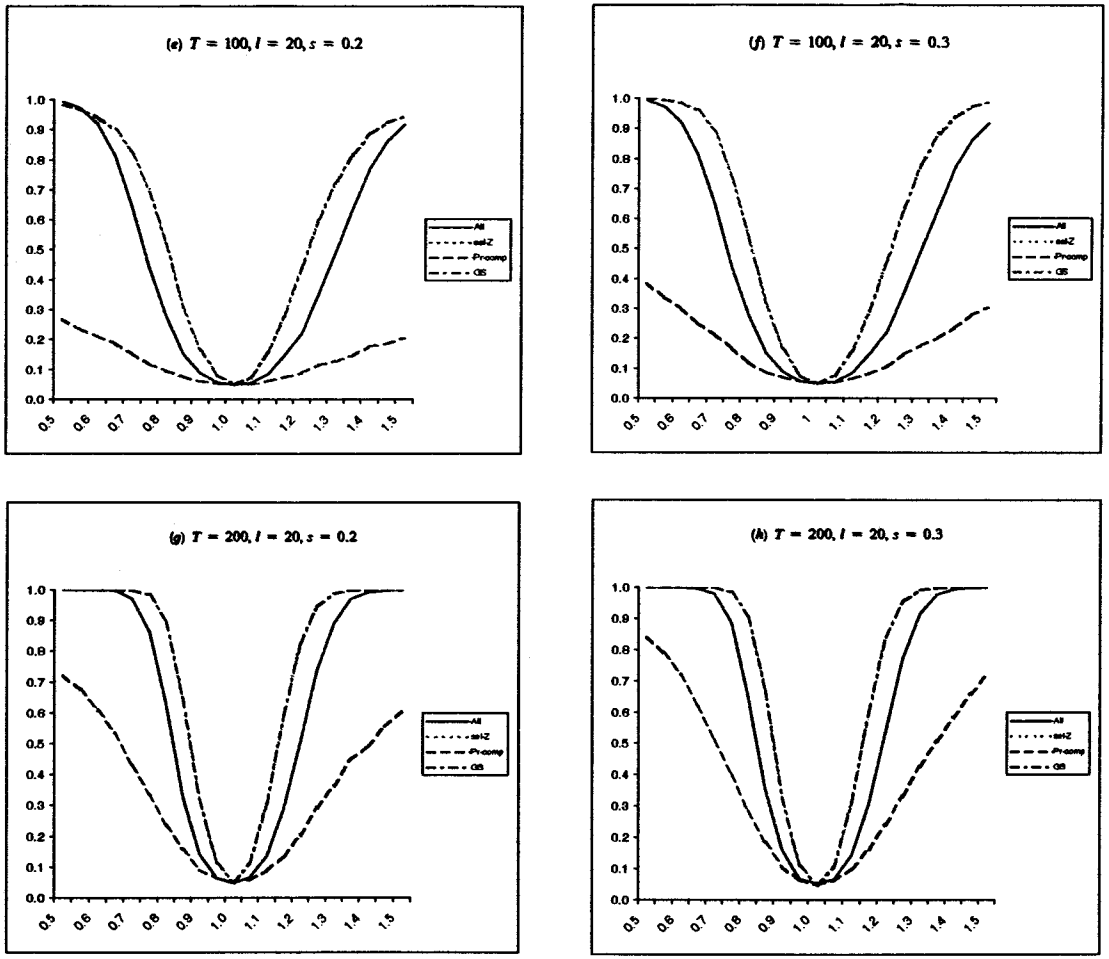


Figure 3 : Power of the *AR* test for H_0: $\beta = 1$ in model (7.1) − (7.2)



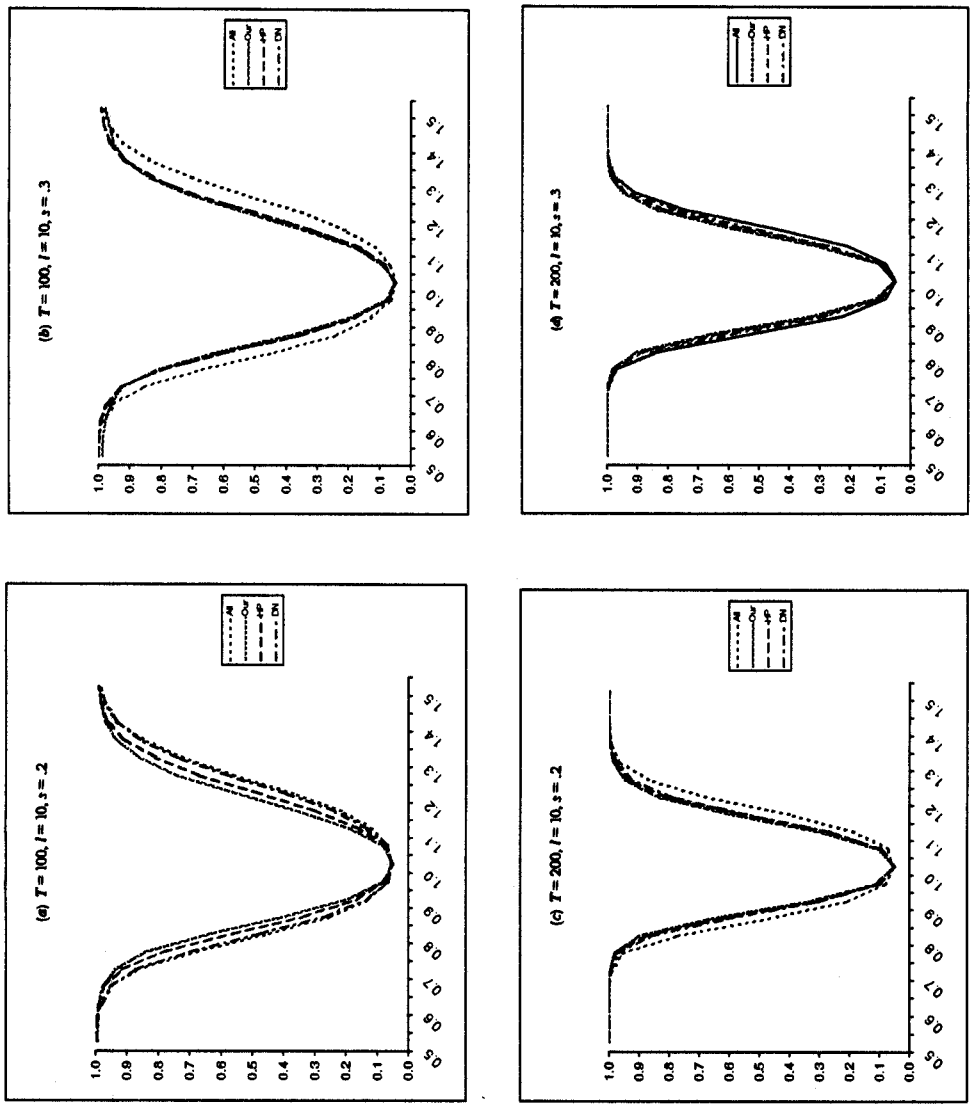
All: AR test using all the potential instruments
Z-sel: AR test with selected instruments among the original ones
Pr-comp.: AR test with selected instruments after principal components decomposition of the original set of instruments
GS: AR test with selected instruments after Gram-Schmidt decomposition of the original set of instruments

Figure 3 (cont´d): Power of the *AR* test for H_0: $\beta = 1$ in model (7.1) – (7.2)



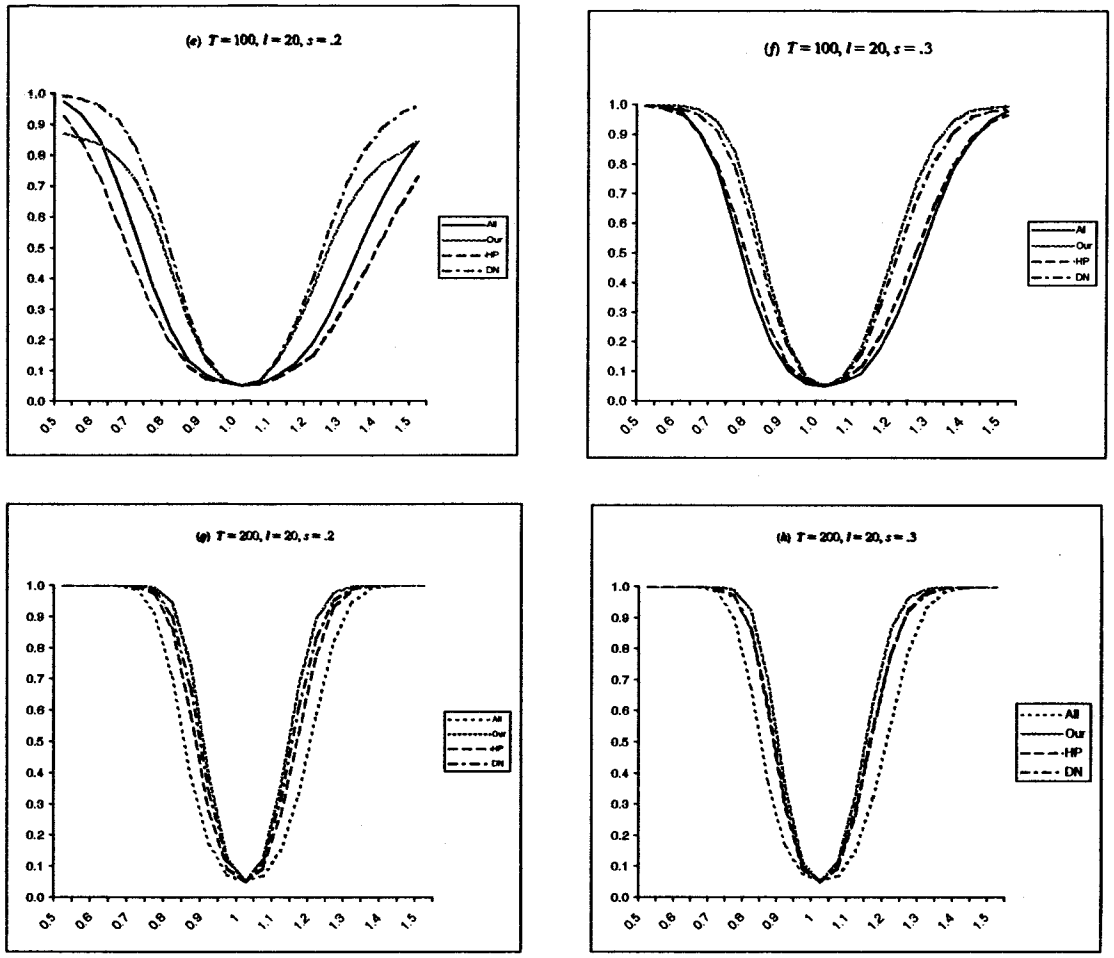
All: AR test using all the potential instruments
Z-sel: AR test with selected instruments among the original ones
Pr-comp.: AR test with selected instruments after principal components decomposition of the original set of instruments
GS: AR test with selected instruments after Gram-Schmidt decomposition of the original set of instruments

Figure 4 : Comparison of different selection methods
Power of the AR test for H_0 β=1 in model (7.1) – (7.2)



All: AR test using all the potential instruments
Z-sel: AR test with selected instruments among the original ones
Pr-comp.: AR test with selected instruments after principal components decomposition of the original set of instruments
GS: AR test with selected instruments after Gram-Schmidt decomposition of the original set of instruments

Figure 4 (cont'd): Comparison of different selection methods
Power of the *AR* test for H_0 β=1 in model (7.1) – (7.2)



All: AR test using all the potential instruments
Z-sel: AR test with selected instruments among the original ones
Pr-comp.: AR test with selected instruments after principal components decomposition of the original set of instruments
GS: AR test with selected instruments after Gram-Schmidt decomposition of the original set of instruments

References

ABDELKHALEK, T., AND J.-M. DUFOUR (1998): "Statistical Inference for Computable General Equilibrium Models, with Application to a Model of the Moroccan Economy," *Review of Economics and Statistics*, LXXX, 520–534.

AMEMIYA, T. (1974): "The Nonlinear Two-Stage Least-Squares Estimator," *Journal of Econometrics*, 2, 105–110.

——— (1975): "The Nonlinear Limited-Information Maximum Likelihood Estimator and the Modified Nonlinear Two-Stage Least Squares Estimator," *Journal of Econometrics*, 3, 375–386.

——— (1977): "The ML and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equations Model," *Econometrica*, 45, 955–968.

ANDERSON, T. W. (1984): *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, New York, second edn.

ANDERSON, T. W., AND H. RUBIN (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. (1999): "Consistent Moment Selection Procedures for Generalized Method of Moment Estimation," *Econometrica*, 67, 543–564.

ANGRIST, J. D., G. W. IMBENS, AND A. B. KRUEGER (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57–67.

ANGRIST, J. D., AND A. B. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earning?," *Quarterly Journal of Economics*, CVI, 979–1014.

ANGRIST, J. D., AND A. B. KRUEGER (1995): "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13, 225–235.

ARNOLD, S. F. (1981): *The Theory of Linear Models and Multivariate Analysis.* John Wiley & Sons, New York.

BASU, S., AND J. G. FERNALD (1995): "Are Apparent Productive Spillovers a Figment of Specification Error," *Journal of Monetary Economics*, 36, 165–188.

——— (1997): "Returns to Scale in U. S. Production : Estimates and Implications," *Journal of Political Economy*, 105(2), 249–283.

BOOT, J. C. G. (1964): *Quadratic Programming : Algorithms-Anomalies-Applications.* North-Holland.

BOUND, J., D. A. JAEGER, AND R. BAKER (1993): "The Cure can be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables," Technical Working Paper 137, National Bureau of Economic Research, Cambridge, MA.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450.

BOWDEN, R. J., AND D. A. TURKINGTON (1984): *Instrumental Variables*, Econometric Society Monographs in Quantitative Economics. Cambridge University Press, Cambridge.

BREUSCH, T., Q. T. S. P., AND D. WYHOWSKI (1999): "Redundancy of Moment Conditions," *Journal of Econometrics*, 91, 89–111.

BURNSIDE, C. (1996): "Production Function Regressions, Returns to Scale, and Externalities," *Journal of Monetary Economics*, 37, 177–201.

BUSE, A. (1992): "The Bias of Instrumental Variables Estimators," *Econometrica*, 60, 173–180.

CABALLERO, R. J., AND R. K. LYONS (1989): "The Role of External Economies in US Manufacturing," Discussion Paper 3033, NBER.

——— (1992): "External Effects in U.S. Procyclical Productivity," *Journal of Monetary Economics*, 29, 209–225.

CHAO, J., AND N. R. SWANSON (2000): "Bias and MSE of the IV Estimators under Weak Identification," Discussion paper, Department of Economics, University of Maryland.

CHOI, I., AND P. C. B. PHILLIPS (1992): "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics*, 51, 113–150.

CRAGG, J. G., AND S. G. DONALD (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford University Press, New York.

——— (1998): "Artificial Regressions," Discussion paper, Departement of Economics, Queen's University.

DONALD, S. G., AND W. K. NEWEY (Forthcoming): "Choosing the Number of Instruments," *Econometrica*.

DUFOUR, J.-M. (1987): "Linear Wald Methods for Inference on Covariances and Weak Exogeneity Tests in Structural Equations," in *Advances in the Statistical Sciences: Festschrift in Honour of Professor V.M. Joshi's 70th Birthday. Volume III, Time Series and Econometric Modelling*, ed. by I. B. MacNeill, and G. J. Umphrey, pp. 317–338. D. Reidel, Dordrecht, The Netherlands.

——— (1989): "Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions," *Econometrica*, 57, 335–355.

——— (1990): "Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica*, 58, 475–494.

——— (1997): "Some Impossibility Theorems in Econometrics, with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1389.

DUFOUR, J.-M., AND J. JASIAK (1993): "Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors," Discussion paper, C.R.D.E., Université de Montréal, 38 pages.

DUFOUR, J.-M., AND J. JASIAK (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, forthcoming.

DUFOUR, J.-M., AND M. L. KING (1991): "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors," *Journal of Econometrics*, 47, 115–143.

DUFOUR, J.-M., AND J. F. KIVIET (1998): "Exact Inference Methods for First-Order Autoregressive Distributed Lag Models," *Econometrica*, 66, 79–104.

DUFOUR, J.-M., AND M. TAAMOUTI (2000a): "Point-Optimal Instruments and Generalized Anderson-Rubin Procedures for Nonlinear Models," Discussion paper, Université de Montréal.

——— (2000b): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," Discussion paper, C.R.D.E., Université de Montréal.

DUFOUR, J.-M., AND O. TORRÈS (1998): "Union-Intersection and Sample-Split Methods in Econometrics with Applications to SURE and MA Models," in *Handbook of Applied Economic Statistics*, ed. by D. E. A. Giles, and A. Ullah, pp. 465–505. Marcel Dekker, New York.

——— (2000): "Markovian Processes, Two-Sided Autoregressions and Exact Inference for Stationary and Nonstationary Autoregressive Processes," *Journal of Econometrics*, 99, 255–289.

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–836.

FRANKEL, J. A., AND D. ROMER (1996): "Trade and Growth: An Empirical Investigation," Discussion Paper 5476, National Bureau of Economic Research, Cambridge, Massachusetts.

——— (1999): "Does Trade Cause Growth?," *The American Economic Review*, 89(3).

GALLANT, A. R. (1987): *Nonlinear Statistical Models.* John Wiley & Sons, New York.

HAHN, J., AND J. HAUSMAN (Forthcoming): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica.*

HALL, A. R., AND F. P. M. PEIXE (2000): *A Consistent Method for the Selection of Relevant Instruments. Discussion Paper.*North Carolina State Unoversity.

HALL, A. R., G. D. RUDEBUSCH, AND D. W. WILCOX (1996): "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283–298.

HALL, R. E. (1990): "Invariance Properties of Solow's Productivity Residual," in *Growth/Productivity/Employment*, ed. by P. Diamond, pp. 71–112. MIT press.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HARRISON, A. (1996): "Openness and Growth: A Time-Series, Cross-Country Analysis for Developing Countries," *Journal of Developement Economics*, 48, 419–447.

HARTLEY, H. O. (1964): "Exact Confidence Regions for the Parameters in Nonlinear Regression Laws," *Biometrika*, 51, 347–353.

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1272.

JOHNSON, N. L., S. KOTZ, AND N. BALAKRISHNAN (1994): *Continuous Univariate Distributions.* John Wiley, New York.

JOHNSTON, J. (1972): *Econometric Methods.* McGraw-Hill.

JORGENSON, D. W., F. M. GOLLOP, AND B. M. FRAUMENI (1987): *Productivity and U.S. Economic Growth.* Harvard University Press.

KING, M. L. (1988): "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218, Comments and reply, 219-255.

KLEIBERGEN, F. (2001a): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," Discussion paper, Tinbergen Institute.

KLEIBERGEN, F. (2001b): "Testing Subsets of Structural Coefficients in the IV Regression Model," Discussion paper, Department of Quantitative Economics, University of Amsterdam.

LEHMANN, E. L. (1986): *Testing Statistical Hypotheses, 2nd edition.* John Wiley & Sons, New York.

MACKINNON, J. G. (1992): "Model Specification Tests and Artificial Regressions," *Journal of Economic Litterature*, 30, 102–46.

MADDALA, G. S. (1974): "Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models," *Econometrica*, 42, 841–851.

MADDALA, G. S., AND J. JEONG (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181–183.

MAGDALINOS, M. A. (1994): "Testing Instrument Admissibility: Some Refined Asymptotic Results," *Econometrica*, 62(2), 373–403.

MANKIW, G. N., D. ROMER, AND D. N. WEIL (1992): "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107(2), 407–37.

MARIANO, R. S. (1982): "Analytical Small-Sample Distribution Theory in Econometrics : The Simultaneous-Equations Case," *International Economic Review*, 23(3), 503–533.

MÁTYÁS, L. (ed.) (1999): *Generalized Method of Moments Estimation.* Cambridge University Press, Cambridge, U.K.

NELSON, C. R., AND R. STARTZ (1990a): "The Distribution of the Instrumental Variable Estimator and its t-ratio When the Instrument is a Poor One," *Journal of Business*, 63, 125–140.

——— (1990b): "Some Further Results on the Exact Small Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

NEWEY, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58(4), 809–837.

——— (1993): "Efficient Estimation of Models with Conditional Moment Retrictions," in *Handbook of Statistics, Vol 11*, ed. by C. R. R. G. S. Maddala, and H. D. Vinod. Elsevier Science.

NEWEY, W. K., AND K. D. WEST (1987): "Hypothesis Testing with Efficient Method of Moments Estimators," *International Economic Review*, 28, 777–787.

OWEN, A. D. (1976): "A Proof That Both the Bias and the Mean Square Error of the Two-Stage Least Squares Estimator are Monotonically Non-Increasing Functions of Sample Size," *Econometrica*, 44(2).

PERRON, B. (1999): "Semi-Parametric Weak Instrument Regressions with an Application to the Risk Return Trade-Off," Discussion paper, Université de Montréal.

PERRON, P., AND G. RODRIGUEZ (1998): "GLS Detrending, Efficient Unit Root Tests and Structural Change," Discussion paper, Department of Economics, Boston University, and CRDE, Université de Montréal.

PETTOFREZZO, A. J., AND M. L. MARCOANTONIO (1970): *Analytic Geometry with Vectors*. Scott, Fosman and Company.

PHILLIPS, P. C. B. (1980): "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing N+1 Endogenous Variables," *Econometrica*, 48(4).

PHILLIPS, P. C. B. (1983): "Exact Small Sample theory in the Simultaneous Equations Model," in *Handbook of Econometrics, Volume 1*, ed. by Z. Griliches, and M. D. Intrilligator, chap. 8, pp. 449–516. North-Holland, Amsterdam.

POTSCHER, B. M. (1983): "Order Estimation in ARMA-Models by Lagrange Multiplier Tests," *Annals of Statistics*, 11, 872–885.

REVANKAR, N. S., AND P. MALLELA (1972): "The Power Function of an F-Test in the Context of a Structural Equation," *Econometrica*, 40, 913–916.

RICHARDSON, D. H., AND D.-M. WU (1971): "A Note on the Comparaison of Ordinary and Two-Stage Least Squares Estimators," *Econometrica*, 39(6).

RODRIK, D. (1995): "Trade and Industrial Policy Reform," in *Handbook of Development Economics*, ed. by J. Behrman, and T. N. Srinivasan, vol. 3A. Elsevier Science, Amsterdam.

SAVIN, N. E. (1984): "Multiple Hypothesis Testing," in *Handbook of Econometrics, Volume 2*, ed. by Z. Griliches, and M. D. Intrilligator, pp. 827–879. North-Holland, Amsterdam.

SAWA, T. (1969): "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923–937.

——— (1972): "Finite-Sample Properties of the K-Class Estimators," *Econometrica*, 40(4).

SCHEFFÉ, H. (1959): *The Analysis of Variance.* John Wiley & Sons, New York.

SHEA, J. (1997): "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics*, LXXIX, 348–352.

SHILOV, G. E. (1961): *An Introduction to the Theory of Linear Spaces.* Prentice Hall, Englewood Cliffs, New Jersey.

SLATER, L. J. (1960): *Confluent Hypergeometric Functions.* Cambridge.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

STARTZ, R., C. R. NELSON, AND E. ZIVOT (1999): "Improved Inference for the Instrumental Variable Estimator," Discussion paper, Department of Economics, University of Washington.

STOCK, J. H., AND J. H. WRIGHT (2000): "GMM with Weak Identification," *Econometrica*, 68, 1097–1126, Econometrica.

TAYLOR, W. E. (1983): "On the Relevance of Finite Sample Distribution Theory," *Econometric Reviews*, 2, 1–39.

ULLAH, A., AND A. L. NAGAR (1974): "The Exact Mean of the Two-Stage Least Squares Estimator of the Structural Parameters in an Equation Having Three Endogenous Variables," *Econometrica*, 42(4).

WANG, J., AND E. ZIVOT (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389–1404.

WHITE, H. (1984): *Asymptotic Theory for Econometricians.* Academic Press, Orlando, Florida.

WRIGHT, J. H. (2000): "Detecting Lack of Identification in GMM," Discussion paper, Board of Governors of the Federal Reserve System.

WU, D.-M. (1973): "Alternative Tests of Independence between Stochastic Regressors and Disturbances," *Econometrica*, 41, 733–750.

ZIVOT, E., R. STARTZ, AND C. R. NELSON (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1144.

Conclusion générale

Dans cette thèse, nous avons développé des méthodes d'inférence exactes dans les modèles structurels. Ces méthodes sont particulièrement utiles au vu du problème des instruments faibles, sujet d'importance dans la recherche actuelle en économétrie. La thèse est composée de trois essais.

Dans le premier essai, nous avons considéré le problème d'inférence dans un modèle structurel à équations linéaires avec possibilité d'instruments faibles. L'objectif est de faire de l'inférence exacte sur des transformations du vecteur des paramètres inconnus. Ce problème est important parce que dans plusieurs cas, les méthodes usuelles permettent de tester des hypothèses spécifiant le vecteur entier des paramètres inconnus mais ne permettent pas de faire l'inférence sur des transformations de ce vecteur. Ce défaut caractérise en particulier le test d'Anderson-Rubin ainsi que plusieurs tests robustes aux problèmes des instruments faibles proposés récemment dans la littérature [Wang and Zivot (1998), Kleibergen (2001a), Stock and Wright (2000)]. Une solution générale à ce problème est la technique de projection décrite dans Dufour (1990, 1997). Cependant, l'application de cette technique est généralement compliquée et nécessite l'utilisation de méthodes numériques. Dans cet essai, nous avons dérivé une solution analytique complète au problème de construction d'intervalles de confiance basés sur la technique de projection à partir de régions de confiance obtenues par inversion de statistiques de type Anderson-Rubin [le test AR et les statistiques de type LR et LM proposées par Wang and Zivot (1998)]. La solution est fournie pour les intervalles de confiance pour les composantes du vecteur des paramètres inconnus et pour des combinaisons linéaires de ces composantes. Nous avons tout d'abord étudié les caractéristiques des régions de confiances basées sur l'inversion des statistiques de type AR sur la base de la géométrie des quadriques [voir Shilov (1961) et Pettofrezzo and Marcoantonio (1970)]. Nous avons montré que l'inversion de ces statistiques se ramènent à une forme simple donnée par $\{\theta : \theta' A\theta + b'\theta + c \leq 0\}$. Ensuite nous avons dérivé de façon analytique les intervalles de confiance par projection. Les calculs exigés pour calculer ces projections sont aussi simple que ceux exigés par la méthode

des moindres carrés ordinaires. Cette solution qui peut être vue comme une extension des intervalles et des ellipsoïdes de confiance usuels, rend la technique de projection très attrayante, surtout dans des situations ou aucune autre solution n'a été proposée jusqu'à date dans la littérature. Nous avons aussi montré que les intervalles de confiance par projection peuvent être interprétés comme des intervalles de type Wald basés sur des estimateurs de type k-class, où les écarts-type sont corrigés en tenant compte du niveau de signification du test. Nous avons également fait le lien avec les régions de confiance "à la Scheffé" largement utilisés en analyse de variance. La technique de projection conduit à des intervalles de confiance conservateurs. Cependant, les simulations que nous avons effectuées montrent que le degré de conservatisme n'est pas exagéré et que les régions obtenues restent informatives même quand elle sont non bornées. Enfin, les résultats que nous avons développés dans cet essai sont valides pour tous modèle où les régions de confiance de la forme ci-dessus peuvent être rencontrées. Ces modèles inclus en particulier les modèles considérés dans Dufour and Jasiak (2001) et quelques modèles non linéaires considérés dans le deuxième essai de cette thèse.

Dans le deuxième essai, nous avons considéré le problème d'inférence statistique dans un modèle structurel non linéaire. Nous avons proposé une approche simple pour construire des tests statistiques exacts. Cette approche généralise l'approche de Hartley (1964) pour les tests d'hypothèses et la construction de régions de confiance dans les régressions non linéaires et la procédure d'Anderson-Rubin spécifique aux modèles structurels à équations linéaires. La première généralisation est faite dans le sens d'inclure des variables explicatives endogènes et la deuxième dans l'extension aux modèles non linéaires. Nous avons ensuite considéré le problème d'optimalité des instruments à utiliser dans la régression artificielle. La notion d'optimalité des instruments que nous avons considéré diffère de celle bien connue dans la littérature [Amemiya (1977)] et qui consiste à minimiser la variance asymptotique d'un estimateur. Dans cet essai, l'optimalité est définie par rapport à la maximisation de la puissance d'un test. Nous avons obtenu les instruments optimaux en calculant la fonction puissance du test et en étudiant sa monotonocité en fonction du nom-

bre et de la valeur des instruments. La dérivation de cette fonction généralise un vieux résultat de Revankar and Mallela (1972)]. Les instruments optimaux dépendent de la valeur de l'alternative, nous parlons d'instruments point-optimaux, [voir King (1988)]. La distribution exacte du test est dérivée sous les hypothèses de normalité et d'indépendance des perturbations, mais le test reste valide asymptotiquement sous des hypothèses similaires à celle habituellement considérées dans la théorie asymptotique. La matrice des instruments optimaux est inconnue de façon générale puisqu'elle dépend des paramètres inconnus et/ou de la distribution des variables endogènes. Nous avons proposé une méthode basée sur les technique de "split-sample" pour estimer cette matrice. Les simulations Monte Carlo montrent que le gain dans la puissance du test est considérable relativement aux tests de type GNR. Il serait intéressant d'essayer d'étendre cette approche pour l'inférence sur les paramètres de plusieurs équations du modèle ainsi qu'à des tests de spécification plus généraux.

Dans le troisième essai, nous avons considéré le problème de sélection des instruments dans le cadre d'un modèle structurel à équations linéaires où la méthode 2SLS pour l'estimation et/ou le test d'Anderson-Rubin sont utilisés. Nous avons d'abord étudié comment la matrice des instruments affecte la qualité de l'estimation et de l'inférence en recourant à plusieurs résultats de l'inférence exacte développés au cours des années 70 et début des années 80. Nous avons considéré par la suite la sélection d'instruments sur la base de ces résultats. Nous nous sommes concentré sur le cas d'une variable explicative endogène. Ce cas est intéressant parce qu'il est fréquemment rencontrés dans la recherche empirique. En plus, les distributions exactes des estimateurs et tests usuels sont relativement maniables. La méthode que nous avons proposée est basée sur la maximisation séquentielle du paramètre de concentration et est simple à appliquer dans la pratique. Les simulations Monte Carlo montrent que son utilisation permet d'améliorer la qualité de l'estimation (surtout en terme de réduction de biais) et de l'inférence par rapport à la méthode qui consiste à utiliser tous les instruments potentiels. Asymptotiquement, nous avons montré que la méthode est convergente dans le sens qu'elle sélectionne uniquement les instruments avec un coefficient non nul dans la vraie DGP de la variable explicative endogène. Enfin

sa comparaison avec les autres méthodes de la littérature montre qu'elle se comporte relativement bien. Les défis futures consistent à étendre cette méthode au cas de plusieurs variables explicatives endogènes et possiblement aussi aux modèles non linéaires.