UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ESTIMATEUR PAR VARIABLES INSTRUMENTALES : UNE APPROCHE ALTERNATIVE ?

MÉMOIRE

PRÉSENTÉ COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉCONOMIQUE
CONCENTRATION EN ÉCONOMIE FINANCIÈRE

PAR

BOCAR BA

OCTOBRE 2011

REMERCIEMENTS

Je tiens à remercier mon directeur Philip Merrigan pour la confiance qu'il m'a témoignée tout au long de mon parcours. Je lui suis infiniment reconnaissant pour sa patience, ses précieux conseils, sa compréhension et son soutien.

Je souhaiterais également remercier Pierre Lefebvre, Steve Ambler, Pierre-Carl Michaud, Anne-Catherine Faye, Lisa Pinheiro, Jimmy Royer, Georges Tanguay, Vadim Marmer, Ruthie Coffman et Victoria Miller pour leur enseignement et leur disponibilité pour répondre à mes interrogations.

Merci à Martine Boisselle pour sa patience et sa gentillesse. Je tiens également à souligner ma gratitude envers le département pour m´avoir laissé une certaine liberté dans le choix de mes cours.

Enfin, je remercie le Groupe d'Analyse, Pierre Lefebvre et Philip Merrigan pour le soutien financier.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

Tableaux Page

RÉSUMÉ

Ce texte propose une façon alternative d'obtenir l'estimateur à variables instrumentales. Dans un premier temps, un ensemble d'estimés candidats est obtenu à l'aide d'une estimation par méthode des moments généralisés. Puis, l'estimateur alternatif est obtenu en combinant les estimés candidats par l'intermédiaire d'une matrice de poids optimale. Les hypothèses, concernant l'estimateur alternatif, garantissant l'efficacité et la normalité asymptotique sont également présentées. De plus, un test permettant de vérifier la validité des instruments est suggéré. Les expériences de Monte Carlo montrent que le biais de l'estimateur proposé augmente avec le nombre d'estimés candidats. Finalement, l'écart type moyen de l'estimateur proposé est plus petit que l'estimateur à variables instrumentales standard quand les instruments ne sont pas faibles. Lorsque la gamme des estimateurs possibles s'élargit, l'estimateur alternatif démontre de meilleures performances que l'estimateur standard dans le cas d'une endogénéité qui est faible ou moyenne.

Les propriétés de puissance dans le cadre de l'alternative de type Pitman-local sont également suggérées. Les simulations de Monte Carlo montrent que la puissance du test en grand échantillon semble satisfaisante au niveau de la taille et de la puissance. Cependant il se peut que ce test souffre d'un biais en petit échantillon.

Finalement, en utilisant l'approche d'Angrist et Krueger (1991) pour l'application empirique, les estimateurs alternatif et standard donnent des résultats similaires; cependant la méthode proposée donne des écart-types plus faibles.

Mots clés : alternative de type Pitman-local, fonction de puissance, méthode des moments généralisés, minimisation de la distance simulation de Monte Carlo, test de Wald, variables instrumentales, tests d'hypothèses.

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

EQM	Erreur Quadratique Moyenne
k	Nombres de régresseurs
l	Nombres d'instruments
LFGN	Loi Faible des Grands Nombres
MMG	Méthode des Moments Généralisés
MCO	Moindres Carrés Ordinaires
n	Nombre d'observations
TSLS	Moindres Carrés à Deux Étapes
TCL	Théorème Central Limite
$\rightarrow_p$	Convergence en probabilite

INTRODUCTION

Ce texte propose une méthode alternative d'estimation de l'estimateur à variables instrumentales. Puisque la méthode des moments généralisés peut être considérée comme une généralisation de la procédure par variables instrumentales, l'idée est d'obtenir une gamme d´estimateurs candidats en résolvant un système MMG à équations multiples, au lieu d'utiliser plusieurs équations séparées. L´estimateur alternatif est ainsi obtenu en combinant ces estimateurs candidats à l´aide d´une matrice de poids. En somme, l´estimateur proposé est donné par la combinaison linéaire des estimateurs candidats, préalablement obtenus grâce à la résolution du système MMG.

La méthodologie proposée pour obtenir la matrice de poids optimale, l´efficacité et la normalité asymptotique $\sqrt{n}$ de cet estimateur alternatif repose sur Chen, Jacho-Chávez et Linton (2009). La matrice de poids optimal est construite à l'aide de l'estimation par minimisation de la distance proposée par Rothenberg (1973). En effet, la matrice de poids optimal est telle que la distance quadratique, entre le véritable paramètre et chaque estimateur de la gamme, est minimisée, compte tenu de la matrice de variance covariance de chaque candidat. Cette approche permet de tester les conditions de moments. Compte tenu de la gamme des estimateurs, la validité des instruments peut être testée conjointement. Les propriétés de puissance de la statistique de Wald sont examinées sous l'alternative de type Pitman-local (1979) où il est pris en compte que la distance mesurée à partir du véritable paramètre rétrécit à mesure que la taille de l´échantillon augmente.

De nombreux exemples de combinaisons d'estimateurs ont été fournis dans la littérature. Sawa (1973) a proposé de combiner linéairement les estimateurs MCO et TSLS afin de réduire le biais en petit échantillon des estimateurs TSLS. L´estimateur Bagging (agrégation bootstrap) de Breiman (1996a, 1996b) est donné par la moyenne des estimateurs calculés par ré-échantillonnage bootstrap. Yang (2004) suggère une manière de combiner des procédures de prévision afin d'obtenir de meilleures performances par rapport aux méthodes initiales; la méthode proposée prend en compte le compromis entre le gain potentiel et le prix de la combinaison. Keller et Olkin (2004) proposent une méthode pour combiner plusieurs estimateurs sans biais à l'aide de matrices de poids qui sont des fonctions de la variance-covariance de chaque estimateur.

L'alternative de type Pitman-local est introduite dans le but d'examiner les propriétés de puissance de la statistique de Wald lorsque la taille de l'échantillon varie; puisque avec l'alternative fixe, la puissance est plate lorsque la taille de l'échantillon augmente. Pesaran (1982) montre que dans l'alternative de type Pitman-local, le test Cox non imbriqué et le test J proposé par Davidson et MacKinnon (1981) ont plus de puissance au niveau local que le test F traditionnel, alors que ces trois tests présentent des tailles correctes et consistent sous une hypothèse alternative fixe. En considérant l'alternative de type Pitman-local et des séquences de création de données, Davidson et MacKinnon (1987) montre que le statistiques LR, Wald et LM sont asymptotiquement équivalentes; d'ailleurs, ils soulignent aussi qu'il y a une "hypothèse alternative implicite " qui donne la puissance la plus élevée pour toute statistique qui suit asymptotiquement une distribution Chi-Carré sous l'hypothèse nulle. En introduisant la fonction de puissance inverse, Andrews (1989) fournit une procédure à appliquer dans la pratique afin de déterminer l'alternative de type Pitman-local. Un autre exemple est fourni par Chen et Fan (1999) où ils déterminent la distribution asymptotique sous l'hypothèse nulle et l'alternative de type Pitman-local en invoquant le Théorème de Limite Central pour les réseaux aléatoires de valeur Hilbert dépendante.

Cet article est organisé comme suit. La section 2 résume l'approche standard. La section 3 présente l'estimation de Chen, Jacho-Chávez et Linton (2009). Les hypothèses pour obtenir l'efficacité et la normalité $\sqrt{n}$-asymptotique de l'estimateur alternatif en grand échantillon sont également présentées dans cette section. Les tests d'hypothèses et la puissance sous l'alternative de type Pitman-local sont fournies dans la section 4. L'article d'Angrist et Krueger (1991) est utilisé en tant qu'application empirique pour la procédure alternative dans la section 5. Les expériences Monte-Carlo sont réalisées dans la dernière section pour vérifier le comportement de l'estimateur proposé dans les cas d'instruments forts et faibles. La taille de la statistique de Wald et sa puissance sous l'alternative de type Pitman-local sont également calculées par des expériences de Monte Carlo. Les preuves sont données dans l'annexe.

CHAPITRE I

L'APPROCHE STANDARD

Supposons que les données $\{W_i : i \leq n\}$ sont indépendamment et identiquement distribuées. Il existe un unique $\theta_0 \in \Theta^1 \subseteq \mathbf{R}^k$ tel que les conditions de moments sur les résidus sont données par

$$E[f(w_i, \theta_0)|X_i] = 0.$$

Où $W_i = \{Y_i, X_i^T\}^T$, $f: \mathbf{R}^p \times \mathbf{R}^k \to \mathbf{R}^l$ et $f(\cdot)$ est le vecteur des résidus. Considérons Z_i le vecteur composé des instruments avec $l \geq k$. Les autres conditions de moments sont

$$E[Z_i f(w_i, \theta_0)] = E[g(w_i, \theta_0)] = 0. \quad (1.1)$$

Soit A_n la matrice de poids qui est $l \times l$ avec $A_n \to_p A$ où A est non aléatoire de rang *l*. Alors l'estimateur MMG $\hat{\theta}$ est donné par

$$\hat{\theta}(A_n) = \operatorname{argmin}_{\theta} \left\| A_n n^{-1} \sum_{i=1}^{n} g(w_i, \theta) \right\|^2. \quad (1.2)$$

Avec $A'A$ est définie positive. Parmi la classe d'es estimateurs MMG, l'estimateur efficient est celui avec la plus petite matrice variance covariance (asymptotiquement). Cet estimateur est fonction de sa matrice de poids tel que

$$A'_n A_n \rightarrow_p [E(g(w_i, \theta_0) g(w_i, \theta_0)')]^{-1} = \Omega^{-1}. \qquad (1.3)$$

D'un point de vue asymptotique, la borne inférieure pour les matrices variance covariance pour la classe des estimateurs MMG est

$$V = \left[\left(E \frac{\partial g(w_i, \theta)}{\partial \theta'} \right)' \Omega^{-1} \left(E \frac{\partial g(w_i, \theta)}{\partial \theta'} \right) \right]^{-1}. \qquad (1.4)$$

Pour plus de détails, voir Amemiya (1974), Hansen (1982) et Newey (1990, 1993).

---

[1] Sauf indication contraire, les notations suivantes sont utilisées. Soit I_l représente la matrice identité de dimension $l \times l$, et $\|z\| = \sqrt{tr(zz')}$ correspond à la norme euclidienne d'un vecteur ou d'une matrice z.

CHAPITRE II

L'APPROCHE ALTERNATIVE

2.1 Estimation: la méthode de Chen, Jacho-Chávez and Linton

La méthode alternative est basée sur l'estimateur de Chen, Jacho-Chávez et Linton (2009). Ils proposent une façon optimale de combiner plusieurs estimateurs potentiels $\tilde{\theta}^j$ (avec j=1,.., *J(n)*) en un estimateur $\tilde{\theta}$. Soit A_n^j, la matrice de poids de dimension *l* x *l* avec $A_n^j \rightarrow_p A^j$ où A^j est une matrice non aléatoire et de rang plein *l*. Dans ce cas, équation (1.1) peut être écrite dans le cas échantillonnal comme

$$n^{-1}\sum_{i=1}^{n} Z_i^j f(w_i,\theta) = o_p\left(n^{-1/2}\right). \tag{2.1.1}$$

Où chaque estimateur $\tilde{\theta}^j$ est donné par

$$\tilde{\theta}^j\left(A_n^j\right) = \mathbf{argmin}_{\tilde{\theta}^j}\left\| A_n^j n^{-1}\sum_{i=1}^{n} Z_i^j f(w_i,\theta)\right\|^2. \tag{2.1.2}$$

En combinant les estimateurs $\tilde{\theta}^j$, on obtient $\tilde{\theta}$ qui est donné par

$$\tilde{\theta} = \sum_{j=1}^{J(n)} \Lambda_{nj}\tilde{\theta}^j. \tag{2.1.3}$$

Où *J(n)* est le nombre d'estimateurs disponibles, le nombre optimal d'estimateurs devrait être fonction de la taille de l'échantillon. La matrice de poids Λ_{nj} doit satisfaire

$$\sum_{j=1}^{J(n)} \Lambda_{nj} = I_k. \tag{2.1.4}$$

Notons que $\tilde{\theta}^j$ sera efficace si A_n^j est choisit de tel sorte que $A'^j_n A^j_n = \left(\hat{H}^j\right)^{-1}$ où $n^{-1/2} \sum_{i=1}^{n} Z_i^j f(w_i, \theta) \sim N(0, H^j)$ par le TCL

2.2 Exemple

Considérons le cas des variables instrumentales linéaires, où les hypothèses suivantes sont considérées :

- $Y_i = X'_i \theta + U_i$ où $\theta \in \mathbf{R}^k$
- $\{(Y_i, X_i, Z_i, Z_i^j) : i = 1, \ldots, n\}$ sont iid
- $E(Z_i U_i) = 0$ et $E(Z_i^j U_i) = 0$
- **rang** $(E(Z_{1i} [X']_{1i}) = k$ et **rang** $(E(Z_{1i}^{j} [X']_{1i}) = k$
- $A_n \rightarrow_p A$, $A_n^j \rightarrow_p A$ et rang $l \geq$ k
- $EX_{r,s}^2 < \infty, EZ_{r,q}^2 < \infty$ pour tout s=1,…, k
- $EZ_{r,q}^{j2} < \infty$ pour tout q=1,...*l*

Sachant (2.1.2), $\hat{\theta}^{\mathrm{MMG}}$ et $\tilde{\theta}^{j,\mathrm{MMG}}$ peuvent être écrits comme

$$\hat{\theta}^{MMG} = \left(\sum_{i=1}^{n} X_i Z'_i (A'_n A_n) \sum_{i=1}^{n} Z_i X'_i \right)^{-1} \sum_{i=1}^{n} X_i Z'_i (A'_n A_n) \sum_{i=1}^{n} Z_i Y_i. \tag{2.2.1}$$

$$\tilde{\theta}^{j,MMG} = \left(\sum_{i=1}^{n} X_i Z'^j_i \left(A'^j_n A^j_n\right) \sum_{i=1}^{n} Z^j_i X'_i\right)^{-1} \sum_{i=1}^{n} X_i Z'^j_i \left(A'^j_n A^j_n\right) \sum_{i=1}^{n} Z^j_i Y_i. \qquad (2.2.2)$$

L'estimateur alternatif sera donc

$$\tilde{\theta}^{MMG} = \sum_{j=1}^{J} \Lambda_{nj}\, \tilde{\theta}^{j,MMG}. \qquad (2.2.3)$$

La matrice de poids sera trouvée en utilisant l'estimateur de minimisation de la distance proposé par Rothenberg (1973). L'idée est de trouvé la matrice de poids qui rendra $\tilde{\theta}^{MMG}$ et $\hat{\theta}^{MMG}$ asymptotiquement équivalents.

2.3 Estimation Proposée

Une approche différente de celle de Chen, Jacho-Chávez et Linton (2009) peut-être utilisée afin de trouver des résultats similaires. La gamme d'estimés potentiels peut-être obtenue en utilisant un système d'équations multiples MMG. Ainsi, cette approche permet de vérifier que les conditions de moments sont valides lorsque le système est sur-identifié. Les S conditions de moments sont données par

$$E(Z_i \otimes f(w_i, \theta)) = 0. \qquad (2.3.1)$$

Où $Z = \left[Z^1_{s_1 \times n} \ldots Z^{J(n)}_{s_{J(n)} \times n}\right]$ et $S = \sum_{j=1}^{J(n)} s_j$. Soit $\bar{A}_n$ la matrice de poids de dimension S x S. L'équation (2.3.1) peut-être écrite comme

$$E\begin{pmatrix} Z^1_i f(w_i, \theta) \\ \vdots \\ Z^{J(n)}_i f(w_i, \theta) \end{pmatrix} = E\big(g(w_i, \theta)\big) = 0. \qquad (2.3.1^b)$$

Où pour tout j, $g^j(w_i, \theta) = Z^j_i f(w_i, \theta)$. Chaque estimateur MMG est obtenu en résolvant

$$\ddot{\theta}(\ddot{A}_n) = argmin_{\theta}\left(n^{-1}\sum_{i=1}^{n} Z_i \otimes f(w_i,\theta)\right)^T \ddot{A}'_n \ddot{A}_n \left(n^{-1}\sum_{i=1}^{n} Z_i \otimes f(w_i,\theta)\right) \quad (2.3.2)$$

Où $\ddot{\theta} = (\ddot{\theta}^1, \dots, \ddot{\theta}^{J(n)})'$. Par analogie avec la partie (3.1), la matrice de poids est tel que

$$\ddot{A}'_n \ddot{A}_n = \begin{bmatrix} A'^1_n A^1_n & \dots & A'^1_n A^{J(n)}_n \\ \vdots & \ddots & \vdots \\ A'^{J(n)}_n A^1_n & \dots & A'^{J(n)}_n A^{J(n)}_n \end{bmatrix}. \quad (2.3.3)$$

En supposant que $\ddot{A}'_n \ddot{A}_n \to_p \ddot{A}'\ddot{A}$ est définie positive, symétrique et le rang $\left(E\left(Z^j_i X'_i\right)\right)$ est égal a k pour tout j. Comme l'indique l'équation (2.1.3), l'estimateur alternatif est simplement une combinaison linéaire des estimés candidats trouvés grâce a l'équation (2.3.1), l'estimateur proposé est donné par

$$\ddot{\theta}^W = \sum_{j=1}^{J(n)} \Lambda_{nj} \ddot{\theta}^j. \quad (2.3.4)$$

Cette méthode peut-être vue comme une généralisation de l'approche de Chen, Jacho-Chávez et Linton (2009), car la procédure par système d'équations simples est un cas particuliers de l'estimateur obtenu par système d'équations multiples quand $A'^r_n A^s_n = \mathbf{0}$ avec $r \neq s$.

De plus, avec cette méthode, les estimateurs candidats j sont reliés entre eux. Enfin, noter que $\ddot{\theta}^W$ sera efficace si $\ddot{A}_n$ est choisit de tel sorte que la matrice de poids est donné par $\ddot{A}'_n \ddot{A}_n = (\psi_n)^{-1}$ où $n^{-1/2}\sum_{i=1}^{n} Z_i \otimes f(w_i,\theta) \sim N(0,\psi)$ par le TCL. Excepté pour le cas où le système est juste identifié où au moins une des équations est sur identifiée avec $\ddot{A}'_n \ddot{A}_n = (\psi_n)^{-1}$ (où ψ_n est bloc diagonale), l'estimation jointe est plus efficace. Cependant, l'approche par équation simple pourrait avoir de meilleures propriétés en petit échantillon. La spécification du modèle est supposée bonne; de plus, noter que les chances de mauvaises spécifications augmentent avec le nombre d'équations.

2.4 Propriétés en Grand Échantillon

Cette section présente les hypothèses nécessaires à l'estimateur proposé pour être efficace et asymptotiquement normal. Cette partie reprend certaines hypothèses suggérer par Chen, Jacho-Chávez et Linton (2009). La fonction de critère stochastique est donnée par

$$Q_n(\theta) = \left(n^{-1} \sum_{i=1}^{n} Z_i \otimes f(w_i, \theta) \right)^T \ddot{A}'_n \ddot{A}_n \left(n^{-1} \sum_{i=1}^{n} Z_i \otimes f(w_i, \theta) \right). \qquad (2.4)$$

Où la fonction $Q_n(\theta)$ est supposée continue et différentiable.

2.4.1 Efficacité

Comme dans Chen, Jacho-Chávez et Linton (2009), le fait que $\ddot{\theta}^W - \theta_0$ soit égal à $o_p\left(n^{-1/4}\right)$ est requis afin d'obtenir la $\sqrt{n}$ normalité asymptotique de $\ddot{\theta}^W$. Pour cela, les hypothèses suivantes sont requises:

- (A1) la matrice de poids satisfait $\sum_{j=1}^{J(n)} \Lambda_{nj} = I_k$ et
  $\sup_n \sum_{j=1}^{J(n)} \| \Lambda_{nj} \| < \infty$.
- (A2) $Q_n(\theta)$ a un unique minimum qui est $\theta_0 \in \Theta \subseteq \mathbf{R}^k$.
- (A3) $\ddot{\theta}_n \in \Theta$ et $Q_n\left(\ddot{\theta}_n\right) \leq \inf_{\theta \in \Theta} Q_n(\theta) + o_p\left(n^{-1/4}\right)$.
- (A4) Il existe une fonction non stochastique $Q(\theta)$ sur Θ , tel que
  $\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| = o_p\left(n^{-1/4}\right)$.
- (A5) Pour tout $\varepsilon > 0, \inf_{\theta \in B(\theta_0, \varepsilon)} Q(\theta) > Q(\theta_0)$.

Comme dans A1, la matrice de poids doit être bornée uniformément. Les hypothèses A3 et A4 sont respectivement des versions plus faibles des conditions sur les Estimateurs Extremum et sur la Convergence Uniforme Faible, où le taux de convergence est plus lent que d'habitude. L'hypothèse A5 est un critère d'identification.

Théorème 1 : En supposant que les hypothèses A1-A5 sont valides,

alors $\tilde{\theta}^{W} - \theta_0 = \left(n^{-1/4}\right)$.

2.4.2 Normalité Asymptotique

Cette section présente les conditions nécessaires pour que l'estimateur alternatif soit asymptotiquement normal. Les démonstrations reposent sur les hypothèses suggérées par Pakes et Pollard (1989), Newey et McFaddden (1994) et Chen, Jacho-Chávez et Linton (2009). Pour obtenir la normalité asymptotique, les hypothèses suivantes sont requises:

- (B1) θ_0 appartient à l'intérieur de Θ .
- (B2) $Q_n(\theta)$ est deux fois continument différentiable pour un certain voisinage $\Theta_0 \subset \Theta$ de θ_0 avec probabilité 1.
- (B3) $\sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) \to_d N(0, \Omega_0)$, où $\Omega_0 = \Gamma'_0 \ddot{A}'\ddot{A}\psi_0\ddot{A}'\ddot{A}\Gamma_0$,

  $\psi_0 = E\left(g(w_i, \theta)g(w_i, \theta)^T\right)$, $\Gamma_0 = E\left(\frac{\partial}{\partial\theta}g(w_i, \theta)\right)$ et $\ddot{A}'_n\ddot{A}_n \to_p \ddot{A}'\ddot{A}$.
- (B4) $\sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta) - B(\theta)\right\| \to_p 0$ pour certain J(n)S x J(n)S, la matrice de valeurs $B(\theta)$ est continue en θ_0 avec $B_0 = B(\theta_0)$ est non singulière. $B(\theta)$ est tel que

  $$[B(\theta)]_{rs} = E\left(\frac{\partial}{\partial\theta_r}g(w_i, \theta)^T\right)\ddot{A}'\ddot{A}E\left(\frac{\partial}{\partial\theta_s}g(w_i, \theta)\right) + E\left(\frac{\partial^2}{\partial\theta_r\partial\theta_s}g(w_i, \theta)^T\right)\ddot{A}'\ddot{A}E\left(g(w_i, \theta)\right)$$

  Pour r, s=1,..., J(n)S x J(n)S et $B_0 = \Gamma'_0\ddot{A}'\ddot{A}\Gamma_0$ est symétrique.
- (B5) $\max_{1\le j\le J(n)}\left\|\tilde{\theta}^j - \theta_0\right\| = o_p\left(n^{-1/4}\right)$.
- (B6) $\frac{\partial}{\partial\theta}Q_n(\tilde{\theta}) = o_p\left(n^{-1/2}\right)$.

- (B7) Il existe une séquence de matrices Λ^0_{nj} tel que

$$\sum_{j=1}^{J(n)} \left\| \left(\Lambda_{nj} - \Lambda^0_{nj} \right) \Gamma^{j^{-1}} \right\| = o_p(1)$$ et $\limsup_n \left\| \Lambda^0_{nj} \Gamma^{j^{-1}} \right\| < \infty$

où $\Gamma^j = E\left(\frac{\partial}{\partial \theta} g^j(w_i, \theta) \right)$.

- (B8[a]) $E(g^j(w_i, \theta) g^l(w_i, \theta)^T) = \psi^j(w_i)$ pour $j = l$ et $E(g^j(w_i, \theta_0) g^l(w_i, \theta_0)^T) = 0_k$ pour $j \neq l$.
- (B8[b]) $A'^j A^l = 0_k$ pour $j \neq l$.
- (B8[c]) Pour tout $j, l = 1, \dots, J(n)$, la matrice $\Sigma_n = \sum_{j=1}^{J(n)} \sum_{j=1}^{J(n)} \Lambda^0_{nj} V_{jl} \Lambda^{0T}_{nl}$ est définie positive et $V_{jl} = \Gamma_0^{-1^j} E\left(g^j(w_i, \theta) g^l(w_i, \theta)^T\right) \Gamma_0^{-1^{l^T}}$.

B1-B4 sont les hypothèses habituelles sur la fonction de critère permettant d'avoir la normalité asymptotique. Sachant B3 et la fonction de critère donné par (2.4), le TCL est appliqué à $n^{-1/2} \sum_{i=1}^{n} g(w_i, \theta_0)$ puisque $E\left(g(w_i, \theta_0)\right) = 0$. De plus, par la LFGN, $n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \theta'} g(w_i, \theta_0)$ devient $E \frac{\partial}{\partial \theta'} g(w_i, \theta_0)$ et $A_n \to_p A$. Le TCL et la LFGN requiert les hypothèses suivantes

$$E\|g(w_i, \theta_0)\|^2 < \infty \text{ et } E\left\| \frac{\partial}{\partial \theta'} g(w_i, \theta_0) \right\| < \infty.$$

Sous B4, la LFGN uniforme sur Θ_0 est appliquée à $\{g(w_i, \theta_0)\}_{i=1}^n$, $\left\{ \frac{\partial}{\partial \theta'} g(w_i, \theta_0) \right\}_{i=1}^n$ et $\left\{ \frac{\partial^2}{\partial \theta_r \partial \theta_s} g(w_i, \theta_0) \right\}_{i=1}^n$ pour $r, s = 1, \dots, l$. où les dérivées premières et secondes de $g(\cdot)$ par rapport à θ sont supposées continues en Θ_0. Γ_0 est de rang plein et A est non singulière. La LFGN uniforme requiert les hypothèses suivantes

$$E \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial \theta'} g(w_i, \theta_0) \right\| < \infty \quad \text{et} \quad E \sup_{\theta \in \Theta_0} \left\| \frac{\partial^2}{\partial \theta_r \partial \theta_s} g(w_i, \theta_0) \right\| < \infty$$

B5-B6 sont des hypothèses modifiées sur les Estimateurs Extremum qui prennent en comptent le Théorème 1.La matrice de poids déterministe doit être uniformément bornée. L'hypothèse B8 permet à l'estimateur alternatif d'être équivalent à celui de Chen, Jacho-Chávez et Linton (2009) quand *J(n)* n'est pas fixé. Noter que la matrice de poids est bloc diagonales sous les hypothèses ($B8^a$)-($B8^b$).

Lemme 1: En supposant (i) $\hat{\hat{\theta}}_n \to_p \theta_0$, *(ii)* $\sup_{\theta \in B(\theta_0, \varepsilon)} |F_n(\theta) - F(\theta)| \to_p 0 \ \forall \varepsilon > 0,$ *(iii)* $F(\theta)$ *est continue en* θ_0. *Alors* $F_n(\hat{\theta}_n) \to_p F(\theta_0)$.

Théorème 2: En supposant que les hypothèses B1-B8 sont valides,

alors $\sqrt{n}\left(\ddot{\theta}_I^W - \theta_0\right) \to_d N(0, \Sigma)$.

2.5 Matrice de Poids Optimal

Chen, Jacho-Chávez et Linton (2009) suggèrent l'utilisation de l'estimateur de minimisation de la distance proposé par Rothenberg (1973) afin d'obtenir la matrice de poids optimale. En utilisant cette approche, un estimateur efficient $\ddot{\theta}^{MD}$ peut être trouvé en minimisant l'espérance de la fonction de distance au carré. Chen, Jacho-Chávez et Linton (2009) proposent deux cas pour cette matrice de poids, dans un cas où *J* est fixé et l'autre où *J*(n) augment quand $n \to \infty$. Ici, nous considérons seulement le premier cas. La fonction de critère est alors donnée par

$$K_n(\theta) = \left[\begin{pmatrix} \ddot{\theta}^1 \\ \vdots \\ \ddot{\theta}^{J(n)} \end{pmatrix} - \theta \otimes i_{J(n)}\right]^T V^{-1} \left[\begin{pmatrix} \ddot{\theta}^1 \\ \vdots \\ \ddot{\theta}^{J(n)} \end{pmatrix} - \theta \otimes i_{J(n)}\right]. \tag{2.5.1}$$

Avec $i_{J(n)}$ qui est vecteur de uns et V est la matrice variance covariance (asymptotique) de dimension $J(n).k$ x $J(n).k$ sachant $\left(\sqrt{n}(\tilde{\theta}^1 - \theta), \dots, \sqrt{n}(\tilde{\theta}^{J(n)} - \theta)\right)$, tel que, $V = (V_{jl})$ pour tout j, l=1,..., $J(n)$:

$$V_{jl} = \Gamma_0^{-1j} E\left(g^j(w_i,\theta) g^l(w_i,\theta)'\right) \Gamma_0^{-1l^T}.$$

De plus, $\tilde{\theta}^j\left(A_n^j\right)$ et $\tilde{\theta}^j\left(A_n^j\right)$ seront efficaces si A_n^j est choisit de tel sorte que $A'^j_n A_n^j = \widehat{V_{jj}}^{-1}$ et $\widehat{V_{jj}}^{-1} \to_p V_{jj}^{-1}$. La condition de premier ordre est

$$[I_k \otimes i_{J(n)}]^T V^{-1} \begin{pmatrix} \tilde{\theta}^1 \\ \vdots \\ \tilde{\theta}^{J(n)} \end{pmatrix} = [I_k \otimes i_{J(n)}]^T V^{-1} [\theta \otimes i_{J(n)}].$$

En supposant que l'estimateur MMG est efficace et que les covariances entre les différents estimateurs sont nulles, l'estimateur $\hat{\theta}^{\mathrm{MD}}$ est donné par

$$\tilde{\theta}^W = \sum_{j=1}^{J(n)} \left(\sum_{l=1}^{J(n)} V_{ll}^{-1} \right)^{-1} V_{jj}^{-1} \tilde{\theta}^j. \qquad (2.5.2)$$

Intuitivement, la matrice de poids sera fonction de la matrice variance covariance. La relation inverse entre ces deux matrices implique que les coefficients ayant les écart-types les plus élevés, auront un poids inférieurs.

CHAPITRE III

TESTS D'HYPOTHÈSES

3.1 Statistique de Wald

La statistique de Wald teste les hypothèses suivantes:

$$H_0\colon \bar{\theta}^W = \theta \quad vs \qquad H_1\colon \bar{\theta}^W = \theta + \frac{\pi}{\sqrt{n}}. \qquad (3.1.1)$$

Sous H_1, $\frac{\pi}{\sqrt{n}}$ donne la distance par rapport à l'hypothèse nulle avec $\pi \in \mathbf{R}^k$ où $\|\pi\| > 0$. La formule asymptotique de la statistique de Wald et sa distribution sous l'hypothèse nulle sont données par

$$Wstat_n = n(\bar{\theta}^W - \theta)^T (\Sigma)^{-1} (\bar{\theta}^W - \theta), \qquad (3.1.2)$$

$$Wstat_n \sim_{H_0} \chi^2_{k,1-\alpha}, \qquad (3.1.3)$$

Où α correspond à la taille du test.

3.2 Puissance sous l'alternative de type Pitman-local

L'hypothèse alternative H_1 en (3.1.1) est construite afin de considérer le cas de l'alternative locale. En effet, les propriétés de puissance du test peuvent être vérifiées lorsqu'il y a une petite déviation par rapport à l'hypothèse nulle H_0. Cette méthode donne une bonne idée sur la puissance du test en échantillon fini. Sous H_1, $\frac{\pi}{\sqrt{n}}$ donne la distance par rapport à l'hypothèse nulle avec $\pi \in \mathbf{R}^k$, $\|\pi\| > 0$ et le taux $\frac{1}{\sqrt{n}}$ est choisi de tel sorte que la statistique sera non dégénéré asymptotiquement quand n$\rightarrow \infty$. Sous H_1, la statistique de Wald n'est plus centrée autour de zéro et sa distribution asymptotique est donnée par

$$Wstat_n \sim_{H_1} \chi^2_{k,1-\alpha}(\pi'\Sigma^{-1}\pi). \qquad (3.2.1)$$

Ainsi,

$$P(Wstat_n > \chi^2_{k,1-\alpha}) \rightarrow \omega\left(\pi'^{\Sigma^{-1}}\pi\right). \qquad (3.2.3)$$

Où ω est une fonction croissante en π tel que

$$\alpha \leq \omega(\pi'^{\Sigma^{-1}}\pi) \leq 1. \qquad (3.2.3)$$

Lorsque le paramètre de non centralité $\pi = 0$, la probabilité de rejeter l'hypothèse nulle devrait être égal à la taille asymptotique α du test. Quand l'hypothèse nulle est fausse, le test basé sur la statistique the $Wstat_n$ rejettera l'hypothèse nulle avec une probabilité asymptotique supérieure à α : le test a une puissance non trivial par rapport à l'alternative locale. La puissance du test sera inversement reliée à la variance de l'estimateur alternatif et positivement relié à la distance par rapport à l'hypothèse nulle.

3.3 Statistique J (Test de Sur identification)

La statistique J permet de vérifier la validité des instruments. Ce test a pour but de vérifier si les conditions de moments données par $(2.3.1^b)$ sont correctement spécifiés[2]. Ainsi, la statistique J teste les hypothèses suivantes:

$$H_0: E\begin{pmatrix} Z_i^1 f(w_i,\theta) \\ \vdots \\ Z_i^{J(n)} f(w_i,\theta) \end{pmatrix} = 0 \quad vs \quad H_1: E\begin{pmatrix} Z_i^1 f(w_i,\theta) \\ \vdots \\ Z_i^{J(n)} f(w_i,\theta) \end{pmatrix} \neq 0. \qquad (3.3.1)$$

Sachant la matrice de poids optimal, la formule de la statistique J asymptotique sera donnée par

$$Jstat_n = n\left(\sum_{i=1}^{n} Z_i \otimes f(w_i,\theta)\right)^T (\psi_n)^{-1} \left(\sum_{i=1}^{n} Z_i \otimes f(w_i,\theta)\right). \qquad \mathbf{(3.3.2)}$$

Où $Z = \left[Z^1_{s_1\times n} \dots Z^{J(n)}_{s_{J(n)}\times n}\right]$, $S = \sum_{j=1}^{J(n)} s_j$ et $\psi^j = E(g^j(w_i,\theta)g^j(w_i,\theta)')$.

Théorème 3: En supposant que l'hypothèses $B8^a$ est valide, alors la statistique J est donnée par

$$Jstat_n = \sum_{j=1}^{J(n)} Jstat_n^j. \qquad (3.3.3)$$

Où pour chaque $j=1,\dots, J(n)$, $Jstat_n^j$ est la statistique de sur identification de $\tilde{\theta}^j$, tel que

$$Jstat_n^j \sim \chi^2_{s_j-k}. \qquad (3.3.4)$$

Théorème 4: En supposant que pour $j=1,\dots, J(n)$, les statistiques $Jstat_n^j$ sont indépendantes, alors

$$Jstat_n \sim \chi^2_{\sum_{j=1}^{J(n)}(s_j-k)} = \chi^2_{S-J(n)k}. \qquad (3.4.4)$$

---

[2] Notez que la matrice de poids n'a aucun impact sur les conditions de moments.

CHAPITRE IV

APPLICATION EMPIRIQUE

4.1 Modèle Théorique

Cette section présente une application de la théorie présentée précédemment. En supposant que

$Y_i = X'_i\theta + U_i.$

où $E(U_i|X_i) = 0$, $E(U_i^2|X_i) = \sigma^2$ et $\theta \in \mathbf{R}^k$. Considérons les ensembles d'instruments Z_i^j tel que $j=\{A;B;C\}$, de dimension $l_j \times n$ and $l_j \geq k$. Les conditions de moments sont données par

$$E\begin{pmatrix} Z_i^A U_i \\ Z_i^B U_i \\ Z_i^C U_i \end{pmatrix} = 0. \qquad (4.1.1)$$

Avec le $rang(E(Z_i^j X'_i)) = k$. L'estimateur TSLS est donné par

$$\tilde{\theta} = argmin_{\tilde{\theta}}\left(n^{-1}\sum_{i=1}^{n} Z_i \otimes U_i\right)^T \left(n^{-1}\sum_{i=1}^{n} Z_i \otimes U_i\right).$$

Où $Z = \left[Z^A_{l_A \times n}\ Z^B_{l_B \times n}\ Z^C_{l_C \times n}\right]$. L'estimateur alternatif est donc donné par

$$\tilde{\theta}^W = \Lambda^{-1}\left(V_A^{-1}\left(X'P_AX\right)^{-1}X'P_A + V_B^{-1}\left(X'P_BX\right)^{-1}X'P_B + V_C^{-1}\left(X'P_CX\right)^{-1}X'P_C\right)Y. \tag{4.1.2}$$

Avec $\Lambda = V_A^{-1} + V_B^{-1} + V_C^{-1}$, $V_j = \sigma^2\left(X'P_jX\right)^{-1}$ et $P_j = Z^j[(Z'^jZ^j)]^{-1}Z^j$. Les résidus sont supposés homoscédastiques et les instruments sont mutuellement orthogonaux, il peut-être montré que la matrice variance covariance de l'estimateur alternatif est donnée par

$$Var(\tilde{\theta}^W) = \sigma^2\Lambda^{-1}\left(V_A^{-1}(X'P_AX)^{-1}V_A^{-1} + V_B^{-1}(X'P_BX)^{-1}V_B^{-1} + V_C^{-1}(X'P_CX)^{-1}V_C^{-1}\right)\Lambda'^{-1}. \tag{4.1.3}$$

4.2 Application

Une application empirique peut être fournie par l'utilisation de l'étude d'Angrist et Krueger (1991) sur les rendements de la scolarité utilisant le trimestre de naissance comme un instrument. En raison des lois de scolarité obligatoires, les personnes ayant des dates de naissance plus tôt dans l'année scolaire ont un peu moins de scolarité que les personnes nées vers la fin de l'année. En effet, la politique sur l'âge de commencement de l'école et les lois de scolarité obligatoires crées une expérience naturelle dans laquelle les individus sont obligés de fréquenter l'école pour différentes longueurs de temps en fonction de leur anniversaire.

Angrist et Krueger (1991) ont établi un échantillon du recensement de 1980 aux États-Unis qui se composait de 329.500 hommes nés entre 1930. L'approche décrite dans le présent document est la même que dans Angrist et Krueger (1991 Tableau VII). Selon Hansen, Hausman et Newey (2006) l'approche d'Angrist et Krueger (1991) est plus susceptible de souffrir d'un problème de nombreux instruments plutôt qu'un problème de faibles instruments tel que propose Bound, Jaeger et Baker (1996) et Staiger et Stock (1997). Le modèle considéré a une constante, neuf années de naissance (Y), huit dichotomiques sur les régions de classification (R), une dichotomique sur la race (Race) et une dichotomique sur la situation maritale (Marié). Enfin l'âge et l'âge au carré sont mesurés en trimestre (respectivement AGEQ et AGEQSQ).

Pour construire l'estimateur alternatif $\tilde{\theta}$, une gamme de 3 estimateurs ($\theta^A, \theta^B, \theta^C$) est considérée. Chaque estimateur de cette gamme est obtenu en utilisant respectivement Z^A, Z^B and Z^C en tant qu'instruments. Le bloc Z^A comprend les trimestres de naissance (dénotée par Q). Un total de 27 variables dichotomiques compose le bloc Z^B et celui-ci est obtenu en faisant interagir les trimestres de naissance avec les 9 dichotomiques d'années de naissance (dénotée par Q * Y). Le dernier bloc Z^C est formé en faisant interagir les trimestres de naissance avec les 8 régions de la classification (dénotée par Q * R). Les estimations sont obtenues en utilisant la procédure décrite à la section 5.2. Cette méthode a été construite de telle sorte que si au moins une équation est sur-identifiée (mais que les équations ne sont pas liés au sens de B8^a-B8^b) alors la procédure suggérée est asymptotiquement équivalente à l'approche de Chen, Jacho-Chávez et Linton (2009)[3].

Le tableau 1 présente les estimations pour la méthode standard et alternative. Les estimations alternatives sont similaires au TSLS standard en termes de grandeur des coefficients et des écarts-type. De plus, les écarts-types de l'approche alternative sont plus petits. Notez que la matrice de variance covariance de la méthode alternative a été calculée sous la supposition d'homoscédasticité. La statistique de Hausman confirme que l'estimateur alternatif est équivalent à celui traditionnel en ne rejetant pas l'hypothèse nulle. Afin de calculer le test de sur-identification proposé par (3.4.4), le test du Chi-Carré de

[3] Les estimations équations-par-équations sont présentées dans les tableaux 1 et 2.

Basmann est utilisé pour vérifier que les conditions de moments sont correctement spécifiées. Dans les deux cas, l'approche standard et alternative, l'hypothèse nulle n'est pas rejetée, donc les instruments proposés sont valables.

Tableau 4.1.

Estimations du Rendement de l'Education pour les hommes nés entre 1930 et 1949: Recensement de 1980. (TSLS Alternatif vs. TSLS Standard).

Variables Indépendantes	TSLS Alternatif	TSLS Standard
Nombre d´années d´éducation	0.0744***	0.0736***
	(0.0197)	(0.0204)
Race (1=noir)	-0.2398***	-0.2412***
	(0.0311)	(0.0324)
Marié (1=marié)	0.2453***	0.2455***
	(0.0049)	(0.0056)
SMSA (1=center city)	-0.1645***	-0.1654***
	(0.0207)	(0.0216)
Âge	-0.0823***	-0.0818
	(0.005)	(0.0616)
Âge au carré	0.0009***	0.0009
	(0.0001)	(0.0007)
Constante	6.7454***	6.7498***
	(0.348)	(1.3592)
Statistique de Basmann [df]	43.5271 [52]	41.6251 [52]

Statistique d´Hausman [df]	6.9054 [23]

Notes: Les écarts-types sont entre parenthèses. La taille de l'échantillon est de 329,509 individus. La variable dépendante est le logarithme des gains hebdomadaires. L´Âge et l´âge au carré sont mesurées en termes de trimestre. Un total de 9 variables dichotomiques sur l´année de naissance et de 8 variables dichotomiques sur la Région-de-résidence sont incluses mais non présentées dans le tableau.*, **, *** Indique le niveau de signification de 90, 95, et 99 pour cent, respectivement.

Tableau 4.2.

Estimations du Rendement de l'Education pour les hommes nés entre 1930 et 1949: Recensement de 1980. (Estimés des Modèles A, B et C).

Variables Indépendantes	Modèle A	Modèle B	Modèle C
Nombre d'années d'éducation	0.6774	0.0600**	0.0886***
	(1.5355)	(0.0290)	(0.0289)
Race (1=noir)	-0.5988	0.0925*	0.0604
	(1.7231)	(0.0480)	(0.0481)
Marié (1=marié)	-0.5305	0.0878**	0.059
	(1.5410)	(0.0431)	(0.0432)
SMSA (1=center city)	-0.4204	0.0809**	0.0576
	(1.2503)	(0.0373)	(0.0374)
Âge	-0.3251	0.0771**	0.0584*
	(1.0039)	(0.0325)	(0.0326)
Âge au carré	-0.2232	0.0687**	0.0552**
	(0.7303)	(0.0274)	(0.0276)
Constante	-0.1683	0.0523**	0.0420*
	(0.5528)	(0.0232)	(0.0233)
	0.6774	0.0600**	0.0886***
Statistique de Basmann [df]	(1.5355)	(0.0290)	(0.0289)
Instruments	Z^A	Z^B	Z^C

Notes: Les écarts-types sont entre parenthèses. La taille de l'échantillon est de 329,509 individus. La variable dépendante est le logarithme des gains hebdomadaires. L'Âge et l'âge au carré sont mesurées en termes de trimestre. Un total de 9 variables dichotomiques sur l'année de naissance et de 8 variables dichotomiques sur la Région-de-résidence sont incluses mais non présentées dans le tableau.*, **, *** Indique le niveau de signification de 90, 95, et 99 pour cent, respectivement.

CHAPITRE V

EXPÉRIENCES DE MONTE CARLO

Les expériences de Monte Carlo sont basées sur le modèle de Carrasco (2009). Dans cette partie, un modèle avec des conditions de moment indépendantes et identiquement distribuées est estimé. Le modèle a les spécifications suivantes:

$$y_i = \delta W_i + \varepsilon_i \qquad (5.1)$$

$$W_i = \gamma e^{-x_i^2} + u_i$$

Où δ=0.1, γ= {0.1, 1} et les résidus $(\varepsilon_i, u_i)'$ ont la distribution suivante

$$\begin{bmatrix} \varepsilon_i \\ u_i \end{bmatrix} \sim iidN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \varphi \\ \varphi & 1 \end{bmatrix}\right)$$

Avec $\varphi \in \{0.2, 0.5, 0.8\}$ correspondant respectivement au cas où l'endogénéité est faible, moyenne et forte. Noter que ces spécifications prennent en considération le cas d'instruments faibles où γ=0.1. Afin de faire face aux problèmes d'endogénéité, les instruments sont construits en utilisant les fonctions de puissance où $z_1 = (x_i\ x_i^2\ x_i^3)$ et pour $2 \leq k \leq 5$, $z_k = (x_i^{2k}\ x_i^{2k+1})$.

5.1 Monte Carlo: Estimation

Les résultats pour l'approche alternative et standard sont rapportés lorsque $\gamma = \{0.1, 1\}$. Les deux méthodes prennent en considération deux échantillons de tailles n = 500, 1000 et 1000 réplications. Chaque estimateur $\tilde{\tilde{\delta}}^j$ et $\tilde{\tilde{\delta}}$ est calculé en utilisant l'approche MMG à deux étapes en utilisant la matrice au poids optimale. Compte tenu de chaque valeur de paramètre d'endogénéité (φ) et du nombre d'estimateurs pondérés, le biais moyen simulé, l'écart-type moyen et l'EQM pour chaque estimateur sont rapportés. Les tableaux 5.3 et 5.4 rapportent le résumé pour les estimateurs standard et alternatif lorsque $\gamma = 1$. L'ampleur des biais augmente avec le nombre d'instruments et diminue lorsque n devient plus grand pour les deux méthodes. De plus, le biais de l'estimateur alternatif est plus petit que le standard. Les variances et l'EQM se comportent très différemment en fonction de la méthode, la taille de l'échantillon et le paramètre d'endogénéité. Cependant, la méthode alternative semble avoir de plus petits biais et EQM, en particulier dans le cas d'endogénéité faibles et moyennes. L'estimateur alternatif démontre de plus petits EQM que l'estimateur standard quand la gamme des estimateurs augmente.

Les tableaux 5.5 et 5.6 présentent les résultats pour les estimateurs standard et alternatif lorsque γ =0,1, le cas où les instruments sont faibles. Lorsque les instruments sont faibles, le biais et l'EQM de l'approche alternative sont beaucoup plus grands que l'approche standard. Le biais augmente encore lorsque la gamme des estimations s'agrandit, mais l'ampleur de cette augmentation est beaucoup plus élevée. Contrairement au cas où $\gamma = 1$, l'EQM démontre que l'estimateur alternatif a des performances bien pire quand la gamme des estimateurs devient plus grande.

Tableau 5.3.

Sommaire des Statistiques de $\tilde{\delta}$ (Approche Alternative) $\gamma=1$

J	Estimateurs Pondérés	Biais	Std.	EQM	Biais	Std.	EQM
		$\varphi=0.2,\ n=500$			$\varphi=0.2,\ n=1000$		
2	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2)$	-0.00135	0.07981	0.00637	-0.00139	0.09462	0.00896
3	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3)$	-0.01610	0.14889	0.02243	-0.00176	0.09528	0.00908
4	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3),\tilde{\delta}^4(z_4)$	-0.01297	0.09817	0.09834	-0.00626	0.08962	0.08966
5	$\tilde{\delta}^1(z_1)$, $\tilde{\delta}^2(z_2)$, $\tilde{\delta}^3(z_3)$, $\tilde{\delta}^4(z_4)$, $\tilde{\delta}^5(z_5)$	0.01934	0.10231	0.10269	-0.02251	0.06367	0.06417
		$\varphi=0.5,\ n=500$			$\varphi=0.5,\ n=1000$		
2	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2)$	-0.01434	0.11541	0.11561	0.00271	0.10330	0.10331
3	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3)$	-0.01608	0.15567	0.15593	-0.00338	0.08262	0.08263
4	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3),\tilde{\delta}^4(z_4)$	-0.05833	0.05972	0.06313	-0.00849	0.06970	0.06977
5	$\tilde{\delta}^1(z_1)$, $\tilde{\delta}^2(z_2)$, $\tilde{\delta}^3(z_3)$, $\tilde{\delta}^4(z_4)$, $\tilde{\delta}^5(z_5)$	-0.11870	0.03611	0.05020	-0.06559	0.06102	0.06532
		$\varphi=0.8,\ n=500$			$\varphi=0.8,\ n=1000$		
2	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2)$	-0.01719	0.07981	0.08011	-0.01531	0.14100	0.14123
3	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3)$	-0.03479	0.10770	0.10892	-0.02142	0.06066	0.06111
4	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3),\tilde{\delta}^4(z_4)$	-0.02977	0.10867	0.10956	-0.02258	0.05355	0.05406
5	$\tilde{\delta}^1(z_1)$, $\tilde{\delta}^2(z_2)$, $\tilde{\delta}^3(z_3)$, $\tilde{\delta}^4(z_4)$, $\tilde{\delta}^5(z_5)$	-0.06310	0.09188	0.09586	-0.03269	0.07109	0.07216

Notes: Nombre de réplications=1000. Une constant a été utilisée mais pas rapportée dans le tableau.

Tableau 5.4.

Sommaire des Statistiques de $\tilde{\delta}$ (Approche Standard) : γ=1

J	Estimateurs	Biais	Std.	EQM	Biais	Std.	EQM
		φ=0.2, n=500			φ=0.2, n=1000		
2	$\tilde{\delta}(z_1,z_2)$	0.0183	0.1244	0.0158	0.0320	0.1053	0.0121
3	$\tilde{\delta}(z_1,z_2,z_3)$	0.0205	0.1177	0.0143	0.0293	0.1024	0.0113
4	$\tilde{\delta}(z_1,z_2,z_3,z_4)$	0.0544	0.1167	0.1196	-0.0147	0.0989	0.0991
5	$\tilde{\delta}(z_1,z_2,z_3,z_4,z_5)$	0.0483	0.1162	0.1185	-0.0474	0.0983	0.1006
		φ=0.5, n=500			φ=0.5, n=1000		
2	$\tilde{\delta}(z_1,z_2)$	-0.0071	0.1276	0.1277	0.0342	0.1031	0.1043
3	$\tilde{\delta}(z_1,z_2,z_3)$	-0.0029	0.1127	0.1127	0.0207	0.1027	0.1031
4	$\tilde{\delta}(z_1,z_2,z_3,z_4)$	-0.0155	0.1005	0.1008	0.0026	0.1208	0.1208
5	$\tilde{\delta}(z_1,z_2,z_3,z_4,z_5)$	-0.0271	0.0978	0.0986	-0.0082	0.1278	0.1278
		φ=0.8, n=500			φ=0.8, n=1000		
2	$\tilde{\delta}(z_1,z_2)$	0.0220	0.0881	0.0886	0.0320	0.1053	0.1064
3	$\tilde{\delta}(z_1,z_2,z_3)$	0.0029	0.0876	0.0876	0.0293	0.1024	0.1033

4	$\hat{\delta}(z_1, z_2, z_3, z_4)$	-0.0121	0.0849	0.0850	-0.0147	0.0989	0.0991
5	$\hat{\delta}(z_1, z_2, z_3, z_4, z_5)$	-0.0378	0.0756	0.0770	-0.0474	0.0983	0.1006

Notes: Nombre de réplications=1000. Une constant a été utilisée mais pas rapportée dans le tableau.

Tableau 5.5.

Sommaire des Statistiques de $\tilde{\delta}$ (Approche Alternative): γ=0.1

J	Estimateurs Pondérés	Biais	Std.	EQM	Biais	Std.	EQM
		φ=0.2, n=500			φ=0.2, n=1000		
2	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2)$	-0.1718	0.8522	0.7558	-0.1430	0.8661	0.7707
3	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3)$	-0.1649	0.1331	0.0449	-0.1100	0.7749	0.6126
4	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3),\tilde{\delta}^4(z_4)$	-0.1724	0.0981	0.0983	-0.5833	1.3857	2.2606
5	$\tilde{\delta}^1(z_1)$, $\tilde{\delta}^2(z_2)$, $\tilde{\delta}^3(z_3)$, $\tilde{\delta}^4(z_4)$, $\tilde{\delta}^5(z_5)$	-0.2064	0.2453	0.1028	-0.1829	0.2432	0.0926
		φ=0.5, n=500			φ=0.5, n=1000		
2	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2)$	-0.4164	0.4405	0.3675	-0.3955	1.1400	1.4562
3	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3)$	-0.4010	0.4562	0.3690	-0.4256	0.2374	0.2375
4	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3),\tilde{\delta}^4(z_4)$	-0.5726	0.1729	0.3577	-0.4365	0.3191	0.2924
5	$\tilde{\delta}^1(z_1)$, $\tilde{\delta}^2(z_2)$, $\tilde{\delta}^3(z_3)$, $\tilde{\delta}^4(z_4)$, $\tilde{\delta}^5(z_5)$	-0.4225	0.0887	0.1863	-0.4891	0.1898	0.2752
		φ=0.8, n=500			φ=0.8, n=1000		
2	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2)$	-0.7278	0.1740	0.5600	-0.6669	0.5193	0.7145
3	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3)$	-0.7239	0.2500	0.5866	-0.6378	0.2974	0.4952
4	$\tilde{\delta}^1(z_1),\tilde{\delta}^2(z_2),\tilde{\delta}^3(z_3),\tilde{\delta}^4(z_4)$	-0.7670	0.3882	0.7391	-0.7223	0.1031	0.5324
5	$\tilde{\delta}^1(z_1)$, $\tilde{\delta}^2(z_2)$, $\tilde{\delta}^3(z_3)$, $\tilde{\delta}^4(z_4)$, $\tilde{\delta}^5(z_5)$	-0.7278	0.1740	0.5600	-0.6669	0.5193	0.7145

Notes: Nombre de réplications=1000. Une constant a été utilisée mais pas rapportée dans le tableau.

Tableau 5.6.

Sommaire des Statistiques de $\tilde{\delta}$ (Approche Standard) : γ=0.1

J	Estimateurs	Biais	Std.	EQM	Biais	Std.	EQM
			φ=0.2, n=500			*φ=0.2, n=1000*	
2	$\tilde{\delta}(z_1,z_2)$	0.0305	0.7860	0.6187	-0.3469	0.6555	0.5501
3	$\tilde{\delta}(z_1,z_2,z_3)$	-0.2316	0.1291	0.0703	-0.1759	0.6758	0.4876
4	$\tilde{\delta}(z_1,z_2,z_3,z_4)$	-0.0206	0.2850	0.0817	0.1533	0.3691	0.1598
5	$\tilde{\delta}(z_1,z_2,z_3,z_4,z_5)$	-0.0380	0.1302	0.0184	0.0313	0.1276	0.0172
			φ=0.5, n=500			*φ=0.5, n=1000*	
2	$\tilde{\delta}(z_1,z_2)$	-0.1377	0.6019	0.3813	-0.3592	1.5167	2.4293
3	$\tilde{\delta}(z_1,z_2,z_3)$	-0.6173	0.5635	0.6986	-0.3676	1.5212	2.4493
4	$\tilde{\delta}(z_1,z_2,z_3,z_4)$	-0.0222	0.1930	0.0378	-0.1376	0.1805	0.0515
5	$\tilde{\delta}(z_1,z_2,z_3,z_4,z_5)$	-0.1397	0.0964	0.0288	-0.2267	0.1084	0.0631
			φ=0.8, n=500			*φ=0.8, n=1000*	
2	$\tilde{\delta}(z_1,z_2)$	-0.4068	0.1720	0.1951	-0.1207	0.1794	0.0467
3	$\tilde{\delta}(z_1,z_2,z_3)$	-0.0197	0.1520	0.0235	-0.5863	0.2930	0.4296
4	$\tilde{\delta}(z_1,z_2,z_3,z_4)$	-0.6673	0.4965	0.6918	-0.4373	0.1470	0.2129
5	$\tilde{\delta}(z_1,z_2,z_3,z_4,z_5)$	-0.2814	0.0891	0.0871	-0.4560	0.2083	0.2513

Notes: Nombre de réplications=1000. Une constant a été utilisée mais pas rapportée dans le tableau.

5.2 Monte Carlo: Tests d'Hypothèses

Le tableau 5.7 rapporte les probabilités du rejet de l'hypothèse nulle et alternative présentées dans (3.1.1) lorsque $\gamma = 1$[4] avec un niveau de signification nominal α de 5%. La spécification dans (5.1) est toujours considérée, mais sans une constante dans la première équation. Seuls les cas où *J = 2,3* sont considérés. Quatre tailles d'échantillon sont utilisées: n = 200, 500, 1000, 1000 et 10000 répétitions. La distance de l'hypothèse nulle est définie en fonction du biais de telle sorte que $\pi_\mu = \mu \times bias$ où $\mu = \{1,1.5,2\}$.

Sans surprise, la puissance des tests augmente en même temps que la distance de l'hypothèse nulle s'agrandit, elle semble converger vers 1 quand $n \rightarrow \infty$. Cependant, les tests ont une bonne taille seulement lorsque *n* est assez grand et dans le cas d'endogénéité faible. Dans le cas d'endogénéité moyenne, les tests ont une taille adéquate que lorsque n = 10000. La statistique peut souffrir d'un biais de petits échantillons. Si le degré de liberté était plus grand que 1, au lieu d'utiliser une distribution Chi-Carré, une statistique-F = $\frac{Wstat_n}{k}$ peut être utilisé. Dans ce cas, l'inférence basée sur le F (n, k) peut fournir des résultats plus robuste en échantillon fini. Ici, la statistique de Wald est équivalente à une statistique-F puisque le degré de liberté est juste le carré d'une statistique *t*.

Notez que la probabilité de rejet sous l'hypothèse alternative rétrécit à mesure que *n* augmente, car dans ce cas la distance de l'hypothèse nulle devient plus petite avec la variation

---

[4] Le cas de l'instrument faible a été considéré, mais non rapporté car la taille semble fausse. Il pourrait être plus approprié d'utiliser une approche d'Anderson et Rubin (1949).

de n. Il peut être démontré que lorsque le paramètre décentralisé π devient plus grand, la distribution devient plus étendue et se déplace vers la droite.

Tableau 5.7.

Probabilités de Rejet

	J=2					J=3			
	Hypothèse Nulle	Hypothèse Alternative				Hypothèse Nulle	Hypothèse Alternative		
N		π_1	$\pi_{1.5}$	π_2	N		π_1	$\pi_{1.5}$	π_2
		φ=0.2					*φ=0.2*		
200	0.111	0.679	0.777	0.826	200	0.235	0.756	0.831	0.871
500	0.061	0.637	0.767	0.828	500	0.155	0.699	0.804	0.854
1000	0.051	0.626	0.751	0.82	1000	0.106	0.665	0.781	0.846
10000	0.04	0.601	0.737	0.803	10000	0.04	0.646	0.756	0.81
		φ=0.5					*φ=0.5*		
200	0.238	0.771	0.842	0.876	200	0.387	0.815	0.873	0.908
500	0.157	0.695	0.795	0.851	500	0.276	0.756	0.83	0.882
1000	0.111	0.668	0.774	0.825	1000	0.199	0.736	0.813	0.851
10000	0.049	0.64	0.757	0.825	10000	0.064	0.663	0.757	0.826
		φ=0.8					*φ=0.8*		
200	0.521	0.864	0.908	0.933	200	0.668	0.915	0.948	0.958
500	0.379	0.777	0.854	0.888	500	0.554	0.867	0.554	0.923
1000	0.267	0.742	0.819	0.855	1000	0.44	0.819	0.882	0.907
10000	0.091	0.665	0.764	0.828	10000	0.161	0.682	0.788	0.835

Notes : Nombre de réplications=1000.

CONCLUSION ET EXTENSIONS

Avec l'approche considérée, les estimateurs candidats sont estimés par des équations multiples MMG. Cette approche permet de construire un test de sur identification, ainsi, la validité des instruments peut être testée conjointement.

Dans l´application d´Angrist et Krueger (1991), les estimateurs alternatif et standard ont des performances similaires. Cependant, l´estimateur proposé possède des écart-types légèrement plus petits que ceux de la méthode standard.

De plus en supposant que le modèle est correctement spécifié, les expériences de Monte Carlo montrent que l'estimateur proposé a de meilleures performances lorsque le nombre d´estimateurs candidats augmentent, dans le cas où les instruments ne sont pas faibles et le niveau d'endogénéité n'est pas fort ; cela peut s'expliquer par le fait qu´un plus large éventail d'estimateurs contient plus d'informations.

En considérant l'alternative de type Pitman-local, il se pourrait que la statistique de Wald souffre d'un biais en petit échantillon, mais cette dernière semble donner des résultats asymptotique satisfaisant.

Tester pour les conditions de moment et les instruments faibles sont des facteurs importants à prendre en compte lors de l'application de cette méthode. Ainsi, il pourrait être intéressant d'étudier les propriétés de puissance et de taille de la statistique AR et de la statistique J proposée dans le cadre l´alternative locale. De plus, la matrice de poids optimal est diagonale par blocs en raison de l'hypothèse d'orthogonalité entre les résidus de chaque ensemble d'estimateurs, ainsi le comportement de l'estimateur proposé pourrait changer sans cette hypothèse. Enfin, notez que la méthode proposée ne tient pas compte du compromis entre le gain potentiel et le prix pour combiner.

APPENDICE

PREUVES MATHÉMATIQUES

Preuve du Théorème 1:

Sachant $\varepsilon > 0$, il existe $\delta > 0$ tel que $\theta \notin B(\theta_0, \varepsilon)$, cela implique que

$Q(\theta) - Q(\theta_0) \geq \delta > 0$ par A4, ainsi

$$P(\tilde{\theta}_n \in B(\theta_0, \varepsilon)) \leq P(Q(\tilde{\theta}_n) - Q(\theta_0) \geq \delta)$$

$$= P \left(Q(\tilde{\theta}_n) - Q_n(\tilde{\theta}_n) + Q_n(\tilde{\theta}_n) - Q(\theta_0) \geq \delta\right)$$

$$= P \left(Q(\tilde{\theta}_n) - Q_n(\tilde{\theta}_n) + Q_n(\theta_0) + o_p\left(n^{-1/4}\right) - Q(\theta_0) \geq \delta\right)$$ par A5

$$\leq P \left(2 sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| + 0_p\left(n^{-1/4}\right) \geq \delta\right) \rightarrow_p 0$$ par A3

Il est donc prouvé que $\tilde{\theta} - \theta_0 = o_p\left(n^{-1/4}\right)$.

Par (2.3.4) $\tilde{\theta}^W = \sum_{j=1}^{J(n)} \Lambda_{nj} \tilde{\theta}^j$. Cela implique que

$$\|\tilde{\theta}^W - \theta_0\| \leq \sum_{j=1}^{J(n)} \|\Lambda_{nj}\| \|\tilde{\theta}^j - \theta_0\| = o_p\left(n^{-1/4}\right)$$, d'après l'inégalité triangulaire et A1.□

Preuve du Lemme 1:

$$\left|F_n\left(\hat{\theta}_n\right) - F(\theta_0)\right| \leq \left|F_n\left(\hat{\theta}_n\right) - F\left(\hat{\theta}_n\right) + F\left(\hat{\theta}_n\right) - F(\theta_0)\right|$$

$\leq |F_n(\hat{\theta}_n) - F(\hat{\theta}_n)| + |F(\hat{\theta}_n) - F(\theta_0)|$, d'après l'inégalité triangulaire.

$\leq sup_{\theta \in B(\theta_0, \delta)} \left|F_n\left(\hat{\theta}_n\right) - F\left(\hat{\theta}_n\right)\right| + |F(\hat{\theta}_n) - F(\theta_0)| \rightarrow_p 0$ □

Preuve du Théorème 2:

D'après les hypothèses B2 et B6, le théorème de la valeur moyenne $\frac{\partial}{\partial \theta} Q_n(\ddot{\theta}_n)$ et sachant que θ_n^* se situent entre $\ddot{\theta}_n$ et θ_0.

$$\frac{\partial}{\partial \theta} Q_n(\ddot{\theta}_n) = \frac{\partial}{\partial \theta} Q_n(\theta_0) + \frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_n^*)(\ddot{\theta}_n - \theta_0) = o_p\left(n^{-1/4}\right) \qquad (T2.1)$$

D'après le lemme 1, B2, B4 et B5

$$\frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta_n^*) = B_0 + o_p(1) \qquad (T2.2)$$

En substituant $(T2.2)$ dans $(T2.1)$

$$\sqrt{n}\left(\frac{\partial}{\partial \theta} Q_n(\theta_0) + (B_0 + o_p(1))(\ddot{\theta}_n - \theta_0)\right) = o_p\left(n^{-1/4}\right)$$

$$\sqrt{n}\left((\ddot{\theta}_n - \theta_0)\right) = \left((B_0 + o_p(1))\right)^{-1} \sqrt{n} \frac{\partial}{\partial \theta} Q_n(\theta_0) + o_p\left(n^{-1/4}\right)$$

$\sqrt{n}\left((\ddot{\theta}_n - \theta_0)\right) \rightarrow_d \mathbb{N}\left(0, B_0^{-1}\Omega_0 B_0^{-1}\right)$, par B3, B4 et le Théorème de Convergence de Cramer.

De plus, par (2.3.4)

$$\ddot{\theta}^{W} = \sum_{j=1}^{J(n)} \Lambda_{nj}\ddot{\theta}^{j}$$

$$\sqrt{n}\left(\ddot{\theta}^{W} - \theta_0\right) = \sqrt{n}\sum_{j=1}^{J(n)} \Lambda_{nj}\left(\ddot{\theta}^{j} - \theta_0\right)$$

$$= \sqrt{n}\sum_{j=1}^{J(n)} \Lambda^{0}_{nj}(\ddot{\theta}^{j} - \theta_0) + \sqrt{n}\sum_{j=1}^{J(n)} \left(\Lambda_{nj} - \Lambda^{0}_{nj}\right)(\ddot{\theta}^{j} - \theta_0)$$

$\sqrt{n}\left(\ddot{\theta}^{W} - \theta_0\right) \to_d N(0,\Sigma)$, d'après l'inégalité triangulaire, B5 et B7

$$\left\|\sqrt{n}\sum_{j=1}^{J(n)} \left(\Lambda_{nj} - \Lambda^{0}_{nj}\right)(\ddot{\theta}^{j} - \theta_0)\right\| \le \sqrt{n}\,max_{1\le j\le J(n)}\left\|\Gamma^{j}(\ddot{\theta}^{j} - \theta_0)\right\| \sum_{j=1}^{J(n)} \left\|\left(\Lambda_{nj} - \Lambda^{0}_{nj}\right)\Gamma^{j^{-1}}\right\|$$

$$= O_p(1) \times o_p(1) \qquad \square$$

Preuve du Théorème 3:

La statistique $Jstat_n$ est donnée par

$$Jstat_n = n\left(\sum_{i=1}^{n} Z' \bigotimes f(w_i,\theta)\right)^{T} (\psi_n)^{-1}\left(\sum_{i=1}^{n} Z' \bigotimes f(w_i,\theta)\right)$$

$$= n\begin{pmatrix} \sum_{i=1}^{n} Z_i^{1} f(w_i,\theta) \\ \vdots \\ \sum_{i=1}^{n} Z_i^{J(n)} f(w_i,\theta) \end{pmatrix}^{T} \begin{pmatrix} \psi^{1} & \dots & 0_{k\times k} \\ \vdots & \ddots & \vdots \\ 0_{k\times k} & \dots & \psi^{J(n)} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n} Z_i^{1} f(w_i,\theta) \\ \vdots \\ \sum_{i=1}^{n} Z_i^{J(n)} f(w_i,\theta) \end{pmatrix}$$

Avec B8[a] , ψ_n devient block diagonale, et $Jstat_n$ peut s'écrire comme

$$Jstat_n = n\left(\sum_{j=1}^{J(n)}\left(\sum_{i=1}^{n} Z_i^j f(w_i,\theta)\right)^T (\psi^j)^{-1}\left(\sum_{i=1}^{n} Z_i^j f(w_i,\theta)\right)\right) \qquad (T3.1)$$

□

Preuve du Théorème 4:

Avec $(T3.1)$ et l'hypothèse sur l'indépendance, le nombre de degré de liberté de $Jstat_n$ est donné par la somme des degrés de libertés de chaque $Jstat_n{}^j$, pour tout j=1,.., $J(n)$.

BIBLIOGRAPHIE

AMEMIYA, T. (1974): "The Non-Linear Two-Stage Least Squares Estimator," *Journal of Econometrics, 2*, 105-110.

ANDERSON, T., ET 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. (1988): "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348-358.

ANDREWS, D. W. K. (1988): "Chi-Squared Diagnostic Tests for Econometric Models: Theory," *Econometrica*, 56(6), 1419-1453.

ANDREWS, D. W. K. (1989): "Power in Econometric Applications," *Econometrica*, 57(5), 1059-1090.

ANGRIST, J. D., ET A. B. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings," *The Quarterly Journal of Economics*, 106(4), 979-1014.

ANGRIST, J. D., G. W. IMBENS, ET A. B. KRUEGER (1999): "Jackknife Instrumental Variables Estimation," *Journal of Econometrics*, 14, 54-67.

BOUND, J., D. JAEGER, ET R. BAKER (1996): "Problems with Instrumental Variables Estimation when the Correlation between Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90,443-450.

BREIMAN, L. (1996a): "Stacked Regressions," *Machine Learning,* 24,49-64.

BREIMAN, L. (1996b): "Bagging predictors," *Machine Learning,* 24, 123-140.

CAMPBELL, J. Y., A.W. LO ET A.C. MACKINLAY (1996): *The Econometrics of Financial Markets.* Princeton University Press.

CAMRON, A. C. ET P.K. TRIVEDI (2005): *Microeconometrics: Methods and Applications.* Cambridge University Press.

CARRASCO, M. (2009): "A Regularization Approach to the Many Instruments Problem," *Forthcoming in the Journal of Econometrics.*

CHEN, X. ET F. FAN (1999): "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91 (2), 373-401.

CHEN, X., D.T. JACHO-CHAVEZ ET O.LINTON (2009): "An Alternative Way of Computing Efficient Instrumental Variable Estimators," Research Paper, LSE STICERD.

CHAUSSE, P. (2009): "Computing GMM and GEL with R," *Forthcoming in the Journal of Statistical Software.*

DAVIDSON, R., ET J. G. MACKINNON (1981): "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49, 781-793.

DAVIDSON, R., ET J. G. MACKINNON (1987): "Implicit Alternatives and the LocalPower of Test Statistics," *Econometrica*, 55, 1305-1329.

DONALD, S. ET W. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161-1191.

HANSEN, C., J. HAUSMAN ET W. NEWEY (2006): "Estimation with Many Instrumental Variables," mimeo, Chicago GSB.

HANSEN, L., P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

HANSEN, L., P., J. HEATON ET A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-280.

HAYASHI, F. (2000): *Econometrics.* Princeton University Press.

KELLER, T. ET I. OLKIN (2004): "Combining Correlated Unbiased Estimators of the Mean of a Normal Distribution," *A Festschrift for Herman Rubin, Lecture Notes-Monograph Series*, 45, 218-227.

MARMER, V. ET T.OTSU (2009): "Optimal Comparison of Misspecified Moment Restriction Models," *Cowles Foundation Discussion Papers,* 1724.

NEWEY,W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809-837.

NEWEY,W. K. (1993): "Efficient Estimation of models with Conditional Moment Restrictions," *Handbook of Statistics*, vol. 11, eds. G.S. MADDALA, C.R. RAO ET H.D. VINODS. North Holland.

NEWEY,W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-1382.

NEWEY,W. K. ET D.F. McFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, vol. IV, eds. D.F. McFADDEN AND R.F. ENGLE III. North Holland.

NEWEY,W. K. ET K.D. WEST (1987): "A Simple, Positive Semi-de_nite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

PAKES, A., ET D.POLLARD (1989): "Simulation and the Asymptotic of Optimization Estimators," *Econometrica*, 57, 1027-1057

PARK, J. Y., ET Y.J. WHANG (2005): "A Test for Martingale Hypothesis," *Studies in Nonlinear Dynamics and Econometrics*, 9 (2), 2.

PESARAN, M. H. (1982): "A Comparison of Local Power of Alternative Tests of Non-Nested Regression Models," *Econometricsa*, 50 (5), 1287-1305.

PITMAN, E.J.G (1979): *Some Basic Theory for Statistical Inference.* London: CHAPMAN AND HALL.

ROTHENBERG, T.J. (1973): "Efficient Estimation with a priori Information," *Cowles Monograph.*

ROTHENBERG, T.J. (1974): "A Note on Two-Stage Least Squares," *Unpublished Manuscript.* UC Berkeley.

ROTHENBERG, T.J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," *Handbook of Econometrics*, vol. II, eds. Z.GRILICHES AND M.D. INTRILIGATOR North Holland.

ROTHENBERG, T.J. (1984): "Hypothesis Testing in Linear Models when the Error Covariance Matrix is Nonscalar," *Econometrica*, 52, 827-842.

SAWA, T. (1973): "Almost Unbiased Estimator in Simultaneous Equation Systems," *International Economic Review*, 14, 97-106.

STAIGER, D., ET J. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

WHITE, H. (1984): *Asymptotic Theory for Econometricians.* Academic Press.

YANG, Y. (2004): "Combining Forecasting Procedure: Some Theoretical Results," *Econometric Theory*, 20, 176-222.