

APPENDICE 2

Aspetti Metodologici

Prima di procedere alla specificazione dei modelli da sottoporre ad analisi di regressione, avendo i nostri dati una dimensione temporale importante (9 anni), abbiamo indagato l'eventuale presenza di trend nelle variabili e la loro natura (stocastica o deterministica) attraverso specifici test sulla radice unitaria per dati panel. La scelta è ricaduta sui test di tipo Fisher proposti da Maddala e Wu (1999) e Choi (2001) poichè essi hanno il vantaggio, rispetto ad altri test famosi come quelli di Levin, Lin e Chu (2002) e Im, Pesaran e Shin (2003), di essere applicabili a *dataset* "sbilanciati" come il nostro¹. Molti studi antecedenti hanno analizzato gli effetti pre-elettorali delle politiche fiscali utilizzando modelli ad effetti fissi dinamici (Schuknecht 2000; Shi e Svensson 2003; Brender e Drazen 2004; Katsimi e Sarantides 2012). La struttura econometrica più utilizzata per indagare sugli effetti delle elezioni sulle variabili fiscali è la seguente:

$$y_{i,t} = \alpha + \beta y_{i,t-1} + \gamma Ele_{i,t-1} + \delta Z_{i,t} + \mu_i + \lambda_t + \varepsilon_{i,t} \quad (1)$$

Dove $y_{i,t}$ è una variabile fiscale (nel nostro caso, la spesa comunale in conto capitale pro – capite e la spesa totale comunale pro – capite); $Ele_{i,t-1}$ è una variabile *dummy* che assume valore 1 nell'anno precedente a quello in cui si tengono le elezioni nel comune i ; $Z_{i,t}$ è il vettore degli effetti specifici di ogni comune e delle variabili di controllo socio-economiche. γ è il nostro parametro di interesse che "cattura" come lo strumento fiscale osservato si comporta in prossimità delle elezioni. Infine, μ_i e λ_t sono rispettivamente gli effetti fissi comunali e gli effetti fissi temporali e $\varepsilon_{i,t}$ è il termine d'errore.

L'inclusione della variabile dipendente ritardata nella (1) introduce una distorsione nel modello che non può essere eliminata tramite una regressione ad effetti fissi. Lo stimatore OLS, per essere non distorto, richiede che ogni variabile esplicativa (formalizzata come un vettore colonna di osservazioni) deve essere incorrelata con il vettore degli errori. Nell'equazione (1) questa assunzione è violata poiché il vettore della variabile dipendente ritardata $\begin{matrix} \rightarrow \\ y_{i,-j} \end{matrix}$ è correlato con il vettore $\begin{matrix} \rightarrow \\ \varepsilon_i \end{matrix}$. Per esempio, per un dato comune i , il t^{th} elemento del vettore $\begin{matrix} \rightarrow \\ y_{i,-1} \end{matrix}$, e il $(t-1)^{th}$ elemento del vettore $\begin{matrix} \rightarrow \\ \varepsilon_i \end{matrix}$, $\varepsilon_{i,t-1}$, sono correlati nella (1).

L'approccio comune per trattare dati non-stazionari è quello di applicare l'operatore differenza in modo da pervenire ad una specificazione dinamica in differenze prime. L'equazione da stimare assumerà così la forma:

$$\Delta y_{i,t} = \alpha + \beta \Delta y_{i,t-1} + \gamma Ele_{i,t-1} + \delta \Delta Z_{i,t} + \lambda_t + \varepsilon_{i,t} \quad (2)$$

Nella (2) abbiamo applicato una trasformazione in differenze prime sia alla variabile dipendente che a tutti i regressori, ad esclusione delle variabili elettorali, misurati inizialmente in livello. Applicando le differenze prime gli effetti fissi comunali vengono eliminati, ma non gli effetti fissi temporali.

L'inclusione della variabile dipendente ritardata tra i regressori elimina spesso il problema della correlazione seriale con il termine di errore, essa introduce però nel modello di regressione una

¹ Nella scelta del numero dei ritardi da includere nel modello che sottopone a verifica la presenza di radice unitarie, al fine di rendere più parsimonioso il modello in termini di numero di parametri da stimare e per evitare il rischio di sovrapparametrizzazione (*overfitting*), è utile ricorrere a criteri che tentano di evitare tale eventualità assegnando un "costo" all'introduzione di ogni parametro addizionale. La scelta del numero dei ritardi è fondamentalmente una questione empirica. Una regola del pollice molto usata è quella di inserire un numero di ritardi pari a circa 1/3 o 1/4 la lunghezza della serie. Il modo di operare migliore è quello di partire con un numero "sufficientemente" ampio di ritardi e ridurlo progressivamente sulla base dei criteri informativi. Il criterio da noi scelto è l'AIC (*Akaike information Criterion*).

potenziale distorsione a causa del venir meno dell'ipotesi di esogeneità forte del termine d'errore $\varepsilon_{i,t}$.

Al fine di correggere tale distorsione, Arellano e Bond (1991), Arellano e Bover (1995) e Blundell e Bond (1998) hanno sviluppato degli stimatori basati sul metodo generalizzato dei momenti, per regressioni con dati panel dinamici. L'idea sottostante è di trovare una variabile strumentale per l'equazione (2) che sia correlata con $\Delta y_{i,t-1}$ ma non con $\Delta \varepsilon_{i,t}$. Arellano e Bond (1991) evidenziano come, sotto l'ipotesi che il termine d'errore $\Delta \varepsilon_{i,t}$ non sia serialmente correlato, il valore della variabile dipendente, ritardato di due o più periodi, soddisfa questi requisiti e potrebbe essere usato come strumento per $\Delta y_{i,t-1}$. Come analizzato in letteratura, la distorsione di questa formulazione è di ordine $1/T$, anche se il numero di Paesi (in questo caso i comuni calabresi) diventa grande. La distorsione dello stimatore ad effetti fissi diminuisce così solamente se $T \rightarrow \infty$ (vedi, tra gli altri, Nickell, 1981 e Kiviet, 1995). Lo stimatore ad effetti fissi è distorto ma consistente, riteniamo però che non sia il metodo più adatto da applicare nel nostro caso.

Sebbene lo stimatore GMM (basato sull'equazione in differenze prime) fornisce delle stime consistenti, Arellano e Bover (1995) e Blundell e Bond (1998) mostrano come, sulla base di simulazioni statistiche, questo stimatore presenta una scarsa precisione in campione finiti. L'intuizione dietro è che quando le variabili esplicative sono persistenti nel tempo, i valori ritardati di queste variabili sono debolmente correlati con le differenze di queste variabili nell'equazione di regressione differenziata. In modo da aumentare la precisione delle stime, Arellano e Bover (1995) e Blundell e Bond (1998) propongono di combinare la regressione differenziata con la regressione originale in livello. Gli strumenti per la regressione in differenze sono quelli precedentemente descritti, mentre gli strumenti per la regressione in livello sono i valori ritardati delle variabili dipendenti differenziate. Lo svantaggio di quest'ultimo stimatore è che esso riduce la dimensione del campione a causa della sensibilità degli strumenti interni (le variabili esplicative ritardate), inoltre, le sue proprietà in campioni di piccola dimensione sono generalmente sconosciute.

Lo stimatore di Arellano e Bond (1991), come primo passo, trasforma tutti i regressori tramite differenziazione e usa il metodo generalizzato dei momenti, viene perciò chiamato *Difference GMM*.

Lo stimatore di Arellano-Bover (1995)/Blundell-Bond (1998) costruisce un sistema di due equazioni, l'equazione originale e quella trasformata, viene perciò denominato *System GMM*.

Entrambi gli stimatori GMM forniscono stime consistenti in panel in cui T è piccolo e N è ampio. Riteniamo così che il ricorso a questi due stimatori per dati panel dinamici sia coerente con le caratteristiche dei nostri dati e molto più appropriato rispetto allo stimatore ad effetti fissi.

Nella nostra specificazione econometrica abbiamo introdotto tra i regressori la *variabile dipendente ritardata di due periodi*; le variabili demografiche: *popolazione*, *densità*² e *indice di dipendenza strutturale*³; *i trasferimenti correnti da altri livelli di governo e il reddito imponibile pro-capite*.

La dimensione della popolazione e la densità influenzano la spesa delle amministrazioni comunali. Un segno positivo di questi coefficienti è coerente con la predizione che un aumento della popolazione comporta un aumento della quantità di beni e servizi offerti dalle pubbliche amministrazioni, mentre un loro segno negativo evidenzierebbe la presenza di economie di scala. Considerando le varie fasce d'età della popolazione, un incremento nel numero di cittadini sotto i 10 anni o un aumento del numero di anziani molto probabilmente implicherebbero un aumento delle spese per servizi sociali; l'eventuale aumento della spesa pubblica potrebbe essere così trainato da programmi sociali.

I trasferimenti correnti e il reddito imponibile pro-capite vengo utilizzati come indicatori delle risorse finanziarie pubbliche e private.

Come *proxy* del ciclo elettorale utilizziamo una variabile *dummy* che assume valore 1 nell'anno precedente alle elezioni comunali e 0 altrimenti. È consigliabile usare l'*anno pre-elettorale* nelle

² concentrazione di individui espressa dal rapporto tra il numero di abitanti e la superficie del territorio (numero di abitanti per chilometro quadrato).

³ frazione della popolazione con meno di 10 anni e più di 65.

analisi di questo tipo, poiché l'anno elettorale può risultare un indicatore distorto: se le elezioni si svolgono "presto" (nei primi mesi dell'anno) tale variabile potrebbe catturare solo l'effetto post-elettorale, mentre se le elezioni si tenessero "tardi" (nella seconda parte dell'anno) essa andrebbe a catturare solo gli effetti pre-elettorali. Poiché la maggior parte delle elezioni documentate dai nostri dati si svolgono nei primi sei mesi dell'anno, riteniamo che la variabile *dummy* anno pre-elettorale è l'indicatore più idoneo per catturare tracce di un ciclo politico-economico di natura fiscale.

Per evidenziare alcuni aspetti del comportamento politico (discussi nella prima parte del saggio), aggiungiamo le variabili *fine mandato* e *affiliazione*.

Per ognuna delle regressioni sono stati eseguiti alcuni test necessari per poter desumere la consistenza delle stime ottenute. Il primo test, fondamentale affinché si possa accettare la regressione nel suo insieme, verifica l'ipotesi nulla che non vi sia autocorrelazione di secondo ordine tra i residui di regressione. L'importanza di questo test dipende dal fatto che l'ipotesi verificata corrisponde ad una delle condizioni sulle quali si basa la costruzione degli stimatori GMM. Il venir meno di questa condizione comporta evidenza a favore della presenza di correlazione tra i regressori utilizzati nella matrice degli strumenti e la componente idiosincratca dell'errore, così da rendere le stime ottenute inconsistenti.

Il secondo test, il test di Hansen, verifica l'ipotesi che le restrizioni in sovra-identificazione siano tutte valide. L'ipotesi testata è che le variabili strumentali siano specificate correttamente, e quindi siano strumenti accettabili ed efficientemente utilizzabili nel modello.

In tutte e quattro le regressioni presentate nella **Tabella 4**, il test di Arellano-Bond sull'autocorrelazione dei residui del secondo ordine ci conferma che non possiamo rifiutare l'ipotesi nulla di assenza di autocorrelazione; non abbiamo perciò problemi da questo punto di vista. Sulla base del test di Hansen gli strumenti utilizzati risultano validi, l'alto valore del p-value, registrato soprattutto nelle regressioni con variabile dipendente la spesa in conto capitale pro-capite, può essere un campanello d'allarme sul fatto che stiamo utilizzando troppi strumenti.

La variabile dipendente ritardata risulta sempre statisticamente significativa, tranne nel caso del 2° lag delle spese totali pro-capite stimate con *il System Gmm*, e registra sempre segno negativo; da ciò si può desumere un uso anticiclico di tale variabile rispetto ai suoi valori passati.

La popolazione e l'indice di dipendenza strutturale risultano in tutte le regressioni non significative. La densità è statisticamente significativa solo nella regressione con *Difference GMM* applicata alle spese in conto capitale pro-capite e il suo coefficiente è negativo; da ciò si può evincere la presenza di economia di scala, ovvero una riduzione dei costi dei beni e servizi forniti dalle amministrazioni comunali in seguito all'aumento della concentrazione di abitanti per km², probabile sintomo questo di un guadagno di efficienza.

I trasferimenti correnti da altri livelli del governo e il reddito imponibile (entrambe le variabili espresse in termini pro-capite) risultano significative solamente nelle due regressioni dove la variabile dipendente sono le spese totali pro-capite. In questo caso, entrambi i coefficienti hanno segno positivo e ciò significa che un aumento delle risorse finanziarie, pubbliche e private, ha un impatto positivo sulle spese totali pro-capite e può contribuire ad un loro aumento.

Concentriamo adesso la nostra attenzione sulle variabili che analizzano il comportamento dei sindaci in carica. La variabile anno pre-elettorale risulta statisticamente significativa e con coefficiente positivo in tutte e quattro le regressioni. Nell'anno che precede le elezioni comunali calabresi si ha così un aumento cospicuo sia delle spese in conto capitale che delle spese totali (nelle quali le prime vi rientrano). Anche la variabile *affiliazione* risulta sempre significativa e in tutte le regressioni registra segno negativo. Questo risultato conferma, anche nel caso dei comuni calabresi, la tesi di Van Houten (2009), secondo la quale i sindaci affiliati a partiti politici nazionali più che attuare strategia opportunistiche volte ad aumentare le probabilità di rielezione tendono invece ad avere l'obiettivo di costruirsi una reputazione di operatori politici "fiscalmente responsabili" (*fiscally responsible*)⁴. La variabile *fine mandato* assume sempre segno positivo ma

⁴ Sottolineiamo, nuovamente, che i sindaci calabresi affiliati a partiti politici nazionali, nel periodo esaminato, sono un numero limitato.

non risulta significativa in nessuna regressione. Tale coefficiente suggerisce che un sindaco non più candidabile tende a “spendere” di più rispetto a sindaci ricandidabili, in modo così da poter agevolare l’elezione del suo candidato prediletto; non possiamo però avallare questa tesi a causa della mancanza di significatività statistica del regressore.