



UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA SPECIALISTICA IN

SCIENZE STATISTICHE ECONOMICHE, FINANZIARIE E AZIENDALI

**LA VERIFICA DELL'IPOTESI DI STRUMENTI DEBOLI IN UN
MODELLO LINEARE**

Relatore:

Ch.mo Prof. Nunzio Cappuccio

Laureando:

Davide Pizzeghello

Matricola: 532701- SEA

Anno Accademico 2007 - 2008

INDICE

INTRODUZIONE

CAPITOLO PRIMO

Stima di un modello lineare in presenza di correlazione tra le esplicative e l'errore: Variabili strumentali e GMM	3
--	---

1.1. Variabili strumentali	3
1.2. Metodo Generalizzato dei Momenti	4

CAPITOLO SECONDO

Gli strumenti deboli nel modello di regressione lineare	7
--	----------

2.1. Il modello di regressione lineare a variabili strumentali con un singolo regressore.	7
2.1.1. Il parametro di concentrazione	8
2.1.2. Un'espressione per lo stimatore TSLS	8
2.2. Approssimazioni asintotiche	11
2.2.1. Espansioni di Edgeworth	11
2.2.2. Teoria asintotica con molti strumenti	12
2.2.3. Teoria asintotica con strumenti deboli	12

CAPITOLO TERZO

Individuazione degli strumenti deboli	15
--	-----------

3.1. La statistica F al primo stadio	15
3.1.1. Accertarsi se gli strumenti sono deboli usando la statistica F al primo stadio in presenza di un solo regressore endogeno	16
3.1.2. Estensione della statistica F al primo stadio a $n > 1$	17

3.2.	R^2 parziale: un regressore endogeno	18
3.3.	R^2 parziale: più regressori endogeni	19
3.4.	Test di Anderson	19
3.5.	Un test con ipotesi nulla di strumenti forti	20

CAPITOLO QUARTO

Inferenza robusta con strumenti deboli	21
---	----

4.1.	Una famiglia di test Gaussiani pienamente robusti	21
4.1.1.	Statistica di Anderson-Rubin	23
4.1.2	Statistica LM	24
4.1.3.	La statistica di Wald	24
4.1.4.	Test rapporto di verosimiglianza condizionale	25
4.2.	Comparazioni di potenza	25
4.3.	Intervalli di confidenza Robusti	27

CAPITOLO QUINTO

Inferenza parzialmente robusta con strumenti deboli	29
--	----

5.1.	Stimatori della classe k	29
5.2.	Stimatori LIML: limited information maximum likelihood	29
5.3.	Stimatore Fuller-K	30
5.4.	Minimi quadrati a due stadi con distorsione aggiustata	30
5.5.	Variabili strumentali di Jackknife	30
5.6.	Comparazioni	31

CAPITOLO SESTO

Esempio empirico	33
-------------------------	----

CONCLUSIONI	49
--------------------	----

BIBLIOGRAFIA	
---------------------	--

LA VERIFICA DELL'IPOTESI DI STRUMENTI DEBOLI IN UN MODELLO LINEARE

INTRODUZIONE

Gli strumenti deboli si verificano quando gli strumenti nella regressione lineare a variabili strumentali (IV) sono debolmente correlati con le variabili endogene incluse. Nel metodo generalizzato dei momenti (GMM), più in generale, gli strumenti deboli corrispondono alla debole identificazione di alcuni o tutti i parametri ignoti. La presenza di debole correlazione porta a statistiche GMM con distribuzioni non normali, anche in grandi campioni, di conseguenza la convenzionale inferenza con IV e GMM non è più utile. Fortunatamente, diverse procedure sono ora disponibili per identificare e trattare gli strumenti deboli nei modelli IV lineari e più in generale in quelli GMM.

La tesi è suddivisa in due parti principali. Nella prima viene presentato il fenomeno degli strumenti deboli da un punto di vista teorico, elencando, a partire da un modello lineare, i problemi che si riscontrano nel caso si sia in presenza di debole correlazione. Dopo una breve argomentazione sulle principali caratteristiche del modello a variabili strumentali e sulle differenze con lo stimatore GMM, si elencheranno i principali test per individuare la possibile presenza di strumenti deboli in un modello lineare, definendo il concetto di parametro di concentrazione e fornendo due definizioni di riferimento. Successivamente, si elencheranno i test più o meno robusti per fare inferenza, in particolare verifiche d'ipotesi sui coefficienti e costruzione di intervalli di confidenza, oltre ad una breve approfondimento sulle approssimazioni asintotiche.

Nella seconda parte viene trattato un esempio empirico per utilizzare i vari test proposti e verificare la presenza di strumenti deboli. Si utilizzerà la funzione di reazione della Banca Centrale americana, il cui modello iniziale è riconducibile ad un modello lineare stimato attraverso lo stimatore a variabili strumentali. In particolare, verranno effettuati prima una serie di test per verificare l'eventuale presenza di eteroschedasticità ed incorrelazione degli errori, oltre alla validità e alla informatività

degli strumenti, in modo da poter decidere se imporre ulteriori limitazioni e stimare con uno stimatore GMM. Una volta definito il modello, si analizzeranno i veri e propri test per la presenza di strumenti deboli, attraverso un test F e lo studio dell'indice di correlazione parziale, ed eventualmente si potranno utilizzare stimatori interamente o solo parzialmente robusti alla debolezza degli strumenti.

CAPITOLO PRIMO

Stima di un modello lineare in presenza di correlazione tra le esplicative e l'errore: Variabili strumentali e GMM

L'applicazione delle variabili strumentali (IV) nel contesto del modello di regressione lineare classico, è abbastanza semplice: se la distribuzione dell'errore non può essere considerata indipendente da quella dei regressori, allora si fa ricorso alle variabili strumentali attraverso un appropriato set di strumenti. Tuttavia, le ricerche applicate spesso devono confrontare diverse possibilità in questo contesto.

Un problema frequente nei lavori empirici è dato dall'eteroschedasticità. Sebbene la consistenza delle stime dei coefficienti di IV non subisca tale presenza, le stime a variabili strumentali degli errori standard sono inconsistenti impedendo una corretta inferenza. Le usuali forme dei test diagnostici per endogenità e restrizioni di sovraidentificazione non sono quindi valide in presenza di eteroschedasticità. Questi problemi possono essere parzialmente indirizzati attraverso l'utilizzo di statistiche ed errori standard robusti o consistenti all'eteroschedasticità. Il convenzionale stimatore IV, sebbene consistente, risulta essere inefficiente.

L'approccio più usuale di fronte ad un caso di eteroschedasticità di forma ignota consiste nell'utilizzo del Metodo Generalizzato dei Momenti (GMM), stimatore che sfrutta le condizioni di ortogonalità per ottenere stime efficienti.

Molti stimatori standard, inclusi IV e OLS, possono essere visti come casi particolari di GMM.

1.1. Variabili strumentali

Sia dato il modello:

$$y_t = \beta'x_t + u_t \text{ con } E[x_t u_t] \neq 0$$

La presenza di correlazione tra le una o più variabili esplicative e l'errore rende le più semplici stime OLS e MQG inconsistenti ed inefficienti.

Per ovviare a questi problemi vengono utilizzate $p \geq k$ variabili strumentali z_t incorrelate con l'errore $E[z_t u_t] = 0$. Di conseguenza vengono suddivise le variabili in $[X_1, X_2]$, con K_1 regressori di cui X_1 endogeni e i $(K-K_1)$ regressori rimanenti assunti come esogeni.

Il set di strumenti è dato da Z , di dim. $n \times L$, questo rappresenta l'insieme completo delle variabili assunte come esogene, cioè tali che $E[Z_i u_i] = 0$. Esso viene suddiviso in $[Z_1, Z_2]$, dove gli L_1 strumenti Z_1 sono strumenti esclusi, e i rimanenti $(L-L_1)$ strumenti $Z_2 \equiv X_2$ sono quelli inclusi o regressori esogeni.

Riassumendo si ha:

Regressori: $X = [X_1, X_2] = [X_1, Z_2] = [\text{Endogeni Esogeni}]$

Strumenti: $Z = [Z_1, Z_2] = [\text{Esclusi Inclusi}]$

La condizione d'ordine per l'identificazione dell'equazione è $L \geq K$: devono esserci almeno tanti strumenti esclusi quanti sono i regressori endogeni. Se $L=K$ l'equazione è detta esattamente identificata; nel caso in cui $L>K$ si parla di sovraidentificazione.

Definita P_Z la matrice di proiezione $Z(Z'Z)^{-1}Z'$, lo stimatore β a variabili strumentali è dato da:

$$\beta_{IV} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y = (X'P_ZX)^{-1}X'P_Zy$$

Questo stimatore va sotto diversi nomi: a variabili strumentali (IV), stimatore generalizzato a variabili strumentali (GIVE), o minimi quadrati a due stadi (TSLS), che riflette il fatto che lo stimatore può essere calcolato attraverso una procedura a due passi.

1.2. Metodo Generalizzato dei momenti

Nel calcolo dello stimatore a variabili strumentali si ipotizza che gli errori siano iid. Tuttavia, quando questa condizione non è soddisfatta tali stimatori producono stime consistenti ma non efficienti. Il termine d'errore ha media zero e matrice di

varianza e covarianze Ω , di dimensione $n \times n$. Di essa si considerano vari casi: omoschedasticità ed incorrelazione, eteroschedasticità ed eteroschedasticità più autocorrelazione.

Gli L strumenti forniscono L condizioni sui momenti,

$$g_i(\hat{\beta}) = Z_i' \hat{u}_i = Z_i' (y_i - X_i \hat{\beta})$$

dove g_i è $L \times 1$. L'esogenità degli strumenti significa che ci sono L condizioni sui momenti, o condizioni di ortogonalità, che saranno soddisfatte con il vero valore di β .

$$E(g_i(\beta)) = 0$$

ciascuna delle L equazioni corrisponde ad un momento campionario che possono essere scritti come

$$\bar{g}(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n Z_i' (y_i - X_i \hat{\beta}) = \frac{1}{n} Z' \hat{u}$$

L'intuizione che sta dietro a GMM consiste nello scegliere uno stimatore per β che risolve $\bar{g}(\hat{\beta}) = 0$.

Se l'equazione da stimare è esattamente identificata, e quindi $L = K$, è possibile risolvere tale equazione, e lo stimatore GMM è di fatto lo stimatore IV.

Nel caso di sovraidentificazione, si ottengono più equazioni che rendono impossibile eguagliare a zero le condizioni sui momenti. In questo caso, si prende una matrice pesata W di dim. $L \times L$ e la si utilizza per costruire una forma quadratica nelle condizioni dei momenti. Si definisce così la funzione obiettivo come:

$$J(\hat{\beta}) = n \bar{g}(\hat{\beta})' W \bar{g}(\hat{\beta})$$

Uno stimatore GMM per β è il valore di $\hat{\beta}$ che minimizza $J(\hat{\beta})$. Derivando rispetto a $\hat{\beta}$ e risolvendo le K condizioni del primo ordine si ottiene lo stimatore GMM

$$\hat{\beta}_{GMM} = (X' Z W Z' X)^{-1} X' Z W Z' y$$

La matrice W è scelta imponendo diverse assunzioni restrittive in base alla struttura del termine d'errore e alla relativa stima della matrice di varianze e covarianze.

L'esogenità degli strumenti è solo uno dei due requisiti necessari agli strumenti per essere validi. Recentemente la rilevanza degli strumenti, ha ricevuto maggiori attenzioni dai teorici che hanno rivolto l'attenzione verso gli strumenti che sono solo debolmente correlati con le variabili endogene incluse in quanto possibile causa di conseguenze negative quali la poca precisione dello stimatore a variabili strumentali, l'elevata distorsione o un'approssimazione normale della sua distribuzione campionaria molto povera.

CAPITOLO SECONDO

Gli strumenti deboli nel modello di regressione lineare.

Molti dei problemi presentati a causa degli strumenti deboli nel modello di regressione lineare a variabili strumentali sono molto ben spiegati nel ambito del classico modello con variabili esogene fissate ed errori normali IID.

Verrà utilizzato quindi questo modello per mostrare come gli strumenti deboli portino ad ottenere lo stimatore dei minimi quadrati a due stadi (TSLS) con una distribuzione campionaria non normale, indipendentemente dalla dimensione del campione.

2.1. Il modello di regressione lineare Gaussiano a variabili strumentali con un singolo regressore.

Dato il modello:

$$y = x\beta + u \quad \text{eq. strutturale}$$

e

$$X = Z\pi + V \quad \text{eq. al primo stadio}$$

dove y e x sono vettori $T \times 1$ delle osservazioni sulle variabili endogene, z è la matrice degli strumenti $T \times K$, e u e v sono vettori $T \times 1$ dei termini di errore. Gli strumenti sono assunti essere non casuali (fissati) in campionamenti ripetuti. Si assume inoltre che gli errori $[u_t, v_t]'$, ($t = 1, \dots, T$) siano IID $N(0, \Sigma)$, dove gli elementi di Σ sono σ_u^2 , σ_v^2 , σ_{uv} , e sia $\rho = \sigma_{uv} / \sigma_u \sigma_v$. La prima equazione rappresenta l'equazione strutturale, e β è il parametro scalare di interesse. L'equazione al primo stadio mette in relazione il regressore endogeno con gli strumenti.

Lo stimatore e la distorsione sono dati rispettivamente da $\hat{\beta}_{SIV} = \frac{z'y}{z'x}$ e da

$$\hat{\beta}_{SIV} - \beta = \frac{z'y}{z'x} - \frac{z'u}{z'(z\pi + v)}$$

2.1.1. Il parametro di concentrazione

Il parametro di concentrazione, μ^2 , è un indice che misura la forza degli strumenti ed è definito come:

$$\mu^2 = \frac{\pi^2 Z'Z}{\sigma_v^2}$$

Un'utile interpretazione di μ^2 è in termini di F , la statistica utilizzata per testare l'ipotesi $\pi = 0$ nell'equazione al primo stadio (statistica F al primo stadio). Sia $\tilde{F}$ la controparte infattibile di F , calcolata usando il vero valore di σ_v^2 . Allora $K\tilde{F}$ è distribuito come una distribuzione χ^2 non centrata con K gradi di libertà e parametro di non centralità μ^2 , ed $E(\tilde{F}) = \frac{\mu^2}{k+1}$. Se la dimensione del campione è grande, allora i valori di F e $\tilde{F}$ sono vicini, quindi $E(\tilde{F}) \cong \frac{\mu^2}{k+1}$. Così, valori più grandi di $\frac{\mu^2}{k}$ traslano la distribuzione del test F al primo stadio, e $F-1$ può essere pensato come uno stimatore di $\frac{\mu^2}{k}$.

2.1.2. Un'espressione per lo stimatore dei minimi quadrati a due stadi.

Lo stimatore TSLS è $\beta^{TSLS} = (X'P_z y) / (X'P_z X)$ dove $P_z = z'(z'z)^{-1}z'$. Attraverso alcune sostituzioni, Rothemberg (1984) ha proposto un'espressione molto utilizzata per lo stimatore TSLS:

$$\mu(\tilde{\beta}^{TSLS} - \beta) = (\sigma_u^2 / \sigma_v^2) \frac{\zeta_u + S_{uv}/\mu}{1 + 2(\zeta_v/\mu) + (S_{vv}/\mu^2)}$$

dove ζ_u e ζ_v , con coefficiente di correlazione pari a ρ e

$$\zeta_v = \frac{\Pi'Zu}{\sigma_u \sqrt{\Pi'Z'Z\Pi}} \sim N(0;1) \quad \zeta_u = \frac{\Pi'Zv}{\sigma_v \sqrt{\Pi'Z'Z\Pi}} \sim N(0;1)$$

$$S_{uv} = \frac{v'P_z u}{\sigma_u \sigma_v}, \quad S_{vv} = \frac{v'P_z v}{\sigma_v^2}$$

Sotto le assunzioni di strumenti fissi ed errori Normali, z_u e z_v sono variabili casuali normali standard con correlazione ρ , e S_{uv} e S_{vv} sono forma quadratiche di variabili casuali normali rispetto alla matrice idempotente P_z .

Poiché le distribuzioni di z_u , z_v , S_{uv} e S_{vv} non dipendono dalla dimensione T del campione, essa entra nella distribuzione di β^{TSLs} solo attraverso il parametro di concentrazione. Se μ^2 è piccolo, allora, i termini z_v , S_{uv} e S_{vv} portano ad una distribuzione non normale della presedente espressione. Al contrario, il termine principale z_u è predominante se μ^2 è grande, producendo così l'usuale approssimazione di β^{TSLs} alla distribuzione normale. Formalmente, μ^2 gioca il ruolo solitamente giocato dal numero di osservazioni: come μ^2 cresce, la distribuzione di $\mu(\hat{\beta}^{TSLs} - \beta)$ è sempre meglio approssimata da una distribuzione normale $N(0; \frac{\sigma_u^2}{\sigma_v^2})$.

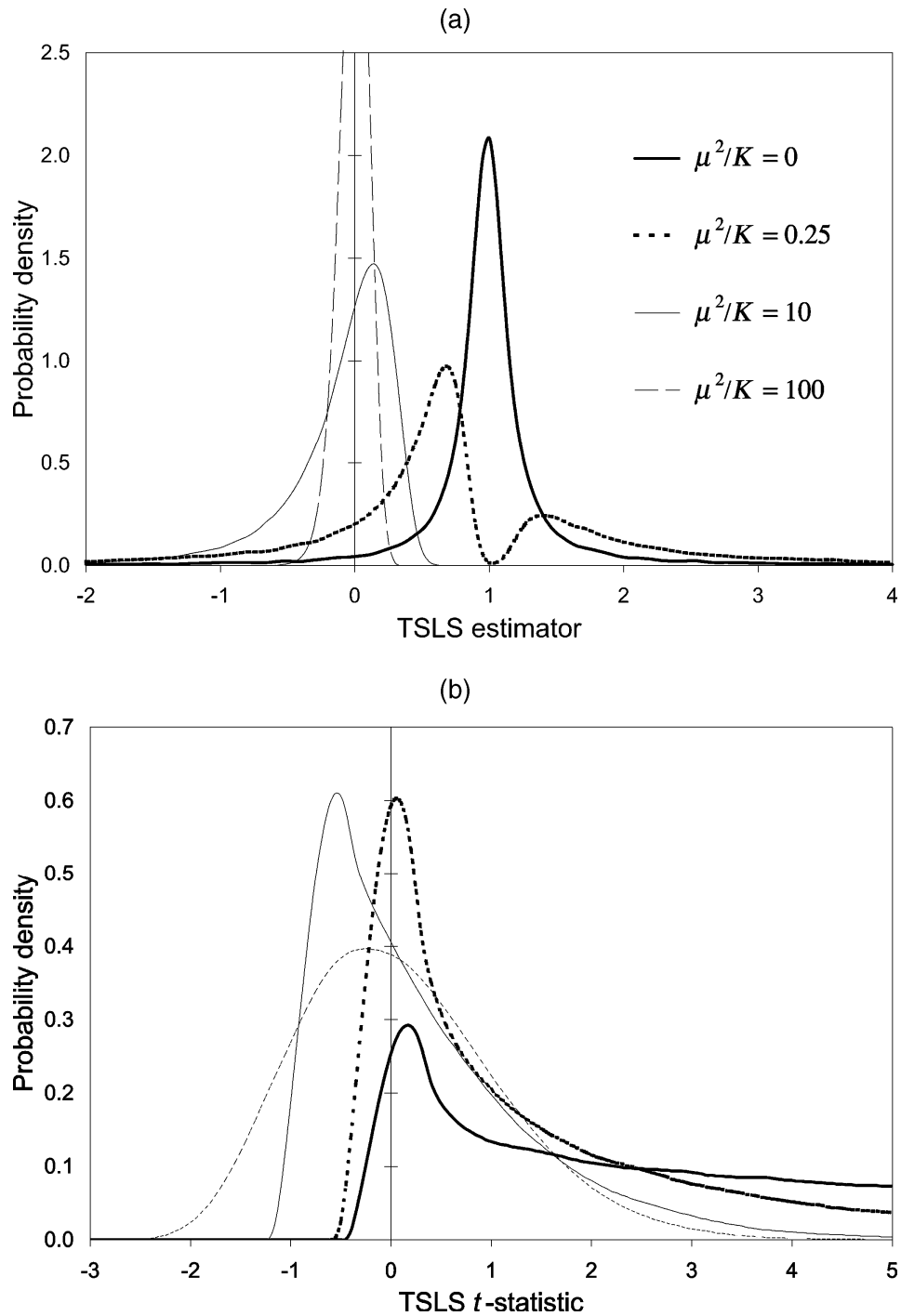
Quindi, affinché l'approssimazione normale alla distribuzione dello stimatore TSLs sia accurata, il parametro di concentrazione deve essere grande.

Quando $\mu^2=0$, o equivalentemente $\Pi=0$, gli strumenti sono non solo deboli, ma addirittura irrilevanti. In questo caso, la media dello stimatore TSLs è il limite in probabilità dello stimatore OLS, $plim(\hat{\beta}^{OLS})$. In particolare, quando $K \geq 3$ allora questa media esiste, $E(\hat{\beta}^{TSLs} - \beta) = plim(\hat{\beta}^{OLS} - \beta) = \sigma_{ux} / \sigma_x^2$.

Quando invece gli strumenti sono rilevanti ma deboli, lo stimatore TSLs è distorto verso $plim(\hat{\beta}^{OLS})$. In particolare, si definisce la "distorsione relativa" di TSLs come la distorsione di TSLs relativa all'inconsistenza di OLS, cioè, $E(\hat{\beta}^{TSLs} - \beta) / plim(\hat{\beta}^{TSLs} - \beta)$. Quando μ^2 è moderatamente grande, la distorsione relativa di TSLs è approssimativamente inversamente proporzionale a $\mu^2/(K-2)$, risultato che vale anche se gli errori non sono normalmente distribuiti.

Le figure qui di seguito mostrano lo stimatore TSLs e la sua statistica t per diversi valori del parametro di concentrazione quando il vero valore di β è 0 e gli tutti gli altri valori sono fissati.

Figura1. Densità della distorsione della sua statistica t per lo stimatore a variabili strumentali con $\beta=0$, $\rho=0.99$, $\sigma_u^2 = \sigma_v^2 = 1$, $p=k=1$, $T=500$. Calcolati tramite una simulazione Monte Carlo



Per piccoli valori di μ^2/k , ad esempio 0.25, le distribuzioni sono sorprendentemente non normali, perfino bimodali. Al crescere di μ^2/k , le distribuzioni assumono un comportamento sempre più simile alla distribuzione normale.

2.2. Approssimazioni asintotiche

Le approssimazioni asintotiche convenzionali in campioni finiti di solito sono calcolate mantenendo fissato il modello e facendo tendere T tende a infinito, tuttavia a volte questo approccio non conduce alla più utile distribuzione approssimata. Questo è il caso degli strumenti deboli; come in fig1, le usuali approssimazioni asintoticamente normali, possono essere abbastanza povere se il parametro di concentrazione è piccolo, anche se il numero di osservazioni è elevato. Per questa ragione, metodi asintotici alternativi vengono utilizzati per analizzare le statistiche IV in presenza di strumenti deboli. Tre metodi simili sono *le espansioni di Edgeworth*, *le approssimazioni con molti strumenti* e *quelle a strumenti deboli*. Questi metodi si propongono di migliorare la qualità delle approssimazioni quando il campione è grande ma il parametro di concentrazione non lo è.

2.2.1. Espansioni di Edgeworth

Un'espansione di Edgeworth è una rappresentazione della distribuzione della statistica di interesse in powers di $1/\sqrt{T}$. Come indicato da Rothemberg (1984) negli strumenti fissi, con modello ad errori normali, un'espansione di Edgeworth $1/\sqrt{T}$ è formalmente equivalente ad un'espansione di Edgeworth $1/\sqrt{\mu}$. In questo senso, le espansioni di Edgeworth fanno meglio della convenzionale approssimazione quando μ è piccolo abbastanza affinché importi il termine in $1/\mu^2$, ma non così piccolo che i termini $1/\mu^3$ e i più alti importino. Rothemberg ha suggerito che l'approssimazione di Edgeworth è "eccellente" per $\mu^2 > 50$ e "adequata" per μ^2 almeno 10, finché il numero degli strumenti è piccolo (meno di μ).

2.2.2. Teoria asintotica con molti strumenti

Sebbene i problemi dei molti strumenti e degli strumenti deboli possano sembrare simili, di fatto sono solo “imparentati”. Con molti strumenti forti, l' R^2 corretto, della regressione al primo stadio sarebbe vicina ad 1, così un piccolo R^2 corretto al primo stadio indica che gli strumenti, presi come gruppo, sono deboli. Bekker(1994) ha formalizzato questa nozione sviluppando approssimazioni asintotiche per una sequenza di modelli con strumenti fissi ed errori normali, nei quali il numero degli strumenti K , è proporzionale alla dimensione del campione e $\frac{\mu^2}{K}$ converge in limite ad un valore costante e finito. Le distribuzioni asintotiche con molti strumenti sono generalmente normali, l'evidenza delle simulazioni suggerisce che queste approssimazioni sono buone sia per moderati che per grandi valori di K , sebbene non catturino la non normalità. Le distribuzioni derivate usando questo approccio generalmente dipendono dalla distribuzione degli errori (Bekker 1999), così alcune procedure che sono giustificate usando i molti strumenti asintotici richiedono aggiustamenti per gli errori non normali. Tuttavia, i risultati su andamento e consistenza sono più robusti rispetto la condizione di non normalità. (chao Swanson 2002)

2.2.3. Teoria asintotica con strumenti deboli

Staiger e Stock (1997) hanno sviluppato una teoria asintotica che si basa sul fatto che, nel modello con p strumenti, il vettore dei coefficienti π dipenda dalla numerosità campionaria.

Dato il modello:

$$y = x\beta + u \quad \text{eq. strutturale}$$

$$x = Z\pi_T + v \quad \text{eq. del primo stadio}$$

$$\pi_T = \frac{c}{\sqrt{T}}$$

π_T è detto Drift di Pitman, necessario per studiare la potenza asintotica del test F sugli elementi del vettore dei coefficienti; in particolare è possibile vedere cosa