

# La Stima Rapida dei Conti Economici Territoriali

Tommaso Proietti

*Dipartimento di Scienze Statistiche, Università di Udine*

*Via Treppo 18, 33100 Udine, e-mail: proietti@dss.uniud.it*

**Sintesi:** la relazione illustra i principali aspetti metodologici concernenti la stima anticipata dei conti regionali degli anni 2000 e 2001. La stima preliminare ed il successivo bilanciamento al totale nazionale sono ottenuti a partire da un modello dinamico per dati panel con eterogeneità regionale.

**Parole chiave:** Bilanciamento; dati panel; eterogeneità regionale; analisi fattoriale; combinazione delle stime.

## 1 Introduzione

La relazione illustra la metodologia utilizzata per la stima rapida dei principali aggregati dei conti economici regionali per gli anni 2000 e 2001. Gli aggregati economici oggetto di stima sono stati i seguenti: i. Valore aggiunto (VA) ai prezzi di base e PIL ai prezzi di mercato, in milioni di euro correnti e costanti; ii. Unità di lavoro (ULA); iii. Consumi finali interni (CFI), in milioni di euro correnti e costanti; iv. Redditi da lavoro dipendente (RLD). Le valutazioni a euro costanti sono espresse ai prezzi dell'anno 1995. La stima delle ULA è stata effettuata dal Dipartimento di Contabilità Nazionale e Analisi Economica dell'Istat.

Per quanto concerne il livello di disaggregazione, le stime sono state effettuate per le 20 regioni italiane, più l'unità extra-regio, laddove presente, e successivamente aggregate per le quattro ripartizioni territoriali; per il VA e i RLD la stima rapida ha riguardato le seguenti sei macrobranche di attività economica della classificazione NACE-Rev.1: 1. Agricoltura, caccia e silvicoltura; pesca e piscicoltura (ACS-PP) 2. Industria in senso stretto (ISS) 3. Costruzioni (COS) 4. Commercio, riparazione di autoveicoli e di beni per la casa; alberghi e ristoranti; trasporti e comunicazioni (CR-AR-TC) 5. Intermediazione monetaria e finanziaria; attività immobiliari, noleggio e attività professionali ed imprenditoriali (IMF-AI-NAPI) 6. Altre attività di servizi (AAS).

Infine, l'anticipazione relativa ai CFI presenta la disaggregazione seguente: a. Spesa per consumi finali delle famiglie, articolata nei 12 capitoli di spesa della classificazione COICOP, Rev.1. b. Spesa per consumi finali delle amministrazioni pubbliche. c. Spesa per consumi finali delle istituzioni senza scopo di lucro al servizio delle famiglie.

Il set informativo utilizzato ai fini della stima contiene: i conti economici territoriali secondo il SEC95 (serie completa dal 1995 al 1999); le unità di lavoro (dipendenti e totali, per branca, serie completa 1995-2000, previsione 2001) i conti economici nazionali (serie completa dal 1970 al 2001); le stime del valore aggiunto ai prezzi di base del settore ACS-PP; indicatore dell'attività edilizia; importazioni ed esportazioni regionali; una batteria di indicatori regionali coincidenti relativi al mercato del lavoro, ai consumi (immatricolazioni, indice delle vendite), al credito, ai giudizi sul livello della produzione, scorte, ordini e clima di fiducia (serie completa 1995-2001); l'indagine rapida sui conti economici delle grandi imprese; (pseudo-panel 1997-2000); la spesa per consumi delle famiglie proveniente dalla rilevazione dei bilanci di famiglia (1997-2001).

La relazione è articolata nel modo seguente: nella sezione successiva verranno considerati due aspetti essenziali del processo generatore delle stime dei conti regionali utilizzato dall'Istat: il ruolo delle unità di lavoro e il bilanciamento delle stime preliminari al totale nazionale desunto dai conti economici nazionali. La classe dei modelli utilizzati e la metodologia statistica di base viene presentata nella sezione 3. In particolare, viene discusso il problema dell'inferenza in presenza di effetti regionali fissi e casuali; per ulteriori dettagli relativi alla stima di massima verosimiglianza e le proprietà degli stimatori, i test diagnostici e di specificazione, si rimanda a Proietti (2002). La sezione 4 illustra l'estrazione di fattori sintetici rappresentativi della crescita regionale, da utilizzare come variabili esplicative delle variazioni degli aggregati regionali oggetto di stima, effettuata a partire da una batteria di indicatori disaggregati territorialmente e disponibili con la tempestività richiesta. Le sezioni 5, 7 e 8 descrivono i metodi utilizzati rispettivamente per la stima del valore aggiunto regionale, dei redditi da lavoro dipendente e dei consumi finali interni, mentre la sezione 6 discute l'impiego dell'indagine rapida sui conti economici delle grandi imprese ai fini della stima del valore aggiunto a prezzi correnti ripartizionale.

## **2 Il ruolo delle ULA ed il bilanciamento nei conti regionali**

Le stime degli aggregati di contabilità regionale sono effettuate dall'Istat mediante l'impiego delle medesime fonti e metodi della contabilità nazionale, anche se in maniera indipendente, dal momento che l'adozione della procedura detta *bottom-up*, che prevede la costruzione della stima del totale nazionale per aggregazione delle stime regionali, è limitata ad alcuni aggregati dell'agricoltura ed alla stima benchmark delle unità di lavoro. L'estensivo uso delle ULA regionali rende la metodologia di costruzione dei conti di tipo *pseudo-bottom-up*, ciò essendo espressione della insufficienza delle informazioni statistiche riferibili direttamente alle unità di attività economica a livello locale. In generale, si può dunque affermare che le stime dei conti nazionali precedono quelle dei conti regionali. Si rimanda a Pascarella (1998), per una trattazione completa di questi aspetti.

Dal punto di vista logico e temporale si possono distinguere due momenti: nel primo vengono ottenute delle stime preliminari; nel secondo queste ultime vengono rese consistenti con il totale nazionale. Con riferimento al primo, in tutte le branche dell'ISS ed in gran parte di quelle dei servizi, fatta eccezione per la parte non *market* afferente alla PA, il calcolo del valore aggiunto, redditi

da lavoro dipendente e investimenti, è effettuato attribuendo alle ULA per fascia dimensionale i valori medi per occupato desunti dalle indagini di base sui conti delle imprese<sup>1</sup>.

Denotato con  $y_t$  il vettore  $R \times 1$  dei flussi regionali concernenti un aggregato, ad esempio il VA, oggetto di stima, e supposto di avere  $m$  classi dimensionali (attualmente  $m = 5$ ) lo schema di calcolo è riassumibile come segue:

$$\hat{y}_t = \sum_{j=1}^m p_{jt} \odot l_{jt} = p_t \odot l_t, \quad p_t = \sum_{j=1}^m p_{jt} \odot (l_{jt} \div l_t), \quad (1)$$

dove  $l_{jt}$  rappresenta il numero di ULA appartenenti ad una specifica classe dimensionale,  $p_{jt}$  è il valore aggiunto per ULA (una misura della produttività del lavoro) delle imprese appartenenti alla medesima classe; i simboli  $\odot$  e  $\div$  rappresentano rispettivamente il prodotto e la divisione elemento per elemento (o di Hadamard).

La (1) rappresenta una stima preliminare; essa viene bilanciata al totale nazionale,  $Y_t = i' y_t$ , dove  $i$  denota un vettore di  $R$  elementi tutti unitari, mediante la ripartizione proporzionale delle discrepanze con il totale nazionale del flusso, generando una stima bilanciata,  $\tilde{y}_t$ , che incorpora l'informazione  $Y_t$ .

L'aggiustamento proporzionale può essere formalizzato nella maniera seguente:

$$\tilde{y}_t = \hat{y}_t \cdot \frac{Y_t}{i' \hat{y}_t} = \hat{y}_t + \hat{y}_t \frac{Y_t - i' \hat{y}_t}{i' \hat{y}_t}. \quad (2)$$

La seconda espressione chiarisce che l'aggiustamento proporzionale può essere interpretato come il migliore previsore lineare basato sul totale nazionale, qualora  $\text{Var}(\hat{y}_t) = \sigma^2 D_{\hat{y}_t}$ ,  $D_{\hat{y}_t} = \text{diag}(\hat{y}_{1t}, \dots, \hat{y}_{Rt})$ . Se  $y_t$  è normale,  $\tilde{y}_t$  minimizza l'errore quadratico medio di stima.

Se, invece, la distribuzione di  $y_t$  è lognormale, occorre che  $\text{Var}(\hat{y}_t)$  vari inversamente a  $\hat{y}_{it}$ , nel senso che chiariremo tra poco. Si supponga  $\ln \widehat{y}_t \sim \mathbf{N}(\mu_t, \Sigma)$ ; si ha allora, per le proprietà della distribuzione lognormale (Johnson e Kotz, 1972),

$$\text{Var}(\hat{y}_t) = D_{\hat{y}_t} [\exp \Sigma - i i'] D_{\hat{y}_t},$$

che è proporzionale a  $D_{\hat{y}_t}$  (come è richiesto dall'aggiustamento proporzionale) nel caso in cui

$$\Sigma = \ln \left( \sigma^2 D_{\hat{y}_t}^{-1} + i i' \right) \approx \sigma^2 D_{\hat{y}_t}^{-1}.$$

In un contesto dinamico, in cui le differenze logaritmiche sono stazionarie, ma non i livelli, si può mostrare che, affinché il bilanciamento proporzionale sia ottimale, occorre che le differenze logaritmiche presentino una forma di eteroschedasticità regionale, tale che la loro varianza varia in senso inversamente proporzionale al peso dell'aggregato regionale.

---

<sup>1</sup>Le ULA sono utilizzate anche per la regionalizzazione di alcune poste rientranti nel calcolo del VA *non market* delle AA.P e dei RLD del settore dell'agricoltura. Per il settore COS solo il VA è stimato prescindendo dalle stime dell'input di lavoro. In linea di massima, la quota delle stime che utilizzano le ULA si aggira attorno al 70%, 75%, e 85% rispettivamente per VA, RLD e gli investimenti fissi lordi.

Il bilanciamento proporzionale è inoltre robusto rispetto alla mancata specificazione e stima di un *drift* nazionale,  $\mu_t = \mu i$  invariante rispetto alle regioni. In generale, una distorsione moltiplicativa comune a tutte le regioni viene neutralizzata dall'aggiustamento proporzionale. Infine, è immediato mostrare che si conseguono stime preliminari a coefficiente di variazione costante se le differenze prime logaritmiche sono omoschedastiche.

### 3 Aspetti metodologici

La stima anticipata degli aggregati dei conti regionali coinvolge due aspetti: la stima preliminare mediante un modello statistico-economico ed il successivo bilanciamento. Essi vengono effettuati contestualmente, in quanto il modello su cui viene fondata la stima preliminare fornisce la matrice di covarianza delle stime, che poi viene impiegata per il bilanciamento al totale nazionale, fornendo il migliore previsore lineare non distorto, sotto l'ipotesi di corretta specificazione del modello, degli aggregati regionali. L'obiettivo che viene perseguito è quello di fornire, accanto alle stime puntuali, una misura della variabilità delle medesime che consenta di asserire la loro accuratezza.

L'intero processo è fondato sull'impiego di un modello di regressione dinamico con effetti regionali fissi o stocastici, formulato solitamente con riferimento alle differenze prime logaritmiche. Viene esplicitamente prevista la possibilità di avere disturbi caratterizzati da eteroschedasticità regionale e da autocorrelazione spaziale, sebbene quest'ultima non si sia rivelata significativa.

In generale, il modello risulta specificato come segue:

$$\Delta \ln y_t = \phi \Delta \ln y_{t-1} + \mu_t + X_t \beta + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \Sigma_\epsilon), \quad t = 1, \dots, T. \quad (3)$$

Si assume che il parametro autoregressivo,  $|\phi| < 1$ , sia comune a tutte le regioni e che  $X_t$  sia esogeno in senso stretto; l'esogeneità implica che la previsione di  $y_t$  può essere effettuata condizionatamente alle previsioni, determinate fuori dal modello, di  $X_t$ .

Gli effetti (*drift*) regionali,  $\mu_t$ , possono essere fissi, nel qual caso  $\mu_t = \mu$  è un vettore di costanti, o casuali, se ad esempio si assume che  $\mu_t \sim \text{NID}(0, \sigma_\mu^2 I)$ ; si veda Baltagi (1995) per una rassegna dei due approcci; ovvero possono essere variabili nel tempo, incorporando, ad. esempio, un trend temporale.

Una volta ottenuta la stima preliminare delle variazioni logaritmiche condizionata ai valori correnti dei regressori,  $\Delta \widehat{\ln} y_{T+l} = \phi \Delta \widehat{\ln} y_{T+l-1} + \widehat{\mu}_{T+l} + X_{T+l} \beta$ ,  $l = 1, 2$ , si ottiene la stima preliminare utilizzando le proprietà della distribuzione lognormale:

$$\widehat{y}_{T+l} = \exp \left( \ln y_T + \Delta \widehat{\ln} y_{T+l} + 0.5 D_\Sigma i \right), \quad (4)$$

dove  $D_\Sigma$  è la matrice diagonale formata dagli elementi sulla diagonale principale della matrice  $\Sigma(l) = \text{Var}(\Delta \widehat{\ln} y_{T+l})$ . Se, inoltre, si denota con  $D^*$  la matrice con gli elementi del vettore (4) sulla diagonale, la matrice di covarianza delle previsioni è fornita da

$$V = D^* [\exp \Sigma(l) - ii'] D^*. \quad (5)$$

La stima bilanciata viene ottenuta mediante proiezione lineare sul totale desunto dai conti economici nazionali:

$$\tilde{y}_{T+l} = \hat{y}_{T+l} + \frac{1}{i'Vi} Vi(Y_{T+l} - i'\hat{y}_{T+l}).$$

Con riferimento a (3) sorgono i seguenti problemi di natura inferenziale: l'eterogeneità della crescita media regionale (specificazione e stima di  $\mu_t$ ), la presenza e la stima di effetti inerziali (parametro autoregressivo), la selezione del set di covariate in  $X_t$ , la stima dell'eteroschedasticità regionale (specificazione e stima di  $\Sigma_\epsilon$ ). I primi due aspetti riguardano il problema di ottenere una stima preliminare non distorta, il terzo riguarda la sua efficienza ed il problema del bilanciamento al totale nazionale.

In Proietti (2002) viene illustrata l'importanza, a fini previsivi, di specificare correttamente la natura del *drift* regionale, in presenza di un panel di regioni dalla limitata estensione temporale e sezionale, come il nostro, che copre quattro anni (1996-1999) con riferimento alle variazioni. Il caso del modello dinamico con effetti fissi o casuali è stato ampiamente studiato e discusso nella letteratura sui dati panel. In particolare, i minimi quadrati ordinari o generalizzati forniscono stime distorte e non consistenti (al crescere della dimensione sezionale del panel) dei parametri  $\phi$  e  $\beta$  (si veda Baltagi, 1995, ed i riferimenti contenuti), che hanno un'importanza cruciale il conseguimento della stima preliminare, ed occorre far ricorso a tecniche che marginalizzano gli effetti regionali (metodo generalizzato dei momenti, si veda Arellano e Bond (1981), o di massima verosimiglianza (Hsiao, Pesaran e Tahmiscioglu, 2001). Entrambe sono essenzialmente fondate sulle differenze prime dei tassi di variazione, che, appunto, eliminano gli effetti regionali costanti.

L'approccio che abbiamo seguito è basato sulla stima di massima verosimiglianza del modello (3), utilizzando la rappresentazione nello spazio degli stati e utilizzando il filtro di Kalman *analitico* (cfr. Durbin e Koopman, 2001), il quale, attraverso l'opportuna specificazione delle condizioni iniziali, consente la marginalizzazione degli effetti regionali, ed pertanto il conseguimento di stime dei parametri  $\phi$  e  $\beta$  consistenti, non distorte, e più efficienti dell'approccio basato sul metodo generalizzato dei momenti.

Al fine di migliorare l'efficienza con cui i parametri invarianti rispetto alle regioni vengono stimati, in qualche caso sono state utilizzate le informazioni aggregate dei Conti Economici Nazionali, disponibili a partire dall'anno 1970, congiuntamente a quelle della contabilità regionale. Questa operazione risulta del tutto ovvia e naturale nel contesto *state space*, richiedendo l'aggregazione sezionale dell'equazione di misurazione.

Inoltre, la rappresentazione del modello viene estesa con facilità al caso in cui gli effetti regionali siano stocastici e variabili nel tempo per la presenza di una componente nominale. Questo risulterà apprezzabile nell'analisi degli aggregati a prezzi correnti.

Per quanto concerne la specificazione della matrice di covarianza di  $\epsilon_t$ , di importanza cruciale per il bilanciamento (si osservi che questa entra nella (5)), abbiamo impiegato il modello fattoriale:

$$\Sigma_\epsilon = \sigma_N^2 ii' + \sigma_*^2 W\gamma, \quad (6)$$

dove  $W$  è la matrice diagonale dei pesi regionali riferiti a  $y_0$  (anno 1995) e  $\gamma$  rappresenta il parametro di eteroschedasticità regionale. La (6) è fondata sulla scomposizione dei disturbi  $\epsilon_t$  in una

componente nazionale, comune a tutte le regioni, e che tende ad indurre una correlazione positiva tra di esse, ed una componente regionale specifica variabile in relazione al peso della regione, elevato ad una potenza  $\gamma$ :  $\epsilon_t = \epsilon_t^{(N)}i + \epsilon_t^*$ , dove  $\epsilon_t^{(N)}$  è un disturbo nazionale. Questa specificazione annida le condizioni sotto le quali si ottengono stime preliminari a coefficiente di variazione costante, che sono soddisfatte per  $\gamma = 0$ , e quelle per le quali si consegue l'aggiustamento proporzionale, che emerge per  $\gamma = -1$ . Il rapporto delle varianze  $\sigma_N^2/\sigma_*^2$  misura l'importanza relativa dei disturbi nazionali.

Infine, il set dei regressori  $X_t$  include, accanto alle ULA regionali ed altre covariate direttamente riferibili all'aggregato oggetto di studio, due fattori sintetici rappresentativi della crescita regionale, discussi nella sezione successiva.

## 4 Estrazione dei fattori rappresentativi della crescita regionale

Recentemente, è emerso un rinnovato interesse circa l'impiego dell'analisi fattoriale dinamica a fini previsivi (cfr. Stock e Watson, 2002). I modelli dinamici utilizzati per la stima rapida degli aggregati dei conti regionali utilizzano, come variabile esplicativa due fattori rappresentativi della crescita regionale, costruiti, con le modalità che ora descriveremo, a partire da un vasto insieme di indicatori regionali coincidenti.

Al tempo  $t$  si dispone di  $K$  indicatori regionali, espressione del livello di attività economica; sia  $x_{rt}$  il vettore  $K \times 1$  che li raccoglie con riferimento alla regione generica, contrassegnata dal pedice  $r$ . E' possibile definire fattori latenti rappresentativi della crescita regionale mediante il seguente modello fattoriale dinamico:

$$\Delta x_{rt} = \Lambda f_{rt} + e_{rt},$$

dove  $f_{rt}$  costituisce un vettore  $M \times 1$ , con  $M$  strettamente minore di  $K$ , che contiene il valore di  $M$  fattori latenti al tempo  $t$  per la regione  $r$ ,  $\Lambda$  è la matrice  $K \times M$  dei pesi (*loading*) fattoriali, invariante rispetto al tempo ed alle regioni, e  $e_{rt}$  rappresenta la componente specifica.

Stock e Watson (2002) mostrano che se i fattori hanno distribuzione stazionaria e la componente specifica  $e_{rt}$  è caratterizzata da correlazione seriale e sezionale "debole",  $f_{rt}$  viene stimato in maniera consistente al crescere del numero  $K$  di indicatori, attraverso l'analisi delle componenti principali. Il risultato è intuitivo, dal momento che il numero dei fattori rimane fisso al crescere di  $K$  e gli indicatori rappresentano misure ripetute di  $f_{rt}$ . La stima dei parametri incidentali in  $\Lambda$  si effettua mediante la scomposizione spettrale della matrice di covarianza dei  $K$  indicatori.

La selezione degli indicatori è basata sulla loro rilevanza e alla possibilità di disporre del dato relativo al 2001 in tempo utile per effettuare la stima rapida, con la disaggregazione regionale richiesta. Tutti gli indicatori di seguito elencati sono disponibili per il periodo 1995-2001. *Intermediazione finanziaria*: 1. depositi bancari per localizzazione della clientela; 2. impieghi bancari per localizzazione della clientela; 3. sportelli bancari; 4. differenziale tassi di interesse a breve termine sui finanziamenti per cassa. *Rilevazione Trimestrale delle Forze di Lavoro*: 1. tasso di attività maschile. 2. tasso di disoccupazione maschile. 3. percentuale di occupati non dipendenti

del settore terziario escluso il commercio. 1. importazioni; 2. esportazioni delle regioni italiane. *Indicatori congiunturali ISAE* 1. saldo livello ordini in generale; 2. saldo livello ordini da interno; 3. saldo livello ordini da estero; 4. grado utilizzo impianti; 5. saldo livello produzione; 6. saldo giacenza prodotti finiti; 7. clima di fiducia delle famiglie; *Immatricolazioni di autovetture nuove* (ACI e Min. Trasporti e Nav.) *Imprese Attive* (Movimprese).

L'estrazione dei fattori si basa sulla scomposizione spettrale della matrice di covarianza delle variazioni relative (assolute per gli indicatori ISAE) rispetto all'anno precedente degli indicatori descritti nella sezione precedente. Per ogni indicatore sono disponibili sei osservazioni annuali (1996-2001) e venti osservazioni regionali. Tutte le variazioni sono state standardizzate sottraendo la media globale (delle 120 osservazioni) e dividendo per la deviazione standard. L'assunzione è che i fattori siano invarianti rispetto alle regioni e al tempo, ed appare irrinunciabile, data la natura dell'informazione disponibile.

L'analisi ha portato alla selezione di due fattori. La percentuale della variabilità complessiva da essi spiegata è del 41%, con il primo fattore che fornisce un contributo pari al 30%. L'ispezione dei *loadings* rivela che il primo fattore sintetizza prevalentemente le informazioni connesse alle variazioni degli indicatori congiunturali dell'ISAE e del commercio con l'estero. Il secondo è definito prevalentemente dalle variazioni del tasso di attività, di occupazione nel terziario, degli sportelli bancari e del clima di fiducia delle famiglie. Il coefficiente di correlazione dei fattori con il tasso di variazione del prodotto lordo ai prezzi di mercato a prezzi costanti risulta pari rispettivamente a 0.19 e 0.16.

Nel passaggio dal 2000 al 2001 la persistenza nel posizionamento delle regioni secondo i due fattori è bassissima. Inoltre, per quanto concerne la dinamica temporale e territoriale, si osserva che  $f_{1t}$  ha una componente comune molto rilevante (ovvero bassa dispersione territoriale), ed una tendenza temporale piuttosto marcata. Il secondo fattore risulta invece caratterizzato rispettivamente da minore dispersione temporale e maggiore dispersione territoriale.

## 5 Il valore aggiunto regionale

### 5.1 La stima rapida del VA a prezzi costanti

La stima del valore aggiunto ai prezzi di base del 1995 per i settori extra-agricoli, eccettuato quello delle costruzioni, risulta dalla combinazione di due stime: la prima viene effettuata mediante una variante modellistica del metodo estrapolativo della produttività del lavoro, mentre la seconda emerge dal modello di regressione dinamico con effetti regionali fissi discusso nella sez. 3. Entrambe le stime poggiano sul medesimo set informativo, basato sulle ULA ed i fattori rappresentativi della crescita regionale, e si differenziano per la specificazione del modello.

La determinazione ad effettuare la combinazione, piuttosto che procedere alla selezione del modello migliore, origina dall'incertezza che caratterizza la modellazione degli effetti dinamici e degli effetti regionali, in presenza di un panel di estensione sezionale e temporale ridotta.

Il metodo estrapolativo della produttività del lavoro si fonda sullo stimatore rapporto, considerato nella sezione 2,  $y_t = p_t \odot l_t$ . Risultando disponibile l'informazione relativa alle ULA,

$l_{T+l}$ , la stima preliminare  $l = 1, 2$  periodi in avanti può essere conseguita attraverso la semplice estrapolazione  $\hat{y}_{T+l} = \hat{p}_{T+l} \odot l_{T+l}$ . Il problema è pertanto estrapolare la produttività del lavoro. L'evidenza preliminare, maturata attraverso un esperimento ricorsivo di previsione un periodo in avanti entro il periodo campionario (1996-1999) che confronta le previsioni con il valore effettivo, ha mostrato che la semplice regola di utilizzare l'ultimo valore osservato,  $\hat{p}_{T+l} = p_T$ , presenta una buona *performance*. Essa risulta consistente con la rappresentazione:

$$\Delta \ln y_t = \Delta \ln p_t + \Delta \ln l_t; \quad (7)$$

dove il logaritmo della produttività del lavoro ammetta una rappresentazione *random walk*:

$$\ln p_t = \ln p_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \Sigma_\epsilon); \quad (8)$$

Tuttavia, il modello (7)-(8) risulta mal specificato se la produttività presenta una componente di *drift*; un modello più generale può pertanto essere specificato contemplando la presenza di *drift*, possibilmente regionali, e la presenza di inerzia sotto forma di una componente autoregressiva a parametro comune, scrivendo:

$$\Delta \ln p_t = m + \phi \Delta \ln p_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \Sigma_\epsilon). \quad (9)$$

dove  $m$  è un vettore di termini costanti. Sostituendo nella (7), si ottiene un modello autoregressivo dinamico a ritardi distribuiti che risulta annidato in

$$\Delta \ln y_t = \phi \Delta \ln y_{t-1} + \mu + \beta_0 \Delta l_t + \beta_1 \Delta \ln l_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \Sigma_\epsilon), \quad (10)$$

rispetto al quale il modello alla base del metodo estrapolativo della produttività del lavoro opera le restrizioni:  $\phi = 0, \beta_0 = 1; \beta_1 = -\beta_0 \phi$ . La seconda è nota come restrizione a fattore comune; inoltre,  $\mu = (1 - \phi)^{-1} m$ .

Come sottolineato nella sezione 3,  $\mu$  gioca un ruolo essenziale; la sua mancata inclusione genera previsioni sistematicamente distorte, anche se esiste una circostanza particolare in cui la distorsione viene corretta in sede di bilanciamento proporzionale, che emerge quando  $\mu$  è invariante rispetto alle regioni, per cui il *drift* diventa *ci*, per  $c$  scalare. L'aggiustamento proporzionale presenta questa proprietà di robustezza. Si osservi che l'assenza di effetti regionali implica l'assenza di processi di convergenza o divergenza tra le regioni italiane nel periodo considerato.

Inoltre, la stima anticipata mediante il modello *random walk* risulta inconsistente con alcune evidenze empiriche, sufficientemente robuste, documentate in Proietti (2002). Essa implica che la variazione attesa del valore aggiunto coincide con quella delle unità di lavoro, mentre l'esperienza maturata sulle serie dei conti regionali e nazionali attesta abbastanza esplicitamente la natura prociclica della produttività del lavoro: questa è l'espressione della più debole correlazione tra le variazioni dell'output e dell'input del fattore lavoro. La prociclicità tende ad essere più debole per il settore dei servizi. Nel breve periodo l'aggiustamento alla situazione ciclica viene effettuato variando maggiormente la produttività, vale a dire l'intensità dell'utilizzo del fattore lavoro, piuttosto che il livello di impiego del medesimo, caratterizzato da maggiori rigidità.

Un'ulteriore evidenza empirica, piuttosto netta, è che  $\text{Var}(\Delta \ln l_t) < \text{Var}(\Delta \ln y_t)$  e questo implica anche  $\text{Var}(\Delta \ln l_t) < \text{Var}(\Delta \ln p_t)$ ; pertanto, la variazione dell'input di lavoro è caratterizzata da maggior liscio rispetto a quella del margine intensivo.

In definitiva, il metodo estrapolativo della produttività possiede una buona *performance* previsiva entro il periodo campionario disponibile, ma soffre di alcune limitazioni: in particolare, le stime anticipate violano la prociclicità della produttività del lavoro, dal momento che la variazione attesa del valore aggiunto risulta pari a quella delle ULA. Inoltre, l'informazione sulle unità di lavoro, caratterizzata da maggiore persistenza (e pertanto più facilmente anticipabile) ha contenuto limitato sulle variazioni dell'output; per meglio cogliere l'eterogeneità del valore aggiunto regionale occorre introdurre i drift regionali e ulteriori variabili esplicative che ne riassumano gli aspetti essenziali.

**Il metodo basato sull'estrapolazione della produttività del lavoro** È stata considerata un'estensione del metodo in questione che fornisce una rappresentazione parsimoniosa del fenomeno della eteroschedasticità regionale, consentendo di ottenere la matrice di covarianza delle stime preliminari da utilizzare per il bilanciamento, e dall'altro utilizza i fattori regionali di crescita come variabile esplicativa delle variazioni della produttività, nella speranza che questi consentano di cogliere l'eterogeneità delle variazioni regionali:

$$\Delta \ln p_t = \mu i + \beta_1 f_{1t} + \beta_2 f_{2t} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \Sigma_\epsilon), \Sigma_\epsilon = \sigma_N^2 i i' + \sigma_*^2 W^\gamma \quad (11)$$

per  $t$  che va dal 1996 al 1999. Come si nota, il *drift*,  $\mu$ , risulta invariante rispetto alle regioni.

**Modello autoregressivo dinamico per dati panel** Il secondo modello, che genera un set di stime alternative da combinare con quelle risultanti dal primo metodo, è un modello di regressione dinamico per dati panel atto a cogliere la presenza di drift regionali, di una componente inerziale nelle variazioni relative del VA settoriale a prezzi costanti. La presenza dei due fenomeni ha effetti di rilievo sulla natura e qualità delle previsioni e purtroppo non risulta facilmente verificabile sulla base del set informativo a nostra disposizione. Il risultato atteso è che le stime saranno caratterizzate da maggiore variabilità, ma minore distorsione; quale delle due componenti dell'errore quadratico medio di stima prevalga non è determinabile a priori, dal momento che soltanto per il 1999 avrebbe senso misurare la performance previsiva del modello.

$$\Delta \ln y_t = \phi \Delta \ln y_{t-1} + \mu + \beta_0 \Delta l_t + \beta_1 \Delta \ln l_{t-1} + \beta_f f_{1t} + \beta_2 f_{2t} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \Sigma_\epsilon). \quad (12)$$

**La combinazione delle stime** Al fine di combinare le stime del valore aggiunto regionale a prezzi costanti che emergono dai modelli trattati nei due paragrafi precedenti, impieghiamo un risultato dovuto a Peña (1997), esplicitabile nella maniera seguente: date due stime indipendenti non distorte,  $\hat{y}_{1t}$  e  $\hat{y}_{2t}$  con matrici di covarianza  $V_1$  e  $V_2$ , non necessariamente non singolari, la stima a varianza minima è formata nella maniera seguente:

$$\hat{y}_t = (V_1^- + V_2^-)^{-1} (V_1^- \hat{y}_{1t} + V_2^- \hat{y}_{2t}),$$

dove  $^-$  denota l'inversa generalizzata; inoltre, la matrice di covarianza di  $\hat{y}_t$  è fornita da  $(V_1^- + V_2^-)^{-1}$ . Non è possibile applicare tale risultato alle stime bilanciate, dal momento che lo spazio nullo delle matrici  $V_1$  e  $V_2$  è coincidente, essendo descritto dal vettore di aggregazione  $i$ , con la conseguenza che  $V_1^- + V_2^-$  non è invertibile in senso stretto. Esso vale dunque solo con riferimento alle stime preliminari ed è stato sfruttato per generare previsioni qualitativamente migliori, in quanto conseguono un errore standard di stima più piccolo.

La stima combinata viene successivamente bilanciata al totale nazionale,  $Y_t$ , nel modo consueto:

$$\tilde{y}_t = \hat{y}_t + (V_1^- + V_2^-)^{-1} i \frac{Y_t - i' \hat{y}_t}{i' (V_1^- + V_2^-)^{-1} i}.$$

## 5.2 Costruzioni

Per la branca delle costruzioni le ULA non giocano alcun ruolo particolare nella stima del valore aggiunto regionale. In effetti, la relazione statistica con le ULA si presenta molto tenue e per la previsione preliminare si è fatto ricorso ad un indicatore del valore dell'edificato (denotato con  $E_t$ ) messo a disposizione dall'Istat. Essendo la variabile esplicativa espressa a prezzi correnti, si è partiti dalla previsione del valore aggiunto a prezzi correnti e si è pervenuti a quella a prezzi costanti mediante deflazione. Il modello stimato è così formulato:  $\Delta \ln y_t^* = \mu + \beta \Delta \ln E_t + \epsilon_t$ ,  $\epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2 I)$ . Si osserva che il parametro di eteroschedasticità non risulta significativo, mentre il termine di *drift* risulta variabile sul territorio.

## 5.3 Le stime a prezzi correnti - Deflazione

Il passaggio dal modello a prezzi costanti a quello a prezzi correnti, e viceversa, viene effettuato in base a considerazioni statistico-economiche. Denotando con  $y_t^*$  il valore aggiunto a prezzi correnti e con  $d_t$  l'indice di prezzo implicito,  $d_t = y_t^* \div y_t$ , segue immediatamente la seguente scomposizione additiva delle variazioni logaritmiche a prezzi correnti:

$$\Delta \ln y_t^* = \Delta \ln y_t + \Delta \ln d_t. \quad (13)$$

Conseguentemente, una volta disponibili le stime del valore aggiunto a prezzi costanti, occorre effettuare soltanto la previsione della componente nominale per i medesimi anni. Ad esempio, con riferimento all'anno  $T + 1 = 2000$ , si ottiene la previsione del logaritmo del valore aggiunto regionale a prezzi correnti come segue:  $\widehat{\ln y_{T+1}^*} = \ln y_T^* + \Delta \widehat{\ln y_{T+1}} + \Delta \widehat{\ln d_{T+1}}$ . Il passaggio ai livelli viene effettuato utilizzando le proprietà della distribuzione lognormale, i cui momenti sono calcolabili a partire da quelli di  $\ln y_{T+1}$ , condizionatamente al set informativo disponibile. La matrice di covarianza delle stime preliminari così ottenuta viene dunque utilizzata al fine di effettuare il bilanciamento al totale nazionale. Viceversa, se sono disponibili le stime a prezzi correnti, come avviene per il settore COS, si procede a sottrarre dalle variazioni a prezzi correnti la variazione del deflatore.

L'inflazione regionale viene modellata con un processo *seemingly unrelated* IMA(1,1) con due fonti di variazione: una comune, dovuta all'inflazione nazionale, ed una specifica:

$$\Delta \ln d_t = \Delta \ln d_{t-1} + \eta_t + \theta \eta_{t-1}, \quad \eta_t \sim \text{NID}(0, \Sigma_\eta) \quad (14)$$

con

$$\Sigma_{\eta} = \sigma_{\eta} i i' + \sigma_{\eta}^2 W^{\delta},$$

dove  $W$  è la matrice diagonale dei pesi del valore aggiunto e  $\delta$  è un parametro di eteroschedasticità.

Per quanto concerne la matrice di covarianza dei disturbi, la stima del parametro di eteroschedasticità regionale sottintende la presenza di una variabilità più accentuata nelle regioni di piccole dimensioni, ma in misura piuttosto moderata. La componente di varianza nazionale è predominante in tutte le serie ad eccezione dell'agricoltura, dei SIFIM e delle imposte sulla produzione e le importazioni.

I differenziali di inflazione regionale non risultano mai significativi. Se questi risultati vengono abbinati al fatto che una quota molto grande della variabilità sia assorbita dalla componente nazionale, è consentito affermare che le dinamiche regionali sono in larga parte comuni.

## 5.4 SIFIM

La stima rapida dei servizi di intermediazione finanziaria indirettamente misurati ricalca il processo di stima utilizzato dall'Istat. Il punto di partenza è rappresentato dalla stima analitica della componente relativa agli impieghi ed ai depositi delle banche, a prezzi correnti, effettuato dal Dipartimento di Contabilità Nazionale e Analisi Economica, sulla base delle informazioni fornite dalla Banca d'Italia concernenti i tassi di interesse attivi e passivi e i depositi e gli impieghi, opportunamente rettificati; il residuo è stato ripartito proporzionalmente al valore aggiunto del settore IMF-AI-NAPI.

Malgrado la componente stimata analiticamente rappresenti circa i due terzi del totale, la stima rapida dei SIFIM rappresenta uno dei punti più controversi del processo di stima indiretta, per l'eccezionale variabilità della componente suddetta. Parimenti problematica è la deflazione delle stime a prezzi correnti; questa viene effettuata utilizzando le stime dei deflatori dei SIFIM regionali desunti dalla sezione 5.3, i quali sono tuttavia caratterizzati da una estrema variabilità.

## 5.5 IVA, imposte nette sui prodotti e sulle importazioni

La stima rapida di questo aggregato a prezzi correnti (nel seguito denotato con  $IVA_t$ ) ha fatto affidamento sul modello di regressione:

$$\Delta IVA_t = \rho_y \Delta VA_t + \rho_m \Delta IMP_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma^2 W^{\gamma})$$

dove  $W$  è la matrice diagonale che contiene i pesi riferiti al 1995. Il valore delle importazioni regionali è stato desunto dalla banca dati del commercio con l'estero CoeWeb (cfr. [www.coeweb.istat.it](http://www.coeweb.istat.it)). Tale rappresentazione assume che l'aggregato sia scomposto in due componenti additive: la prima commisurata al valore aggiunto,  $VA_t$ , e la seconda proporzionale al valore delle importazioni,  $IMP_t$ ; si assume che i coefficienti di proporzionalità siano invarianti nel tempo e questo produce un modello lineare nelle differenze dei livelli che non include l'intercetta.

Per la stima rapida dei flussi regionali del 2000 e del 2001 sono state utilizzate le stime bilanciate del valore aggiunto ai prezzi di base al netto SIFIM. Malgrado il buon adattamento del modello,

la somma dei flussi regionali sovrastima il totale regionale del 10% e del 4% il totale desunto dai conti nazionali rispettivamente per il 2000 ed il 2001, rivelando che la regionalizzazione delle imposte sui prodotti e le importazioni si presenta piuttosto problematica.

Lo stesso modello è stato impiegato per la previsione preliminare delle imposte nette a prezzi costanti, esprimendo a prezzi costanti le importazioni impiegando il deflatore delle importazioni desunto dai conti economici nazionali. Questo sotto l'assunzione che gli scambi delle regioni con l'estero avvenga al medesimo sistema dei prezzi. Questa soluzione si mostra di gran lunga superiore rispetto alla deflazione delle stime effettuati a prezzi correnti, data l'elevatissima variabilità dei deflatori. A dispetto del peggiore adattamento rispetto al modello a prezzi correnti, la discrepanza percentuale delle stime regionali rispetto al totale nazionale risulta contenuta, non superando il 2% in entrambi gli anni.

## **5.6 Prodotto interno lordo ai prezzi di mercato**

Le stime del valore aggiunto settoriale ai prezzi di base, dei SIFIM e dell'IVA, imposte indirette nette sui prodotti e imposte sulle importazioni, vengono aggregate per fornire la stima del PIL regionale e ripartizionale ai prezzi di mercato. La varianza dell'errore di stima viene anche essa ottenuta mediante la somma delle varianze delle singole componenti, dal momento che la stima di ciascuna è effettuata indipendentemente. L'eccezione è rappresentata dai SIFIM e dalle imposte indirette, che ricevono come input le stime del valore aggiunto ai prezzi di base; le stime corrispondenti saranno pertanto correlate con quelle del VA ai prezzi di base.

## **6 L'impiego dell'indagine rapida per la stima del valore aggiunto ripartizionale**

L'indagine rapida (IR) sui risultati economici delle imprese medio-grandi fa riferimento a circa 4000 imprese con più di 100 addetti ed è intesa ad avere una significatività a livello di ripartizione; l'informazione viene regionalizzata mediante l'accoppiamento delle unità rispondenti con l'indagine sul sistema dei conti delle imprese. I dati in nostro possesso si basano sulle risposte ricevute entro il termine utile per il loro impiego e riguardano 7000 unità funzionali, per le quali si dispone degli addetti, dei dipendenti, delle spese per il personale, degli investimenti, del valore aggiunto originario (al netto di imposte e contributi) e del valore aggiunto ai prezzi di mercato.

Le informazioni sono disponibili per gli anni 1998-2000, disaggregate per regione e per 25 branche della classificazione Nace-Rev.1. Per le imprese che appartengono ad una branca regionale, si dispone del dato relativo alle stessa unità riferito all'anno precedente. L'Istat ha messo a disposizione il valore aggiunto operativo e il numero di unità di lavoro lorde con la stessa disaggregazione della classe dimensionale corrispondente desunto dai conti regionali per gli anni 1997-1999 e 2000, limitatamente alle ULA.

L'IR potrebbe essere utilizzata al fine di cogliere ed estrapolare le tendenze della produttività a livello ripartizionale e settoriale che caratterizzano le unità di attività economica di grandi dimensioni. A tal fine, in Proietti (2002) abbiamo proposto e validato in base alla capacità predittiva,

diversi stimatori basati sulla logica del campionamento post-stratificato, ipotizzando che le imprese rispondenti in ciascuna cella di classificazione per branca e regione costituiscano un campione casuale tratto da una popolazione le cui caratteristiche sono note dai conti economici regionali disaggregati per classe dimensionale. Dal momento che la relazione tra gli aggregati IR e dei conti regionali risulta più stabile con riferimento alla produttività del lavoro, e che le variazioni delle unità di lavoro della popolazione sono disponibili, la preferenza si è orientata verso gli stimatori “rapporto” (metodo estrapolativo della produttività del lavoro).

Va rilevato, tuttavia, che l’impiego dell’Indagine Rapida ai fini della stima rapida del valore aggiunto regionale a prezzi correnti presenta aspetti problematici che di seguito vengono qui riassunti: in primo luogo, essa consente la stima dei tassi di variazione soltanto a livello ripartizionale e per le cinque macrobranche extra-agricole, con riferimento alle quali, tuttavia, la copertura è ben lungi dall’essere completa; in particolare, per il settore AAS non rientrano nella rilevazione, per ovvi motivi, le unità delle branche L e P, mentre per il settore IMF-AI-NAPI, risulta esclusa l’importante componente dell’Intermediazione monetaria e finanziaria (branca J). In secondo luogo, le osservazioni sono caratterizzate dal fenomeno dell’iperdispersione, per affrontare il quale abbiamo proposto un criterio operativo per censurare le variazioni più estreme, che comunque presenta rilevanti margini di arbitrio. Risulta inoltre difficile asserire l’affidabilità delle stime: gli errori standard, calcolati sotto l’ipotesi che l’insieme delle imprese che rientrano nell’indagine rapida configuri un campione stratificato dell’universo delle unità con 100 addetti e oltre, sono manifestamente distorti verso il basso, come è facile verificare se si confronta l’errore effettivo di stima per gli anni 1998 e 1999 con il margine atteso sotto quell’ipotesi, e quindi non rappresentano una chiara guida circa l’affidabilità delle stime. Infine, ai fini della stima rapida del valore aggiunto regionale complessivo occorre effettuare la stima per le classi dimensionali inferiori, che andrebbe operata autonomamente utilizzando i metodi indiretti esposti nei capitoli successivi, dal momento che la relazione statistica tra le variazioni di produttività delle grandi e delle piccole è purtroppo molto debole, e pertanto il contenuto informativo della rapida nei confronti della disaggregazione territoriale del VA delle unità piccole risulta decisamente limitato.

## 7 Redditi da lavoro dipendente

La stima rapida dei redditi da lavoro dipendente è stata effettuata sulla base di un modello che tiene conto del processo statistico di generazione dei dati utilizzato dall’Istat: con modalità del tutto analoghe alla stima del valore aggiunto, i redditi da lavoro dipendente fanno largo ricorso alle unità di lavoro dipendenti, dal momento che alla loro determinazione si perviene sommando le stime ottenute separatamente per  $m = 5$  classi dimensionali di unità di attività economica locale, moltiplicando i redditi per ULA dipendente, desunti dalle indagini di base sui conti delle imprese, per le ULA dipendenti della medesima classe.

Con riferimento all’ $i$ -esima regione, il procedimento può essere così formalizzato:

$$y_i = \sum_{j=1}^m r_{ij} l_{ij} = \bar{r}_i \cdot l, \quad \bar{r}_i = \sum_{j=1}^m r_{ij} \frac{l_{ij}}{l_i}$$

dove  $r_{ij}$  rappresenta i redditi per unità di lavoro dipendente della classe dimensionale  $j$  della regione  $i$  ( $m = 5$ ) e  $l_{ij}$  il numero di ULA dipendenti della classe, e  $l = \sum_j l_j$ .

Introducendo il riferimento temporale e raccogliendo  $\bar{r}_i, i = 1, \dots, R$ , nel vettore  $\bar{r}_t$  possiamo scrivere in notazione vettoriale:

$$y_t = \bar{r}_t \odot l_t. \quad (15)$$

Considerando le differenze prime logaritmiche di ambo i membri della (15) e denotando  $\mu_t = \Delta \ln \bar{r}_t$ , si ottiene:

$$\Delta \ln y_t = \mu_t + \Delta \ln l_t$$

$\mu_t$  rappresenta dunque la variazione relativa delle retribuzioni per ULA dipendente e costituisce una componente latente per il set informativo a nostra disposizione. Occorre pertanto formulare un modello dinamico che descriva il suo comportamento temporale.

In un primo momento la nostra attenzione si è incentrata nel modello *random walk* multivariato a componenti di varianza (fattore nazionale e fattori specifici regionali). Tuttavia, la statistica di Ljung-Box evidenziava per gran parte dei settori la presenza di autocorrelazione significativa; la responsabilità andava ascritta all'autocorrelazione negativa a ritardo 1. Questa evidenza empirica ha suggerito la specificazione di modello IMA(1,1) per  $\mu_t$ , con parametro media mobile scalare, qui denotato con  $\theta$ :

$$\begin{aligned} \Delta \ln y_t &= \mu_t + \Delta \ln l_t + \beta_{f1} f_{1t} + \beta_{f2} f_{2t}, \\ \mu_t &= \mu_{t-1} + \eta_t + \theta \eta_{t-1}, \quad \eta_t \sim \text{NID}(0, \sigma_N^2 ii' + \sigma_*^2 W^\gamma). \end{aligned} \quad (16)$$

Tale formulazione implica che  $\ln y_t \sim \text{IMA}(2, 1)$ , vale a dire ha una rappresentazione stazionaria nelle differenze seconde con una componente media mobile. La matrice  $W$  contiene sulla diagonale principale la quota relativa dei redditi da lavoro dipendente di ciascuna regione. Infine, la (16) annovera tra le variabili esplicative i fattori rappresentativi della crescita regionale,  $f_{kt}, k = 1, 2$ .

Il parametro media mobile ha il ruolo di diminuire la persistenza dei disturbi, che ammonta a  $(1 + \theta)$ ; inoltre, risultando negativo e significativamente diverso da zero, le previsioni del tasso di variazione dei redditi medi per ULA dipendente sono effettuate mediante livellamento esponenziale:

$$\tilde{\mu}_{t+1|t} = \lambda \tilde{\mu}_{t|t-1} + (1 - \lambda)(\Delta \ln y_t - \Delta \ln l_t), \quad (17)$$

dove  $\lambda = -\theta$ ,  $\lambda \in (0, 1)$ , è la costante di livellamento. Pertanto, la previsione un periodo in avanti effettuata al tempo  $t$  è una media aritmetica ponderata della previsione del valore corrente fatta al tempo  $t - 1$  e dell'osservazione corrente. Manipolando la (17), si ha anche che la previsione del tempo  $t - 1$  viene corretta per un frazione pari a  $(1 - \lambda) = 1 + \theta$  dell'errore di previsione. Tale frazione, detta appunto persistenza, perché misura la quota parte dell'errore di previsione che si propaga alla previsione del tasso di variazione dei redditi, è inferiore all'unità per tutte le branche.

Il contributo della componente nazionale alla varianza complessiva del termine di disturbo è solitamente molto elevato. In generale, si può affermare che le dinamiche regionali dei redditi da lavoro dipendente sono in larga parte comuni. Il parametro di eteroschedasticità regionale risulta negativo e prossimo a uno; di conseguenza, il bilanciamento al totale nazionale avverrà in misura quasi proporzionale, mentre il coefficiente di variazione delle previsioni varia inversamente con il peso della regione.

Complessivamente l'adattamento del modello è soddisfacente e le discrepanze percentuali rispetto al totale nazionale si presentano contenute.

## 8 I consumi finali interni

**Spese delle famiglie a prezzi correnti** La stima anticipata delle spese delle famiglie a prezzi correnti è basata su un modello di regressione statico che spiega le variazioni logaritmiche delle spese relative ai 12 capitoli ( $\Delta \ln c_t^*$ ) in funzione delle variazioni registrate dall'indagine sui bilanci delle famiglie italiane ( $\Delta \ln bf_t$ ), dall'indice ripartizionale delle vendite al dettaglio ( $\Delta \ln v_t$ ), dai fattori rappresentativi della crescita regionale ( $f_{1t}$  e  $f_{2t}$ ), dalla variazione delle immatricolazioni di autoveicoli nuovi ( $\Delta \ln Imm_t$ ) e del clima di fiducia delle famiglie ( $\Delta \ln GF_t$ ). La natura statica del modello è imposta dalla limitata estensione delle serie  $bf_t$ , disponibili dal 1997.

$$\begin{aligned} \Delta \ln c_t^* &= \mu i + \beta_{BF} \Delta \ln bf_t + \beta_v \Delta \ln v_t + \beta_1 f_{1t} + \beta_2 f_{2t} + \beta_I \Delta \ln Imm_t + \beta_{GF} \Delta \ln GF_t + \epsilon_t, \\ \epsilon_t &\sim \text{NID}(0, \sigma_N^2 ii' + \sigma_*^2 W^\gamma) \end{aligned} \quad (18)$$

Per la stima del modello sono disponibili 40 osservazioni. La bontà dell'adattamento del modello (18) è del tutto apprezzabile per 7 capitoli di spesa, come rivela il coefficiente di determinazione che risulta comunque non inferiore al 30%.

Per i capitoli di spesa considerati, la stima a prezzi costanti si ottiene mediante deflazione, utilizzando gli indici dei prezzi dei comuni capoluoghi di regione.

**Spese delle famiglie a prezzi costanti** I consumi a prezzi costanti sono stati modellati utilizzando un processo ARIMA(1,1,0) con parametro autoregressivo e *drift* invariante rispetto alle regioni, e con un nucleo di regressione che presenta tra le variabili esplicative i fattori rappresentativi della crescita regionale, il tasso di variazione degli autoveicoli immatricolati, limitatamente al capitolo CP07, e del grado di fiducia delle famiglie. Il modello di riferimento viene rappresentato dall'equazione seguente:

$$\Delta \ln c_t = \mu i + \phi \Delta \ln c_{t-1} + \beta_1 f_{1t} + \beta_2 f_{2t} + \beta_I \Delta \ln Imm_t + \beta_{GF} \Delta \ln GF_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma_N^2 ii' + \sigma_*^2 W^\gamma). \quad (19)$$

Le corrispondenti stime a prezzi correnti sono ottenute moltiplicando la stima a prezzi costanti per l'indice dei prezzi al consumo per le famiglie di operai e impiegati in base 1995 dei comuni capoluogo di regione.

**La combinazione delle previsioni** Le stime delle spese delle famiglie ottenute nel paragrafo precedente possono essere espresse a prezzi correnti, moltiplicandole per gli indici regionali dei prezzi al consumo relativi ai capitoli corrispondenti. Per sette capitoli di spesa si dispone pertanto di due stime, la prima ottenuta dal modello di regressione statico (18) e la seconda dal modello autoregressivo (19); queste possono essere combinate al fine di conseguire una previsione più efficiente. I guadagni in termini di precisione delle stime sono sostanziali: ad esempio, per le stime a prezzi correnti relative alla regione Piemonte la combinazione ed il bilanciamento fanno

congiuntamente registrare una riduzione dell'errore standard di stima pari all'82% con riferimento alla regressione statica. Il contributo marginale del bilanciamento è pari all'8%.

**La spesa delle AA.PP e ISP** La spesa delle AA.PP a prezzi costanti viene stimata facendo ricorso ad indicatori di input e di output dei servizi pubblici. Il modello di regressione su cui è fondata la stima anticipata ricalca questa logica, ed è formulato nel modo seguente:

$$\Delta \ln c_t = \mu_i + \beta_P \Delta \ln \text{pop}_t + \beta_L \Delta \ln l_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma_N^2 ii' + \sigma_*^2 W^\gamma) \quad (20)$$

dove  $\text{pop}_t$  denota la popolazione regionale e  $l_t$  rappresenta le ULA delle branche delle altre attività di servizi. L'adattamento del modello è soddisfacente, il coefficiente di determinazione attestandosi su valori consueti per regressioni sezionali come la presente (0.21). Le previsioni a prezzi costanti vengono espresse a prezzi correnti attraverso la semplice moltiplicazione per l'indice nazionale dei prezzi implicito nei consumi delle AA.PP. desunto dai conti economici nazionali. Con riferimento alle ISP il modello 20 non conduce a stime significative e la previsione è di tipo *random walk* a componenti di varianza, essendo fondata sulla constatazione che, sulla base del set informativo disponibile, le differenze logaritmiche delle spese per consumi delle ISP sono IID con drift invariante rispetto alle regioni. La stima anticipata viene espressa a prezzi correnti moltiplicandola per l'indice dei prezzi implicito desunto dalla contabilità nazionale.

**Consumi finali interni** Le stime delle spese per consumi finali delle famiglie, delle amministrazioni pubbliche e delle istituzioni senza scopo di lucro al servizio delle famiglie vengono aggregate per fornire la stima dei consumi finali interni, in milioni di euro correnti e a prezzi 1995. La varianza dell'errore di stima viene anche essa ottenuta mediante la somma delle varianze delle singole componenti, dal momento che la stima di ciascuna è effettuata indipendentemente.

## Riferimenti bibliografici

- Arellano, M., e Bond, S. R. (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". *Review of Economic Studies*, 58, 277-297.
- Baltagi, B. H. (1995). *Econometric Analysis of Panel Data*. Chichester: John Wiley and Sons.
- Durbin, J., e Koopman S.J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.
- Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K. (2001), "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods", *Journal of Econometrics* (in corso di pubblicazione).
- Johnson, N. L., e Kotz, S. (1972), *Distributions in statistics: continuous multivariate distributions*, John Wiley and Sons, New York.

- Pascarella C. (1998), "I conti economici territoriali, prospettive per gli anni 2000", in: *Atti della IV Conferenza Nazionale di Statistica*, Roma, 11-13 Novembre 1998;
- Penã D. (1997), "Combining Information in Statistical Modeling", *The American Statistician*, 51, 326-332.
- Proietti, T. (2002), *La stima rapida dei conti economici regionali degli anni 2000-2001*, Relazione finale, Dipartimento di Scienze Statistiche, Università di Udine.
- Stock, J.H., e. Watson, M.W. (2000), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, 20, 2, 147-162.