

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in

Statistica per le Tecnologie e le Scienze



RELAZIONE FINALE

**Benchmarking e disaggregazione temporale con
indicatori di riferimento: un pacchetto per Gretl**

Relatore Prof. Tommaso Di Fonzo
Dipartimento di Scienze Statistiche

Laureando: Alberto Maronilli
Matricola N. 1124569

Anno Accademico 2018/2019

Indice

1. Introduzione	5
1.1. Presentazione del lavoro	7
2. Il software Gretl	8
3. Le tecniche di benchmarking e di disaggregazione temporale	10
3.1. Naive benchmarking	12
3.2. Pro-rata benchmarking	13
3.3. Denton	14
3.3.1. Original Denton	15
3.3.2. Modified Denton	16
3.4. Varianti del metodo di Denton	17
3.4.1. Additive First Differences (AFD)	17
3.4.2. Additive Second Differences (ASD)	18
3.4.3. Proportional First Differences (PFD)	18
3.4.4. Proportional Second Differences (PSD)	19
3.5. Cholette - Dagum	20
3.6. Chow – Lin.	23
4. Il pacchetto <i>Benchmarking</i>	25
4.1. La struttura del pacchetto	25
4.2. Operazioni preliminari	27
4.2.1. Esempio	28
4.3. Utilizzo del pacchetto <i>Benchmarking</i>	31

4.3.1. Tramite GUI	31
4.3.2. Tramite script	33
4.4. Esempio di output	35
Appendice A: Disaggregazione temporale in Gretl: la funzione <code>chowlin()</code>	43
Appendice B: Codice della funzione <i>Benchmarking</i>	45
Riferimenti Bibliografici	51

Capitolo 1

Introduzione

In numerosi settori della ricerca economica applicata sono spesso necessarie informazioni statistiche aventi cadenza inferiore all'anno. I dati annuali infatti mediano andamenti che possono spesso risultare parecchio diversi nell'arco dei dodici mesi e oscurano fluttuazioni di rilevante interesse per l'analisi economica a breve termine.

La contabilità nazionale produce un gran numero di serie economiche che, se a cadenza trimestrale, forniscono una visione più completa dell'andamento dei principali fenomeni economici e offrono una ricca base informativa, utilizzabile con profitto nello studio della congiuntura e per previsioni e politiche di breve periodo.

I conti trimestrali inoltre, riflettendo "il graduale formarsi di quei risultati economici che sono sintetizzati, pur con i dovuti aggiustamenti, nel consuntivo annuo" (Filosa, 1984, p. l.) e basandosi su fonti statistiche usate in larga misura anche per la stima dei dati annuali, rappresentano un test pressoché continuo sulla realizzazione dei bilanci economici annuali, e allo stesso tempo stimolano più frequenti controlli dei criteri di stima e della qualità dei dati.

L'approccio ovvio per la stima di serie economiche trimestrali è quello diretto, consistente quindi in stime delle grandezze di interesse proveniente da rilevazioni appositamente predisposte. Agire in questo modo può però, per motivazioni legate a costi o alla natura del fenomeno, fornire dei valori poco precisi oppure non essere possibile. Per fare fronte a questa problematica si fa quindi ricorso a metodi indiretti, che fondano le proprie valutazioni su basi informative che comprendono una serie storica a cadenza inferiore (e quindi più facilmente ottenibile) della variabile di interesse e, talvolta, altre informazioni utili a ricostruire la dinamica infra-annuale della serie.

Le tecniche di disaggregazione temporale, benchmarking e riconciliazione possono essere utilizzate in questo senso per ottenere consistenza spaziale e temporale per

insiemi di serie storiche. Questi metodi sono ampiamente utilizzati in tutto il sistema statistico europeo e non solo nella produzione di statistiche ufficiali.¹

In questo elaborato ci si concentrerà soprattutto sulle tecniche di *benchmarking*. Queste vengono utilizzate quando si hanno due o più serie storiche a disposizione per una stessa variabile obiettivo con diverse frequenze.

Frequentemente può capitare che le due serie storiche non “vadano d’accordo”: per esempio le somme annuali di osservazioni trimestrali di una variabile potrebbero non essere uguali alla corrispondente serie storica annuale. Il benchmarking temporale serve proprio per risolvere questo problema, stimando delle nuove serie storiche che siano temporalmente consistenti con i dati annuali e riescano allo stesso tempo a preservare i movimenti a breve termine della serie indicatrice trimestrale.

Attualmente sono note diverse tecniche per l’attuazione del benchmarking che sono state sviluppate in diversi software. Uno di questi, ormai diffuso a livello accademico e non, che non dispone ancora di una adeguata copertura delle principali tecniche di benchmarking proposte in letteratura è Gretl che, pur essendo pensato per l’analisi econometrica, propone una scelta molto povera di tecniche di benchmarking e di disaggregazione temporale con indicatori di riferimento.

¹ Per approfondire l’argomento, consultare “European Statistical System (ESS) guidelines on temporal disaggregation, benchmarking and reconciliation” (<https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-06-18-355>)

1.1 Presentazione del lavoro

Per ovviare alle lacune del software Gretl indicate nell'introduzione, è stato sviluppato un pacchetto di funzioni, chiamato "*Benchmarking*". Esso permette di utilizzare alcune delle più diffuse tecniche di benchmarking e di disaggregazione temporale con indicatori di riferimento all'interno del software Gretl ed eventualmente valutare se la tecnica scelta sia adatta ai dati che si sta utilizzando.

In questo lavoro, dopo una breve introduzione al software Gretl, verranno illustrati i metodi di benchmarking e di disaggregazione temporale con indicatori di riferimento implementati dal pacchetto di funzioni "*Benchmarking*".

In particolare in questo elaborato sono state affrontate cinque diverse tecniche di benchmarking. Saranno descritti inizialmente gli approcci più semplici al problema, consistenti nel naive e nel pro-rata benchmarking. Successivamente verrà spiegato l'approccio di Denton, sia nella sua formulazione originaria del 1971 sia nella formulazione modificata da Cholette nel 1984. Nel fare ciò ci si è basati anche sul lavoro di Di Fonzo (2003). Verrà poi esposta la tecnica formulata da Cholette e Dagum (1994) illustrandone anche una sua variante ampiamente utilizzata dalle agenzie statistiche, cioè il "Proportional Cholette-Dagum method with AR(1) error". Infine l'ultima tecnica illustrata, utilizzata sia per il benchmarking che per la disaggregazione temporale è quella di Chow-Lin (1971), ampiamente utilizzata dagli istituti ufficiali di statistica (ad esempio l'Istat), per la produzione di conti economici trimestrali.

Nel capitolo 4 verrà esposta la struttura del pacchetto Benchmarking, ne verranno illustrate le principali caratteristiche e si spiegherà come utilizzarlo, sia tramite l'interfaccia grafica per l'utente di Gretl (GUI) che tramite riga di comando dal terminale di questo programma. Verrà poi anche riportato un esempio di utilizzo del pacchetto.

Nell'appendice A verrà presentata e analizzata criticamente la funzione `chowlin()` già presente nel software Gretl e ne verrà proposta una variante in grado di ovviare ad alcune debolezze che attualmente ne limitano le possibilità di uso.

Infine nell'appendice B verrà riportato tutto il codice scritto per il pacchetto *Benchmarking*.

Capitolo 2

Il software Gretl

Gretl (<http://gretl.sourceforge.net/>) è un software statistico open-source specializzato principalmente per l'econometria. Il suo nome è un acronimo per Gnu Regression, Econometrics and Time-series Library.

Inizialmente creato da Allin Cottrell sviluppando il lavoro di Ramu Ramanathan, il progetto oggi è seguito da un team di sviluppo che è considerevolmente più grande rispetto alle origini.

Il software dispone di una interfaccia grafica e può essere utilizzato congiuntamente a vari altri programmi, quali TRAMO/SEATS, R, Stata, Python, Octave, Ox e Julia. È stato scritto utilizzando il linguaggio di programmazione C, la sua interfaccia grafica è stata creata tramite il toolkit GTK+ e utilizza il programma gnuplot per creare i grafici. Come complemento all'interfaccia grafica è anche disponibile un programma client a riga di comando.

Il pregio principale di questo software è la sua capacità di essere allo stesso tempo user-friendly e sofisticato. La sua intuitiva interfaccia grafica utente rende molto semplice predisporre i propri dati per le varie analisi econometriche disponibili: Gretl mette a disposizione un'ampia varietà di tecniche di stima basate sui minimi quadrati, sia per singole equazioni che per sistemi di equazioni, compresi i modelli VAR e i modelli a correzione di errore. Comprende anche vari stimatori di massima verosimiglianza (ad es. probit, ARIMA, GARCH), mentre altri metodi avanzati sono programmabili dall'utente sfruttando le opportunità offerte dal programma.

Numerosi articoli dedicati a questo programma sono apparsi nel "Journal of Applied Econometrics" e nel "Journal of Statistical Software".

Per la sua natura open source e la vasta gamma di tecniche econometriche offerte, Gretl viene ampiamente utilizzato anche per l'insegnamento a livello universitario e compare come software di riferimento in alcuni libri di testo.

Nonostante queste caratteristiche, il grado di copertura delle tecniche statistiche offerto da Gretl non è paragonabile a quello di R, anch'esso un software open-source. Tuttavia

per facilitare la comunicazione tra i due programmi, Gretl offre la possibilità di salvare i dataset nei quali si lavora in un formato adatto per le analisi in R, lanciare una sessione di R con il dataset di gretl automaticamente caricato nel workspace di R e incorporare script di R all'interno di script di Gretl.

Infine, va ricordato che Gretl possiede un proprio linguaggio di scripting, chiamato Hansl. Lo scopo principale di questo linguaggio di programmazione è quello di automatizzare le attività più ripetitive e rendere agevole per gli utenti, anche non necessariamente scrittori di codici professionali, la creazione di funzioni e procedure che implementano tecniche non disponibili in Gretl. Hansl non si pone come un linguaggio di programmazione completamente sviluppato, come per esempio C o Python, ma dovrebbe essere considerato più come un linguaggio specifico e orientato soprattutto al dominio dell'econometria. Nonostante questo, Hansl possiede diverse funzionalità che supportano il lavoro avanzato come programmazione strutturata, ricorsione, strutture dati complesse e così via.

Per quanto riguarda la problematica che verrà affrontata in questo lavoro, l'unica procedura di disaggregazione temporale presente in Gretl allo stato attuale consiste nella funzione `chowlin()` che permette di convertire vettori di dati annuali o trimestrali in vettori di dati trimestrali o mensili. Un approfondimento sulla funzione `chowlin()` si trova nell'appendice A.

Capitolo 3

Le tecniche di benchmarking e di disaggregazione temporale

Il benchmarking temporale affronta il problema di combinare una serie di dati ad alta frequenza con una serie di dati meno frequenti per una determinata variabile, al fine di ottenere un risultato coerente.

In questo capitolo verrà discusso l'utilizzo del benchmarking per derivare dati trimestrali che sono consistenti con dei benchmark annuali e coerenti con l'evoluzione a breve termine della serie indicatrice trimestrale.

L'accuratezza delle stime trimestrali dipende dall'accuratezza della serie annuale di benchmark e della serie preliminare. Per questo motivo un prerequisito per la qualità della stima trimestrale è di fare affidamento ad informazioni che misurano esattamente ciò che sta accadendo nell'economia, sia in tempi normali che durante periodi di cambiamenti rapidi ed inaspettati. Il ruolo del benchmarking consiste nel combinare nel miglior modo possibile le serie annuali e trimestrali a disposizione.

Dal punto di vista dell'utilità economica il benchmarking viene utilizzato soprattutto per il confronto tra dati annuali e trimestrali, ma queste tecniche possono essere applicate anche ad altri tipi di confronti come per esempio tra dati annuali e mensili o tra dati trimestrali e mensili. Anche se in questo capitolo per facilitare l'esposizione dei concetti si utilizzeranno serie annuali e trimestrali, queste tecniche possono essere utilizzate per il benchmarking di qualsiasi tipo di confronto tra generici dati ad alta frequenza e dati a bassa frequenza.

Il benchmarking ha due principali scopi:

- Distribuzione trimestrale (o interpolazione) di dati annuali per costruire serie temporali trimestrali che rispettano i vincoli del Benchmark
- Estrapolazione per ricavare le stime dei trimestri per i quali i corrispondenti dati aggregati annuali non sono ancora disponibili (forecasting)

Idealmente, sia la distribuzione che l'estrapolazione di serie trimestrali deve essere basata su serie indicatrici che sono statisticamente ed economicamente correlate alle variabili annuali considerate. Il termine serie indicatrice è adottato in un senso lato in questo contesto: esso può indicare sia una misurazione sub-annuale della stessa variabile obiettivo che una variabile proxy che approssima fortemente il comportamento trimestrale della variabile obiettivo.

In questo contesto il metodo di benchmarking ideale deve essere quindi in grado di riuscire a conseguire questi 3 obiettivi:

- Stimare dati trimestrali che siano temporalmente consistenti con i dati annuali, quindi per esempio assicurarsi che la somma o la media dei dati trimestrali coincida con i dati annuali.
- Preservare il più possibile i movimenti di breve periodo presenti nell'indicatore sotto le restrizioni fornite dal dato annuale.
- Fare in modo che, nel caso del forecasting, la somma dei quattro trimestri dell'anno sia il più possibile vicina all'ignoto dato annuale futuro.

3.1 Naive Benchmarking

La variante *naive* è l'approccio più semplice e intuitivo alla problematica del benchmarking. Nel caso di variabili di flusso, applicando questa tecnica le stime trimestrali sono ottenute semplicemente sommando ai valori della serie preliminare indicatrice una porzione fissa della discrepanza tra i dati preliminari aggregati annualmente e i dati annuali di benchmark.

Formalizzando il problema, nel naive-benchmarking le stime trimestrali saranno ottenute in questo modo:

$$\tilde{y}_{T,j} = p_{T,j} + \frac{1}{s}(y_{0T} - p_{0T}), \quad T = 1, \dots, N, \quad j = 1, \dots, s \quad (1)$$

dove:

$\tilde{y}_{T,j}$ è il valore stimato per il j-esimo trimestre del T-esimo anno

$p_{T,j}$ è il valore della serie preliminare al j-esimo trimestre del T-esimo anno

s è il numero di dati ad alta frequenza per ogni dato a bassa frequenza (nel nostro caso 4)

y_{0T} è il valore della serie di benchmark per il T-esimo anno

p_{0T} è il valore della serie preliminare aggregata annualmente al T-esimo anno (pari a $\sum_{j=1}^s p_{T,j}$ nel caso di distribuzione)

Applicando questa tecnica il vincolo fondamentale di aggregazione temporale del benchmarking verrà rispettato, ma l'utilizzo di una porzione fissa della discrepanza tra i dati preliminari aggregati annualmente e i dati annuali di benchmark potrebbe portare ad ottenere delle stime che non preservano efficacemente i movimenti trimestrali della serie preliminare. Per questo e per altri motivi che verranno spiegati in seguito l'utilizzo del naive benchmarking è fortemente sconsigliato.

3.2 Pro-rata Benchmarking

Un altro metodo di benchmarking che in parte ovvia al problema della variante naive è il pro-rata benchmarking. In questo caso le stime finali sono ottenute dalla somma tra valori della serie preliminare ed un termine di aggiustamento rappresentato da una porzione della discrepanza tra il noto benchmark e la somma dei valori preliminari proporzionale al peso degli stessi valori della serie preliminare.

Formalizzando il problema e utilizzando la notazione precedente, nel pro-rata benchmarking le stime trimestrali sono ottenute in questo modo:

$$\tilde{y}_{T,j} = p_{T,j} + \frac{p_{T,j}}{p_{0T}}(y_{0T} - p_{0T}), \quad T = 1, \dots, N, \quad j = 1, \dots, s. \quad (2)$$

Questa espressione, semplificata, può anche essere riscritta come:

$$\tilde{y}_{T,j} = p_{T,j} \left(\frac{y_{0T}}{p_{0T}} \right), \quad T = 1, \dots, N, \quad j = 1, \dots, s \quad (3)$$

La stima finale si ottiene dunque dal prodotto tra i valori della serie preliminare per il rapporto tra la serie di benchmark e la serie preliminare aggregata annualmente (indicato in letteratura come *BI ratio – benchmark to indicator ratio*). Nel caso però questo rapporto passando da un anno all'altro sia soggetto a cambiamenti, il pro-rata benchmarking subisce il cosiddetto "step problem" (a cui è soggetta, con dinamiche un po' differenti, anche la variante naive).

Quando questo accade, nei dati stimati finali viene introdotta una discontinuità nei tassi di crescita tra l'ultimo trimestre di ogni anno e il primo trimestre dell'anno successivo. In altre parole la variazione tra l'ultimo trimestre di un anno e il primo trimestre del successivo rifletterà la differenza nei vari BI ratio piuttosto che la reale dinamica tra trimestri successivi. Maggiori saranno le differenze nei valori dei rapporti tra la serie annuale e la serie indicatrice aggregata, maggiori saranno le discontinuità rilevate nei tassi di crescita.

Per questo motivo il pro-rata benchmarking rimane un'opzione valida solamente quando questi rapporti rimangono costanti nel tempo, in caso contrario - quindi nella stragrande maggioranza dei casi - la sua applicazione porterà a un risultato finale discutibile in quanto a qualità.

3.3 Il metodo di Denton

Un metodo, più complesso dei precedenti, ideato appositamente per superare il primo citato step problem è il metodo proposto da Denton (1971). Nel suo approccio al benchmarking, Denton specifica una funzione obiettivo con cui viene formulato un problema di ottimizzazione quadratica vincolata.

Definiamo n il numero di dati trimestrali da stimare, N il numero dei dati annuali di benchmark, s il numero delle osservazioni per ogni dato di benchmark, $y = [y_1, y_2, \dots, y_n]'$ il vettore delle n osservazioni trimestrali da stimare, $p = [p_1, p_2, \dots, p_n]'$ il vettore degli n dati preliminari e $y_0 = [y_{01}, y_{02}, \dots, y_{0N}]'$ il vettore degli N dati annuali di benchmark. Specifichiamo una funzione di penalizzazione $f(y, p)$ ed esprimiamo il problema come quello di scegliere y in modo da minimizzare $f(y, p)$ soggetta al vincolo

$$y_0 = Jy \quad \text{dove} \quad J = \begin{bmatrix} j & 0 & \dots & 0 \\ 0 & j & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & j \end{bmatrix} \quad (4)$$

Con j vettore riga s -dimensionale nel quale ogni elemento equivale a 1 e 0 vettore riga s -dimensionale nullo. J ha dimensione $(N \times n)$ ed è chiamata matrice di aggregazione temporale. Essa è tale per cui, per ogni vettore di dati trimestrali z , Jz fornisce le somme annuali di z .

Nel caso si debba procedere all'estrapolazione, cioè se n è maggiore di $N \times s$, alla matrice J dovranno essere affiancati un totale di $n - N \times s$ vettori colonna nulli.

Viene considerata una classe di funzioni obiettivo di tipo quadratico

$$(y - p)'M(y - p) \quad (5)$$

ossia una forma quadratica delle differenze tra i valori della serie storica originale e quella della serie storica aggiustata, con M matrice simmetrica $(n \times n)$, che verrà specificata in seguito.

Il problema di ottimizzazione può essere risolto usando la funzione lagrangiana

$$\mathcal{L} = (y - p)'M(y - p) - 2\lambda'(y_0 - Jx) \quad (6)$$

dove $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_N]'$ è un vettore $N \times 1$ di moltiplicatori di Lagrange. La soluzione si ottiene prendendo le derivate parziali di \mathcal{L} rispetto agli elementi di y e λ , uguagliandole a zero, e risolvendo. La soluzione può essere espressa nella forma (Di Fonzo, 2003):

$$\begin{bmatrix} \tilde{y} \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} M & J' \\ J & 0 \end{bmatrix}^{-1} \begin{bmatrix} Mp \\ y_0 \end{bmatrix} \quad (7)$$

Per la soluzione di questo sistema e la specificazione della matrice M , oltre all'approccio originale di Denton (1971) ne è stato formulato anche un altro, ad opera di Cholette (1984).

3.3.1 Original Denton

Come vedremo nel seguito, nella formulazione originale del problema la matrice M è sempre non-singolare. Sfruttando quindi un noto lemma di inversione di matrice partizionata (Helmut W., 1978) abbiamo

$$\begin{bmatrix} M & J' \\ J & 0 \end{bmatrix}^{-1} = \begin{bmatrix} M^{-1} - M^{-1}J'(JM^{-1}J')^{-1}JM^{-1} & M^{-1}J'(JM^{-1}J')^{-1} \\ (JM^{-1}J')^{-1}JM^{-1} & -(JM^{-1}J')^{-1} \end{bmatrix} \quad (8)$$

La soluzione può quindi essere riscritta come

$$\begin{aligned} \tilde{y} &= [M^{-1} - M^{-1}J'(JM^{-1}J')^{-1}JM^{-1}]Mp + M^{-1}J'(JM^{-1}J')^{-1}y_0 \\ &= p - M^{-1}J'(JM^{-1}J')^{-1}Jp + M^{-1}J'(JM^{-1}J')^{-1}y_0 \\ &= p + M^{-1}J'(JM^{-1}J')^{-1}(y_0 - Jp) \end{aligned} \quad (9)$$

Per la definizione di M la prima soluzione che propone Denton è una funzione di penalizzazione basata sulle differenze prime della serie originale e di quella aggiustata:

$$f(y, p) = \sum_{t=1}^n (\Delta y_t - \Delta p_t)^2 = \sum_{t=1}^n [\Delta(y_t - p_t)]^2 \quad (10)$$

Dove Δ è l'operatore di differenza prima. La funzione obiettivo includerà quindi anche il termine $\Delta(y_1 - p_1) = (y_1 - p_1) - (y_0 - p_0)$, in cui il deponente 0 qui si riferisce non al dato aggregato annuale ma all'ultimo periodo dell'anno precedente all'anno 1. Il periodo 0 è fuori dal range della serie che deve essere stimata, per cui Denton (1971) assume $y_0 = p_0$, da cui deriva:

$$\Delta(y_1 - p_1) = y_1 - p_1. \quad (11)$$

Il vettore delle differenze prime può allora essere espresso come $D(y - p)$ dove D è una matrice (nxn) delle differenze prime approssimate pari a

$$D = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \quad (12)$$

La forma quadratica che verrà minimizzata, soggetta ai vincoli annuali è ora $(y - p)'D'D(y - p)$. In cui M è quindi uguale a $D'D$.

3.3.2 Modified Denton

La modifica proposta da Cholette (1984) al procedimento di Denton riguarda la rimozione dell'ipotesi $y_0 = p_0$. Cholette ha mostrato che operando come proposto da Denton (1971) i risultati sono soggetti ad una ingiustificata distorsione nei valori stimati per il primo anno, che poi si ripercuote sui successivi, allontanando i risultati dall'obiettivo del massimo parallelismo tra la serie preliminare e quella aggiustata.

Utilizzando sempre la funzione obiettivo basata sulle differenze prime, la specificazione proposta da Cholette semplicemente si astiene dal postulare $y_0 = p_0$ e porta alla seguente funzione obiettivo:

$$f(y, p) = \sum_{t=2}^n [\Delta(y_t - p_t)]^2 \quad (13)$$

Soggetta all'usuale vincolo $y_0 = Jy$.

In questo caso il vettore delle differenze prime può essere espresso come $D_1(y - p)$, dove D_1 è una matrice $((n-1) \times n)$ costruita in questo modo

$$D_1 = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \quad (14)$$

La matrice M nella forma quadratica da minimizzare quindi sarà uguale a $D_1' D_1$.

Va notato che in questa versione modificata del metodo di Denton la matrice M è singolare, e dunque non è possibile semplificare la soluzione (7) tramite le espressioni (8) e (9). Il sistema (7) va dunque risolto senza l'ausilio di alcuna semplificazione.

Il Modified Denton comporta l'inversione di una matrice $(N+n) \times (N+n)$ mentre per la tecnica di Denton originale la matrice da invertire, di dimensione $n \times n$ è più piccola. Negli anni 80 ciò poteva avere qualche impatto in termini di complessità delle procedure di calcolo. Ma con la attuale potenza di calcolo, questo non appare più come un problema da prendere in considerazione.

3.4 Altre varianti del metodo di Denton

In base alla definizione della matrice M si possono ottenere alcune varianti del metodo di Denton, tra cui compare la sua versione più diffusa, che utilizza le differenze prime proporzionali.

Inoltre si osserva che ponendo la matrice M uguale alla matrice Identità otteniamo il naive benchmarking, mentre ponendo la matrice M uguale ad una matrice che contiene sulla diagonale principale i valori della serie preliminare otteniamo il pro-rata benchmarking.

In particolare abbiamo che nel naive benchmarking la funzione di penalizzazione sarà definita come

$$f(y, p) = \sum_{t=1}^n (y_t - p_t)^2 \quad (15)$$

Con $M = I_n$

Mentre nel pro-rata benchmarking avremo che

$$f(y, p) = \sum_{t=1}^n \frac{(y_t - p_t)^2}{p_t} \quad (16)$$

Con $M = \hat{p}^{-1}$ dove \hat{p} è una matrice ($n \times n$) contenente sulla diagonale principale il vettore dei dati preliminari.

$$\hat{p} = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_n \end{bmatrix} \quad (17)$$

3.4.1 Additive First Differences (AFD)

Se si considera la funzione di penalizzazione definita in (10) o in (13) a seconda che si stia utilizzando la formulazione originaria di Denton (1971) o quella di Cholette (1984), la variante che si ottiene del metodo di Denton è detta "Additive First Differences".

3.4.2 Additive Second Differences (ASD)

La proposta originaria di Denton per questa variante considera la seguente funzione di penalizzazione:

$$\begin{aligned}
 f(y, p) &= \sum_{t=1}^n [\Delta^2(y_t - p_t)]^2 \\
 &= \sum_{t=1}^n \{[(y_t - p_t) - (y_{t-1} - p_{t-1})] - [(y_{t-1} - p_{t-1}) - (y_{t-2} - p_{t-2})]\}^2 \quad (18)
 \end{aligned}$$

con $y_{-1} = p_{-1}$ e $y_0 = p_0$.

In questo caso la matrice M è uguale a $D'D'DD$ con D definita dalla (12).

La variante di Cholette (1984) considera invece:

$$\begin{aligned}
 f(y, p) &= \sum_{t=3}^n [\Delta^2(y_t - p_t)]^2 \\
 &= \sum_{t=3}^n \{[(y_t - p_t) - (y_{t-1} - p_{t-1})] - [(y_{t-1} - p_{t-1}) - (y_{t-2} - p_{t-2})]\}^2 \quad (19)
 \end{aligned}$$

In questo caso la matrice M è uguale a $D_1'D_2'D_2D_1$ con D_1 definita dalla (14) e

$$D_2 = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \quad \text{è una matrice } ((n-2) \times (n-1)). \quad (20)$$

3.4.3 Proportional First Differences (PFD)

Nella formulazione originaria la funzione di penalizzazione nel caso delle prime differenze proporzionali prende la forma:

$$\begin{aligned}
 f(y, p) &= \sum_{t=1}^n \Delta \left(\frac{y_t - p_t}{p_t} \right)^2 = \sum_{t=1}^n \left(\frac{y_t - p_t}{p_t} - \frac{y_{t-1} - p_{t-1}}{p_{t-1}} \right)^2 \\
 &= \sum_{t=1}^n \left(\frac{y_t}{p_t} - \frac{y_{t-1}}{p_{t-1}} \right)^2 \quad (21)
 \end{aligned}$$

con $y_0 = p_0$.

In questo caso la matrice M è uguale a $\hat{p}^{-1}D'D\hat{p}^{-1}$.

Nella variante di Cholette (1984) la funzione obiettivo è data da:

$$\begin{aligned}
 f(y, p) &= \sum_{t=2}^n \Delta \left(\frac{y_t - p_t}{p_t} \right)^2 = \sum_{t=2}^n \left(\frac{y_t - p_t}{p_t} - \frac{y_{t-1} - p_{t-1}}{p_{t-1}} \right)^2 \\
 &= \sum_{t=2}^n \left(\frac{y_t}{p_t} - \frac{y_{t-1}}{p_{t-1}} \right)^2
 \end{aligned} \tag{22}$$

e la matrice M è uguale a $\hat{p}^{-1} D_1' D_1 \hat{p}^{-1}$.

3.4.4 Proportional Second Differences (PSD)

Nella formulazione originaria la funzione di penalizzazione nel caso delle seconde differenze proporzionali prende la forma:

$$\begin{aligned}
 f(y, p) &= \sum_{t=1}^n \Delta^2 \left(\frac{y_t - p_t}{p_t} \right)^2 \\
 &= \sum_{t=1}^n \left[\left(\frac{y_t - p_t}{p_t} - \frac{y_{t-1} - p_{t-1}}{p_{t-1}} \right) - \left(\frac{y_{t-1} - p_{t-1}}{p_{t-1}} - \frac{y_{t-2} - p_{t-2}}{p_{t-2}} \right) \right]^2 \\
 &= \sum_{t=1}^n \left[\left(\frac{y_t}{p_t} - \frac{y_{t-1}}{p_{t-1}} \right) - \left(\frac{y_{t-1}}{p_{t-1}} - \frac{y_{t-2}}{p_{t-2}} \right) \right]^2
 \end{aligned} \tag{23}$$

Anche in questo caso dalla funzione di penalizzazione sono stati eliminati tutti i valori al di fuori dal range impostando $y_{-1} = p_{-1}$ e $y_0 = p_0$.

La matrice M è uguale a $\hat{p}^{-1} D' D' D D \hat{p}^{-1}$.

Secondo la modifica di Cholette in questa variante la funzione di penalizzazione è definita da:

$$\begin{aligned}
 f(y, p) &= \sum_{t=3}^n \Delta^2 \left(\frac{y_t - p_t}{p_t} \right)^2 \\
 &= \sum_{t=3}^n \left[\left(\frac{y_t - p_t}{p_t} - \frac{y_{t-1} - p_{t-1}}{p_{t-1}} \right) - \left(\frac{y_{t-1} - p_{t-1}}{p_{t-1}} - \frac{y_{t-2} - p_{t-2}}{p_{t-2}} \right) \right]^2 \\
 &= \sum_{t=3}^n \left[\left(\frac{y_t}{p_t} - \frac{y_{t-1}}{p_{t-1}} \right) - \left(\frac{y_{t-1}}{p_{t-1}} - \frac{y_{t-2}}{p_{t-2}} \right) \right]^2
 \end{aligned} \tag{24}$$

e la matrice M è uguale a $\hat{p}^{-1} D_1' D_2' D_2 D_1 \hat{p}^{-1}$.

3.5 Il metodo di Cholette-Dagum

Un limite della tecnica di Denton è il fatto di non prendere in considerazione le possibili presenze di autocorrelazione ed eteroschedasticità.

Per ovviare a ciò, Cholette e Dagum (1994) hanno proposto una tecnica di benchmarking basata su un modello di regressione esteso, da stimare tramite i minimi quadrati generalizzati. Il metodo di Cholette-Dagum mette a disposizione una struttura molto flessibile per il benchmarking. Essa si basa su un modello statistico che permette la presenza di distorsione ed errori autocorrelati nella serie indicatrice, e la presenza di benchmark non vincolanti. La serie risultante è calcolata come soluzione dei minimi quadrati generalizzati di un modello di regressione con effetti deterministici ed effetti stocastici.

Il modello di regressione sul quale si basa il metodo di Cholette-Dagum consiste nelle seguenti due equazioni:

$$p_t = a + \theta_t + e_t, \quad E(e_t) = 0, \quad E(e_t e_{t-k}) \neq 0, \quad t = 1, \dots, n \quad (25)$$

$$y_{0j} = \sum_{t \in J} \theta_t + w_j, \quad E(w_j) = 0, \quad E(w_j w_{j-k}) \neq 0, \quad j = 1, \dots, N \quad (26)$$

Nella equazione (20) p_t indica i valori della serie preliminare, θ_t indica il valore non osservato della variabile da stimare e a indica un parametro di distorsione.

I parametri θ_t e a devono essere stimati.

Le stime di θ_t saranno la serie stimata finale. A seconda della natura della variabile in questione, θ_t può seguire un modello stocastico o deterministico. e_t indica gli errori presenti nelle osservazioni con una specifica struttura di covarianza. Perciò l'equazione (25) afferma che l'osservazione del vero valore della variabile è contaminata da errore e distorsione.

Nell'equazione (26) invece y_{0j} indica il dato annuale di benchmark. Se il benchmark non è soggetto ad errore, e quindi $w_j = 0$ $\sigma_j^2 = 0$, esso viene considerato completamente affidabile e vincolante; in caso contrario, è non vincolante. L'equazione (26) quindi indica che anche le osservazioni annuali della variabile obiettivo possono essere contaminate da errori, che risultano di solito molto più piccoli e con una specifica struttura di covarianza. Viene assunto che e_t e w_j siano mutuamente indipendenti, in quanto in questo schema si assume che i dati annuali e trimestrali provengono da due fonti diverse.

Il sistema di equazioni (25) (26) può essere riscritto i termini matriciali in un'unica equazione:

$$\begin{bmatrix} p \\ y_0 \end{bmatrix} = \begin{bmatrix} 1 & I \\ 0 & J \end{bmatrix} \begin{bmatrix} a \\ \theta \end{bmatrix} + \begin{bmatrix} e \\ w \end{bmatrix} = X \begin{bmatrix} a \\ \theta \end{bmatrix} + \begin{bmatrix} e \\ w \end{bmatrix} \quad (27)$$

$$E(e) = 0, \quad E(w) = 0, \quad E(e e') = V_e, \quad E(w w') = V_w, \quad E(e w') = 0$$

Dove 1 è un vettore di 1 ($n \times 1$), I è una matrice identità ($n \times n$), 0 è un vettore nullo ($N \times 1$) e J è una matrice di aggregazione ($N \times n$). L'equazione (27) indica che la serie da stimare θ si adatta sia alle osservazioni annuali che trimestrali ed è tale che i disturbi hanno matrici di covarianza note V_e e V_w .

Il modello (22) può anche essere riscritto come

$$Y = X\beta + u, \quad E(u) = 0, \quad E(u u') = V \quad (28)$$

Dove $Y' = [p' y_0']$, $\beta' = [a \theta']$, $u' = [e' w']$, V è una matrice diagonale a blocchi diagonali con blocchi V_e e V_w e dove X è la matrice dei regressori già implicitamente definita nella (27).

La soluzione dei minimi quadrati generalizzati della (28) porta a

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (29)$$

Se V è la vera matrice di covarianza dei disturbi u , la matrice di covarianza delle stime di $\hat{\beta}$ è data da

$$\text{cov } \hat{\beta} = (X'V^{-1}X)^{-1} \quad (30)$$

La sostituzione delle partizioni di X , V e $\hat{\beta}$ nella (29) e la trasformazione delle matrici porta a

$$\begin{bmatrix} \hat{a} \\ \hat{\theta} \end{bmatrix} = \begin{bmatrix} 1'V_e^{-1}1 & 1'V_e^{-1} \\ V_e^{-1}1 & (V_e^{-1} + J'V_w^{-1}J) \end{bmatrix}^{-1} \begin{bmatrix} 1'V_e^{-1} & 0 \\ V_e^{-1} & J'V_w^{-1} \end{bmatrix} \begin{bmatrix} p \\ y_0 \end{bmatrix} \quad (31)$$

$$\begin{bmatrix} \hat{a} \\ \hat{\theta} \end{bmatrix} = \begin{bmatrix} V_{aa} & V_{a\theta} \\ V_{\theta a} & V_{\theta\theta} \end{bmatrix} \begin{bmatrix} 1'V_e^{-1} & 0 \\ V_e^{-1} & J'V_w^{-1} \end{bmatrix} \begin{bmatrix} p \\ y_0 \end{bmatrix} \quad (32)$$

Dove V_{aa} e $V_{\theta\theta}$ sono la varianza stimata di \hat{a} e la matrice di covarianza di $\hat{\theta}$.

Tramite inversione per parti delle matrici della (31) ed altre trasformazioni algebriche V_{aa} , $V_{a\theta}$ e $V_{\theta\theta}$ possono essere riscritte come

$$V_{aa} = 1/[1'J'(JV_eJ' + V_w)^{-1}J1] = h \quad (33)$$

$$V_{a\theta} = V'_{\theta a} = -h1' + h1'J'(JV_eJ' + V_w)^{-1}JV_e \quad (34)$$

$$V_{\theta\theta} = [V_e - V_eJ'(JV_eJ' + V_w)^{-1}JV_e] + [I - V_eJ'(JV_eJ' + V_w)^{-1}J]1h1' \\ \times [I - V_eJ'(JV_eJ' + V_w)^{-1}J]' \quad (35)$$

Questo porta a

$$\hat{a} = -h1'J'(JV_eJ' + V_w)^{-1}(y_0 - Jp) \quad (36)$$

$$\hat{\theta} = p^* + V_eJ'(JV_eJ' + V_w)^{-1}(y_0 - Jp^*), \quad p^* = [p_1 - \hat{a} \dots p_n - \hat{a}] \quad (37)$$

La stima finale della serie trimestrale è data dalla (37) e la sua matrice di covarianza dalla equazione (35). In assenza di distorsione nel modello la stima finale della serie è data

dalla (37) dove al posto di p^* viene utilizzato semplicemente p con matrice di covarianza ridotta al primo termine della (35).

Un particolare caso del metodo di Cholette-Dagum ampiamente utilizzato dalle agenzie statistiche ed implementato nel pacchetto *Benchmarking* è il “Proportional Cholette-Dagum method with AR(1) error”.

In questa variante del metodo di Cholette-Dagum le equazioni (25) e (26) prendono la seguente forma:

$$p_t = \theta_t + e'_t \quad (38)$$

$$y_{0t} = \sum_{t \in j} \theta_t \quad (39)$$

Con $e'_t = \phi e'_{t-1} + v_t$ (40)

Dove e'_t è l'errore trimestrale autocorrelato ed eteroschedastico. La relazione proporzionale tra y_t e p_t è ottenuta assumendo che l'errore trimestrale in (38) sia in una relazione proporzionale con la serie preliminare, in particolare $e'_t = e_t/p_t$.

Il valore del parametro autoregressivo ϕ è molto importante per l'estrapolazione. Il suo valore indica quanto velocemente la serie stimata θ_t converge alla serie indicatrice y_t . Valori di ϕ vicini allo 0 portano ad una rapida convergenza, mentre valori vicini ad 1 tendono a mantenere il *BI ratio* invariato per l'anno successivo.

Il manuale dell'International Monetary Found per i conti nazionali trimestrali (2017) suggerisce un valore per ϕ di 0.84 per le trimestralizzazioni di valori annuali. Questo valore infatti assicura che circa il cinquanta per cento dell'errore osservato nell'ultimo trimestre venga eliminato dopo un anno.

Lo stesso risultato si ottiene ponendo il valore di ϕ pari a 0.79 per il benchmarking di serie mensili a partire da benchmark trimestrali e pari a 0.94 per il benchmarking di serie mensili a partire da benchmark annuali.

3.6 Il metodo di Chow-Lin

Un altro metodo ampiamente utilizzato per il benchmarking e la disaggregazione temporale, anch'esso basato su un modello di regressione, è il metodo di Chow-Lin(1971).

Tale tecnica assume un modello di regressione tra le vere (e non osservate) osservazioni trimestrali y_t e un insieme di q serie trimestrali ad esse legate $x_{1,t}, \dots, x_{q,t}$:

$$y_t = \sum_{j=1}^q \beta_j x_{j,t} + u_t \quad t = 1, \dots, n \quad (41)$$

con

$$u_t = \rho u_{t-1} + v_t \quad (42)$$

Dove

- y_t è il valore trimestrale da stimare;
- β_j è il coefficiente di regressione per il j-esimo indicatore;
- $x_{j,t}$ è il j-esimo indicatore trimestrale;
- u_t è un errore casuale che si assume segua un AR(1), con v_t innovazioni indipendenti e identicamente distribuite;
- n è il numero di trimestri, che include possibilmente le estrapolazioni ($q \geq 4y$).

Siccome y_t non è osservato, il modello precedentemente indicato non può essere stimato. Tuttavia, Chow e Lin assumono che la stessa relazione tra y_t e l'indicatore trimestrale si mantenga vera alla frequenza annuale. Perciò, questo modello viene temporalmente aggregato:

$$y_{0T} = \sum_{j=1}^q \beta_j x_{0T} + u_T^a \quad T = 1, \dots, N \quad (43)$$

dove

$$y_{0T} = \sum_{t=4T-3}^{4T} y_t \quad \text{è la nota serie storica annuale}$$

β_j è il coefficiente di regressione per il j-esimo indicatore (che è dunque costante tra le frequenze)

$$x_{0T} = \sum_{t=4T-3}^{4T} x_t \quad \text{è il j-esimo indicatore annualmente aggregato}$$

u_T^a è l'errore annuale ARMA(1,1) derivato dal modello trimestrale AR(1).

Chow e Lin (1971) derivano il migliore stimatore lineare non distorto di y_t stimando simultaneamente il vettore di coefficienti di regressione $\hat{\beta}$ e il coefficiente AR $\hat{\rho}$ del modello (43). La serie stimata \hat{y}_t consiste di due componenti: una proveniente dall'effetto di regressione $\sum_{j=1}^p \hat{\beta}_j x_{j,t}$ e uno dal residuo trimestrale stimato \hat{u}_t . Il modello di regressione può contenere effetti deterministici (costanti, trend, ecc) e indicatori relativi, che dovrebbero essere scelti in base alle caratteristiche della serie annuale da disaggregare.

Data X matrice di regressione, J matrice di aggregazione definita come in (4), V matrice di varianza-covarianza del modello (41), $X_0 = JX$ matrice di regressione aggregata e V_0 matrice di varianza-covarianza del modello (43), l'espressione finale dello stimatore secondo Chow-Lin (1971) sarà quindi definita da:

$$\hat{y} = X\hat{\beta} + VJ'(JVJ')^{-1}(y_0 - X_0\hat{\beta}) \quad (44)$$

Dove

$$\hat{\beta} = (X_0'V_0^{-1}X_0)^{-1}X_0'V_0^{-1}y_0 \quad (45)$$

Nella stima di serie trimestrali la più utilizzata combinazione di regressori è un termine costante più un indicatore. La stima di ρ può essere ottenuta tramite massima verosimiglianza o tramite minimi quadrati pesati. La stima del valore di ρ è preferibile che sia positiva per preservare i movimenti originali della componente di regressione. Tuttavia, dal momento che la stima di ρ viene fornita dai dati osservati e non scelta dall'utente, può capitare che questa risulti negativa e che quindi la componente dell'errore domini i movimenti a breve termine della serie risultante.

Questo metodo può essere utilizzato anche per la disaggregazione temporale senza l'utilizzo di indicatori di riferimento, per esempio utilizzando un indice temporale come variabile indicatrice nel modello di regressione. Un approfondimento su come questo metodo è stato implementato all'interno del software Gretl si trova nell'appendice A.

Nel pacchetto *Benchmarking* il modello di regressione utilizzato per applicare questo metodo consiste in un termine costante più un indicatore di riferimento. Inoltre è possibile scegliere se inserire la stima di ρ manualmente oppure ottenerla dai dati osservati.

Capitolo 4

Il pacchetto *Benchmarking*

In questa sezione verrà descritto il codice e l'utilizzo del pacchetto di gretl per il Benchmarking e la disaggregazione temporale di serie storiche con indicatori di riferimento. Prima verrà spiegato in cosa consiste il codice e che forma deve avere il dataset per funzionare correttamente, poi come utilizzarlo tramite GUI e tramite script. Infine verranno mostrati gli output generati dal pacchetto con l'utilizzo di un esempio.

Il corretto funzionamento delle varie tecniche di benchmarking implementate nel pacchetto è stato verificato confrontando i risultati forniti da quest'ultimo con quelli della libreria di matlab "Temporal Disaggregation" di Quilis E. M. (2019) e della funzione di excel XLBPM 2.0 di Marco Marini (2014).

4.1 La struttura del pacchetto

Il pacchetto consiste in una funzione pubblica chiamata *Benchmarking* e tre funzioni private.

La funzione pubblica contiene sia le indicazioni per l'interfaccia grafica per l'utente (GUI) che gli algoritmi per le tecniche di benchmarking. Delle funzioni private, una permette di convertire un vettore colonna in una matrice diagonale, una viene utilizzata quando si sceglie Chow-Lin(auto) come funzione obiettivo da minimizzare e le altre sono dedicate alla creazione di grafici.

Gli argomenti della funzione pubblica fondamentale per il funzionamento del pacchetto sono i seguenti:

```
function series Benchmark (series low "series of temporally aggregated
constraints",
series high "Preliminary series to be benchmarked",
int k[0:2:0] "Aggregation order" {"3: quarterly/monthly", "4:
annual/quarterly", "12: annual/monthly"},
```

```

int j[0:3:0] "Type of temporal aggregation" {"sum", "average",
"sample of the last", "sample of the first"},
int t[0:11:8] "Benchmarking variants" {"Naive benchmarking",
"Pro-rata benchmarking", "Original Denton: additive first
differences", "Original Denton: additive second differences",
"Original Denton: proportional first differences", "Original
Denton: proportional second differences", "Modified Denton:
additive first differences", "Modified Denton: additive second
differences", "Modified Denton: proportional first
differences", "Modified Denton: proportional second
differences", "Dagum-Cholette", "Chow-Lin (manual)", "Chow-Lin
(auto)"},
scalar phi[0:0.999:0.84] "Autoregressive parameter (used only
with Dagum-Cholette and Chow-Lin)",
bool st[1] "Show text",
bool sg[0] "Show graphs")

```

I numeri tra parentesi quadre, dove presenti, indicano i valori validi per la variabile mentre il testo tra le virgolette indica le etichette che verranno utilizzate per la GUI:

- `series low` è la serie storica a bassa frequenza che contiene i valori da disaggregare;
- `series high` è la serie storica ad alta frequenza utilizzata come serie preliminare o come indicatore di riferimento;
- `int k` indica il tipo di conversione che viene fatta tra la serie originale a frequenza più bassa e quella a frequenza più alta;
- `int j` indica il tipo di aggregazione temporale;
- `int t` indica la tecnica di benchmarking;
- `scalar phi` è il valore dato al parametro autoregressivo quando l'utente sceglie Dagum-Cholette o Chow-Lin (manual)
- `bool st` indica se mostrare (default) o meno eventuali messaggi di warning e un output di testo sulle serie trattate;
- `bool sg` indica se mostrare o meno (default) dei grafici informativi sulle serie trattate.

Evocare questa funzione senza assegnarla ad una variabile produrrà solamente degli output grafici e dei print sulla console di gretl. In caso contrario, oltre agli output precedenti verrà creata (o sostituita, se viene utilizzato il nome di una variabile di tipo `series` già presente nel dataset in cui si sta lavorando) una serie storica, con il nome della variabile assegnata, contenente la serie disaggregata calcolata dalla funzione.

4.2 Operazioni preliminari

Utilizzare il pacchetto Benchmarking nel contesto di gretl pone subito un problema. Infatti le tecniche di benchmarking operano su serie storiche di frequenze diverse ma “in un dataset di serie storiche di gretl, tutte le serie devono avere la stessa frequenza” (Gretl user’s guide, 2018, pp. 22).

Lo stesso gretl però ci permette di aggirare il problema in due diversi modi: o tramite l’utilizzo di una *midas list* (un insieme di serie che contengono i valori per periodo di una singola serie a più alta frequenza, sviluppate per l’implementazione delle tecniche di stima di modelli a frequenza mista, detto appunto MIDAS) oppure tramite dei comandi di gretl che permettono di convertire serie da una frequenza più bassa ad una più alta senza alterare l’informazione in esse contenuta.

Tuttavia con le *midas list* saremmo costretti ad operare in un dataset definito alla frequenza della serie aggregata, il che risulta piuttosto scomodo nel nostro caso. Si è dunque deciso di sviluppare il pacchetto operando su un dataset alla frequenza della serie preliminare. All’interno di questo dataset dovranno essere presenti la serie preliminare e la serie da disaggregare: quest’ultima espressa alla frequenza del dataset ripetendo i suoi valori nei vari sottoperiodi.

Poiché gretl non permette di gestire l’importazione dei dati o la definizione della struttura del dataset all’interno delle funzioni create per i pacchetti, la creazione di un dataset con le sopra citate caratteristiche dovrà rimanere a carico dell’utente. Nel seguito viene presentata una procedura che permette di ottenere il dataset di cui abbiamo bisogno in modo abbastanza semplice.

Se le serie storiche da utilizzare non sono presenti all’interno di un database di gretl, si può procedere in questo modo:

- Importare i dati della serie da disaggregare come serie storica alla sua frequenza originale
- Convertire il dataset alla frequenza della serie preliminare ripetendo i valori a frequenza inferiore usando il menu Dati / Espandi dati..., e selezionando “Ripeti i valori di frequenza inferiori”
- Aggiungere i dati della serie preliminare andando su File / Aggiungi dati...

Nel caso sia necessario estrapolare oltre che disaggregare, e quindi la serie preliminare non è in linea con la serie da disaggregare, prima di aggiungere i dati della serie preliminare sarà necessario aggiungere le osservazioni in più tramite Dati / Aggiungi osservazioni...

Se invece i dati provengono da un database presente all’interno di gretl (ai quali si può accedere usando il menu File / Database / Gretl) basterà importare prima i dati della serie preliminare e poi quelli della serie aggregata selezionando, quando richiesto, “Ripeti i valori di frequenza inferiore”.

Se il campione sul quale si sta lavorando non inizia insieme alla serie da disaggregare è necessario impostare un intervallo che inizia insieme a quest'ultima tramite Campione / Imposta intervallo... .

4.2.1 Esempio

Supponiamo di avere a disposizione due file .gdt (gretl data set) contenenti due serie storiche osservate a cadenze temporali differenti e di voler creare un dataset per utilizzare il pacchetto *Benchmarking*. Il primo file "VAtrimestrale.gdt" contiene la serie storica trimestrale del valore aggiunto dell'industria italiana dal primo trimestre del 1995 all'ultimo trimestre del 2017. Il secondo file "indproit.gdt" contiene la serie storica mensile dell'indice di produzione industriale italiano dal gennaio del 1995 all'ottobre del 2018.

Si procederà in questo modo:

Dopo aver aperto una nuova istanza di gretl si dovrà prima di tutto caricare nel dataset il file contenente la serie storica da disaggregare, nel nostro caso "VAtrimestrale.gdt", come mostrato in figura 1.

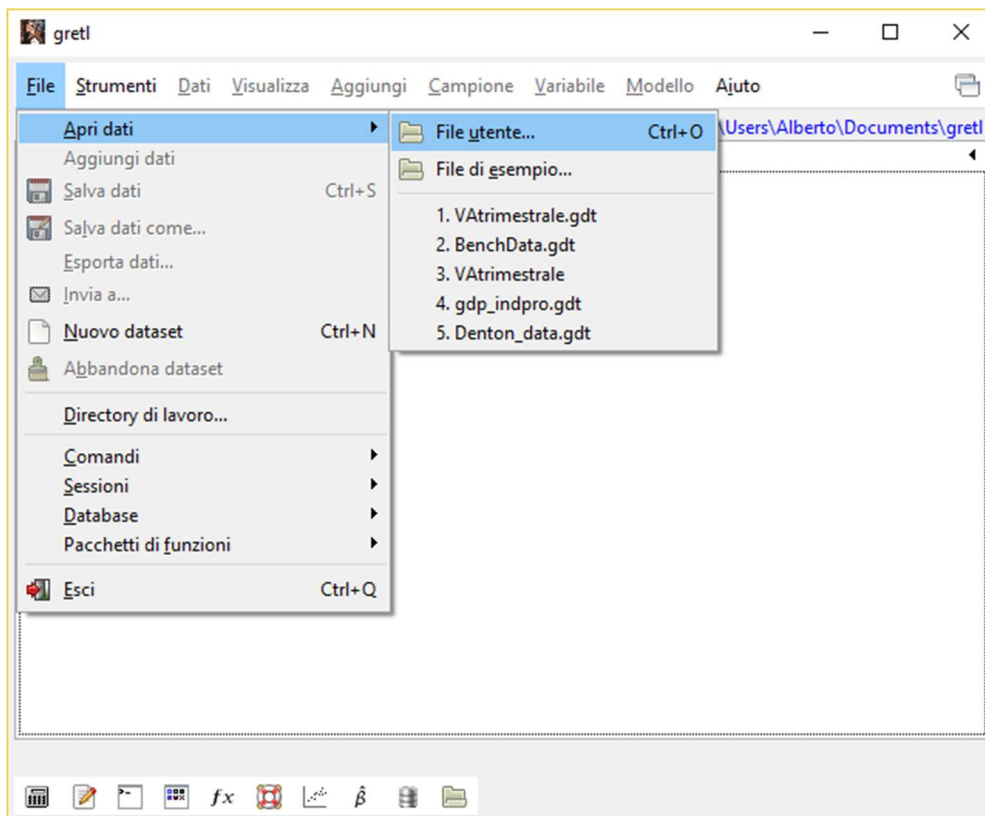


Figura 1

Fatto questo il dataset corrente dovrà essere convertito alla frequenza della serie preliminare come mostrato in figura 2.

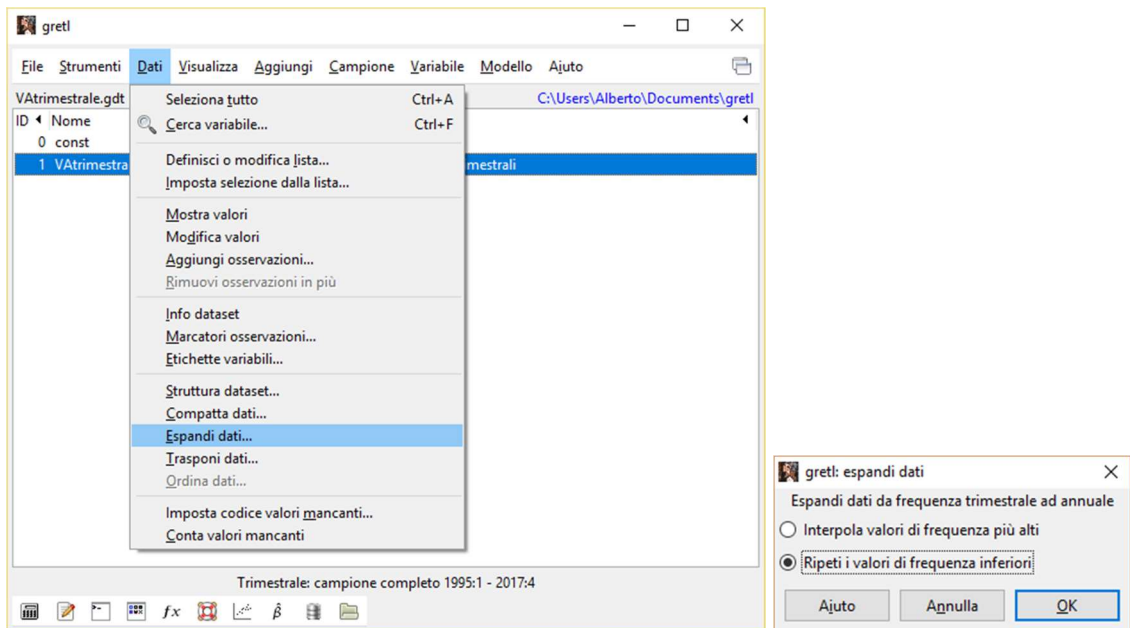


Figura 2

Ora notiamo che la serie preliminare termina ad ottobre del 2018 mentre quella da disaggregare termina a dicembre del 2017. Per poter aggiungere al dataset la serie preliminare avremo quindi bisogno di aggiungere al campione di gretl altre 10 osservazioni, e per fare questo procediamo come in figura 3.

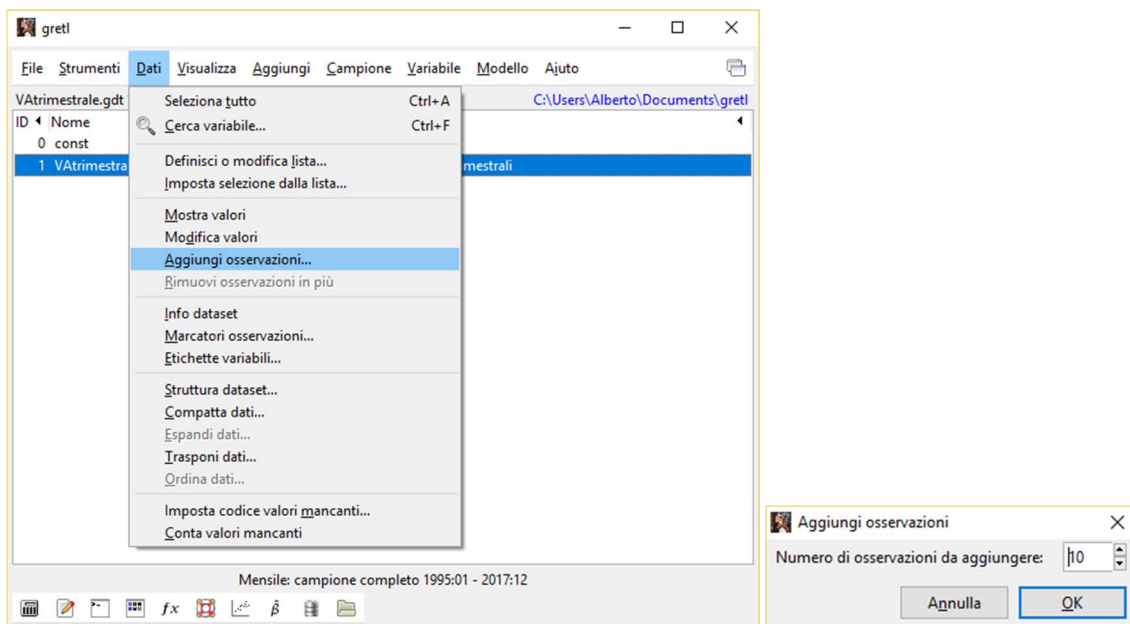


Figura 3

Si potrà a questo punto aggiungere al dataset la serie storica contenuta nel file "indproit.gdt" come mostrato in figura 4.

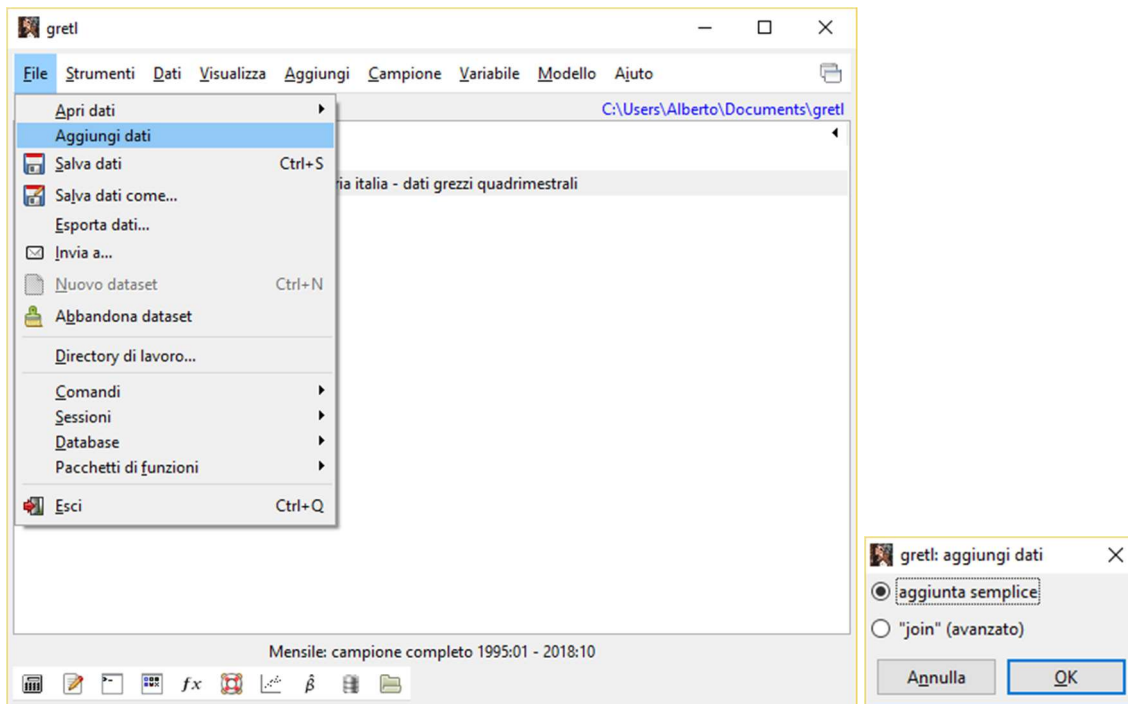


Figura 4

Se non siamo interessati all'extrapolazione, o più in generale abbiamo bisogno di cambiare l'intervallo del campione, possiamo procedere come in figura 5.

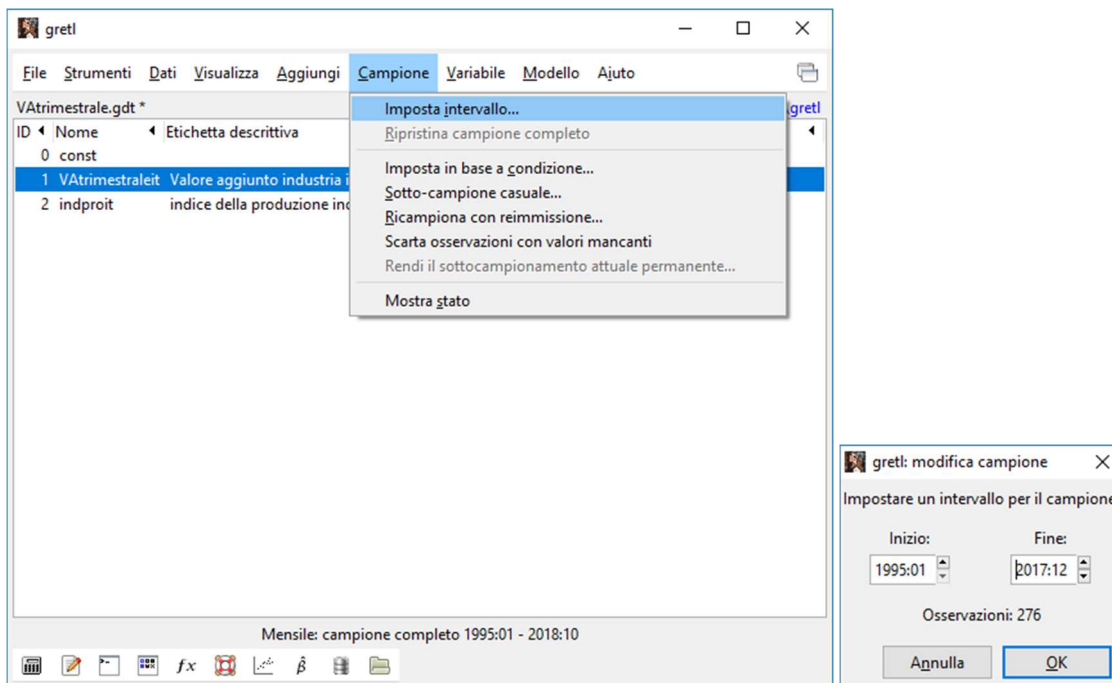


Figura 5

4.3 Utilizzo del pacchetto Benchmarking

Ottenuto un dataset con le caratteristiche precedentemente descritte ed installato il pacchetto *Benchmarking* nella propria versione di Gretl, si potrà iniziare ad utilizzarlo tramite GUI o tramite uno script di hansl.

Se questo pacchetto sarà in futuro caricato sui server di Gretl, per essere utilizzato sulla propria macchina basterà aprire il menu File / Pacchetti di funzioni / Sul Server, dalla finestra che comparirà cercare il pacchetto *Benchmarking*, e, dopo averlo selezionato, fare click con il tasto destro su di esso e scegliere Installa.

Al momento invece, per installare il pacchetto nella propria versione di Gretl bisogna procedere in questo modo:

Se si lavora in Windows si dovrà utilizzare la combinazione di tasti Windows+R, quindi digitare %appdata% e premere invio. Dalla finestra che si aprirà si dovrà entrare nella cartella "gretl" e copiare all'interno della cartella "functions" il file Benchmarking.gfn.

Se si lavora in Mac dal Finder, si sceglierà Vai / Vai alla cartella. Nella finestra di dialogo che comparirà, digitare ~/Library/Application Support e fare click su Vai. Dalla finestra che si aprirà si dovrà entrare nella cartella "gretl" e copiare all'interno della cartella "functions" il file Benchmarking.gfn.

Al successivo avvio di Gretl il pacchetto risulterà installato.

4.3.1 Tramite GUI

Installato il pacchetto nella propria versione di gretl, per utilizzarlo tramite GUI si dovrà, dalla finestra principale, andare su File / Pacchetti di funzioni, selezionare "sul sistema locale" e, dalla schermata che apparirà, fare doppio click sul pacchetto "Benchmarking". Una volta fatto questo, si aprirà una finestra simile a quella mostrata in figura 6.

Come si può vedere dalla figura 1 possiamo specificare le seguenti opzioni dalla finestra GUI della funzione:

- **Series of temporally aggregated constraints** – La serie storica a bassa frequenza che contiene i valori da disaggregare
- **Preliminary series to be benchmarked** – La serie storica ad alta frequenza utilizzata come serie preliminare o indicatore di riferimento
- **Aggregation order** – Il tipo di conversione che viene fatta tra la serie originale a frequenza più bassa e quella a frequenza più alta. È possibile scegliere tra trimestrale/mensile, annuale/trimestrale e annuale/mensile

- **Type of temporal aggregation** – Tipo di aggregazione temporale. I valori della serie a bassa frequenza possono rappresentare la somma o la media dei valori disaggregati oppure il primo o l'ultimo elemento dei sottoperiodi
- **Benchmarking variants** – Tecnica di benchmarking che verrà utilizzata. È possibile scegliere tra naive, pro-rata, le 4 varianti dell'original Denton e del modified Denton, Dagum-Cholette, Chow-Lin (manual) e Chow-Lin (auto). Utilizzando Chow-lin (manual) come valore del parametro di innovazione verrà utilizzato quello passato dall'utente nella finestra, mentre utilizzando Chow-Lin (auto) il valore del parametro di innovazione invece verrà calcolato tramite minimi quadrati pesati.
- **Autoregressive parameter (used only with Dagum-Cholette and Chow-Lin)** – Il valore dato al parametro autoregressivo nel caso si utilizzi la variante di Dagum-Cholette oppure di Chow-Lin (manual)
- **Show text** – se attivo, eseguendo il programma verranno riportate nella console alcune informazioni riguardo le serie analizzate ed eventuali warning se presenti
- **Show graphs** – se attivo, verranno riportati in output alcuni grafici riguardanti le serie trattate (nel caso di serie storiche molto lunghe che coprono numerosi decenni, la creazione dei grafici potrebbe aumentare significativamente il tempo di esecuzione del programma)

Se assegnato un valore di uscita, verrà aggiunta le serie storica calcolata nel database di gretl, in caso contrario saranno visualizzati solamente gli output che l'utente ha indicato tramite "show text" e "show graphs".

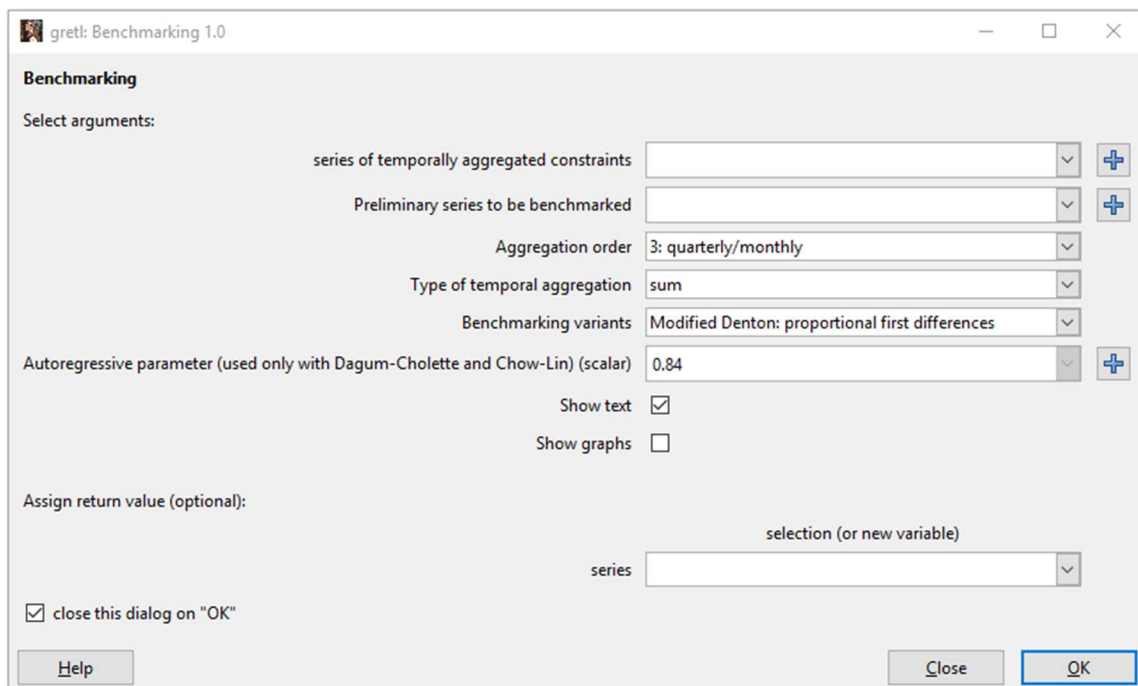


Figura 6: finestra principale del pacchetto Benchmarking per gretl

4.3.2 Tramite Script

Il pacchetto Benchmarking può anche essere usato all'interno di uno script di hansl. Si veda ad esempio il codice seguente:

```
include Benchmarking.gfn
open fedstl.bin
data indpro
data gdpc96
smp1 1947:1 2018:6
Y=Benchmark(gdpc96, indpro, 0, 0, 8, 0.84, 1, 0)
```

In questo esempio la prima riga di comando rende disponibili a gretl le funzioni del pacchetto Benchmarking. Successivamente viene aperto il database **fedstd1**, presente di default nelle distribuzioni di gretl. L'istruzione **data indpro** importa nel dataset la serie storica che utilizzeremo come preliminare e **data gdpc96** quella che contiene i valori aggregati. Si noti che importando per prima la serie preliminare gretl imposta automaticamente il dataset alla frequenza di questa serie: se non viene aggiunta l'opzione **--interpolate** al comando **data**, gretl importa automaticamente le serie di una frequenza più bassa del dataset attuale ripetendone i valori. È questo il caso di **gdpc96** che originariamente sarebbe una serie trimestrale ma viene convertita automaticamente a mensile con i valori trimestrali ripetuti per i tre mesi pertinenti. Tramite **smp1 1947:1 2018:6** ci spostiamo in un intervallo del campione che inizia con la serie da disaggregare e infine aggiungiamo al dataset la serie storica disaggregata **Y** tramite la il comando **Y=Benchmark(gdpc96, indpro, 0, 0, 8, 0.84, 1, 0)**.

I parametri passati alla funzione **Benchmark** indicano che in sostanza **Y** conterrà la stima mensile coerente con la serie trimestrale **gdpc96** e usando **indpro** come serie preliminare. Il terzo argomento pari a 0 indica che come ordine di aggregazione temporale si considera una conversione da trimestrale a mensile, il quarto parametro pari a 0 indica che i valori della serie da disaggregare rappresentano la somma dei valori da stimare, il quinto parametro pari a 8 indica che come tecnica di benchmarking sarà utilizzato il modified Denton con funzione obiettivo basata sulle prime differenze proporzionali, il sesto parametro pari a 0.84 indica il valore dato al parametro autoregressivo nel caso siano utilizzate le tecniche Dagum-Cholette o Chow-Lin(manual) (ma in questo caso un qualsiasi valore tra 0 e 1 non avrebbe portato cambiamenti), il settimo parametro pari a 1 indica che siano riportati nella console gli output di testo ed infine l'ultimo parametro pari a 0 evita che ci sia la produzione di grafici.

```

gretl: risultato dei comandi
gretl versione 2019a
Sessione corrente: 2019-02-21 13:08

? clear
? include Benchmarking.gfn
C:\Users\Alberto\AppData\Roaming\gretl\functions\Benchmarking.gfn
Benchmarking 1.0, 2018-11-08 (Alberto Maronilli)
? open fedstl.bin
C:\Program Files\gretl\db\fedstl.bin
? data indpro
Serie importata con successo
Campione completo dei dati: 1919:01 - 2018:12 (n = 1200)

2 variabili elencate:
  0) const    1) indpro

? data gdpc96
Serie importata con successo
3 variabili elencate:
  0) const    1) indpro    2) gdpc96

? smpl 1947:1 2018:6
Campione completo dei dati: 1919:01 - 2018:12 (n = 1200)
Campione attuale: 1947:01 - 2018:06 (n = 858)

? Y=Benchmark(gdpc96,indpro,0,0,8,0.84,1)
*****
TEMPORAL DISAGGREGATION METHOD: Modified Denton (PFD)
*****
Series of temporally aggregated constraints:    gdpc96
Series used as preliminary:                    indpro
Aggregation order:                             3: quarterly/monthly
Type of aggregation:                            sum
-----
Number of low frequency observations:    282
Number of high frequency observations:    858
Number of extrapolations:                12
-----
Generata la serie Y (ID 3)

```

Figura 7: Risultati dello script di comandi che usa il pacchetto Benchmarking

4.4 Benchmarking al lavoro: stima del valore aggiunto dell'ISS mensile italiano

In questa sezione si illustra come il pacchetto *Benchmarking* per gretl può essere utilizzato per stimare la serie storica mensile del valore aggiunto dell'industria in senso stretto (ISS) italiana. Come serie indicatrice verrà utilizzata la serie storica mensile dell'indice di produzione industriale italiano.

Apriamo il file `benchdata.gdt`, che contiene un dataset di gretl con all'interno le serie storiche di nostro interesse già nella formattazione definita al capitolo 4.2. Impostiamo un range più breve per apprezzare meglio i futuri risultati (in particolare lo impostiamo dal 1995:1 al 2001:12) ed avviamo il pacchetto di funzioni "Benchmarking".

Dalla GUI impostiamo gli argomenti della funzione come in figura 8. Nel nostro caso utilizziamo come serie di benchmark la variabile "VAAnnualeit" e come serie preliminare la variabile "indproit". Come ordine di aggregazione viste le frequenze delle due variabili scegliamo 12: annual/monthly e come tipo di aggregazione temporale scegliamo somma in quanto la variabile aggregata corrisponde alla somma dei valori che vogliamo stimare. In questo caso come metodo di Benchmarking proviamo ad utilizzare la variante originale del Denton, con funzione di penalizzazione basata sulle prime differenze proporzionali. L'argomento "autoregressive parameter" scegliendo questo metodo non verrà utilizzato, per questo lo lasciamo al valore di default. Spuntiamo entrambe le caselle "Show text" e "Show graphs" per visualizzare tutti gli output che può generare il programma e infine assegniamo il valore di ritorno VAmensileit per aggiungere al dataset una nuova variabile contenente la serie storica disaggregata.

Una volta premuto OK si chiuderà la finestra della selezione degli argomenti e si apriranno 6 diverse finestre, una con un output di testo e le altre cinque contenenti i grafici.

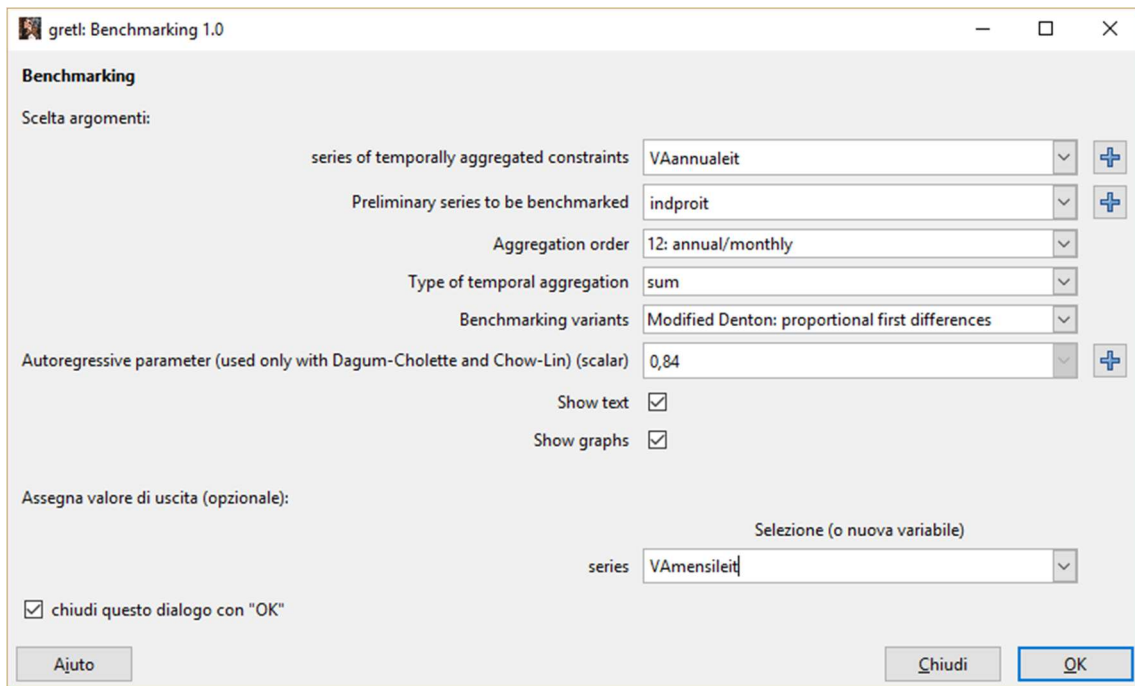


Figura 8: schermata principale del pacchetto benchmarking con gli argomenti scelti per l'esempio

In figura 9 è riportato l'output di testo della funzione di benchmark, qui saranno riportate le opzioni scelte per l'esecuzione del programma, il numero di osservazioni ad alta e bassa frequenza analizzate dalla funzione e il numero di estrapolazioni. Se nell'utilizzo del pacchetto verrà scelta la variante, Cholette-Dagum, Chow-Lin (manual) o Chow-Lin(auto) verrà riportato anche il valore del parametro passato alla funzione o stimato da quest'ultima.

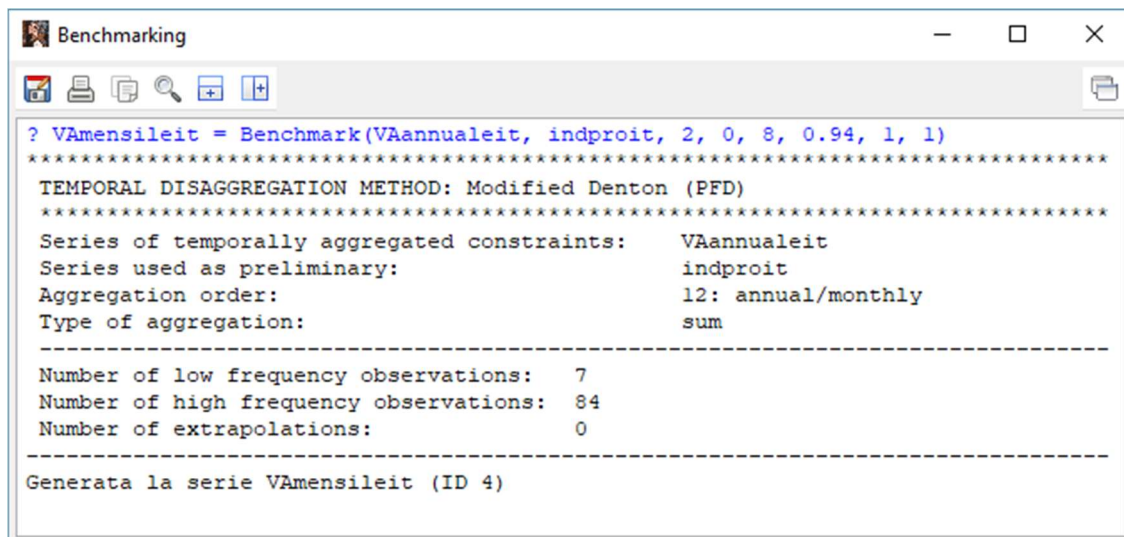


Figura 9: Output su console della funzione Benchmark

In figura 10 è riportato il primo dei grafici generati dal programma. Qui troveremo il grafico della serie storica della variabile originale e della sua stima disaggregata. Se il tipo di aggregazione scelto è uguale a sum i valori della variabile originale saranno riportati nel grafico divisi per l'ordine di aggregazione, negli altri casi troveremo il valore originale.

Nel secondo grafico riportato (Figura 11) troviamo i tassi di crescita della serie stimata disaggregata e della serie preliminare. I tassi di crescita qui sono calcolati come $(y_t - y_{t-1})/y_{t-1}$. In questo caso i tassi di crescita della serie stimata ricalcano quasi perfettamente quelli della serie preliminare.

Il terzo grafico che ci viene presentato (figura 12) è uno scatterplot dei tassi di crescita della serie stimata rispetto a quelli della serie indicatrice. Nel grafico è indicata anche l'equazione della retta di regressione calcolata tramite i minimi quadrati tra le due variabili. Nel titolo il valore di Rho indica la correlazione che sussiste tra le due variabili.

Gli ultimi due grafici (Figura 13 e figura 14) riportano, con la stessa presentazione del secondo e del terzo, i tassi di crescita annuali della serie stimata e della serie indicatrice e lo scatterplot dei tassi di crescita della serie stimata rispetto a quelli della serie preliminare. I tassi di crescita annuali qui sono calcolati come $(y_t - y_{t-k})/y_{t-k}$. Con k uguale a 12 nel caso di benchmarking di serie mensili e a 4 nel caso di benchmarking di serie trimestrali.

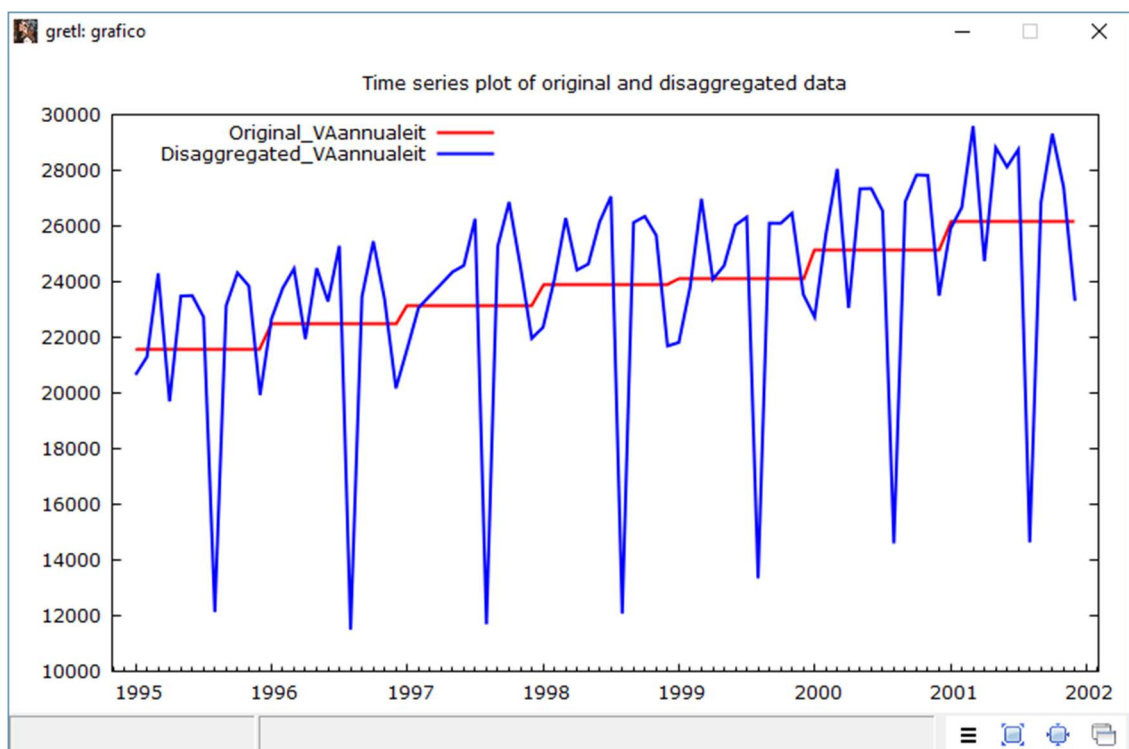


Figura 10: Serie storiche della variabile originale e della sua stima disaggregata

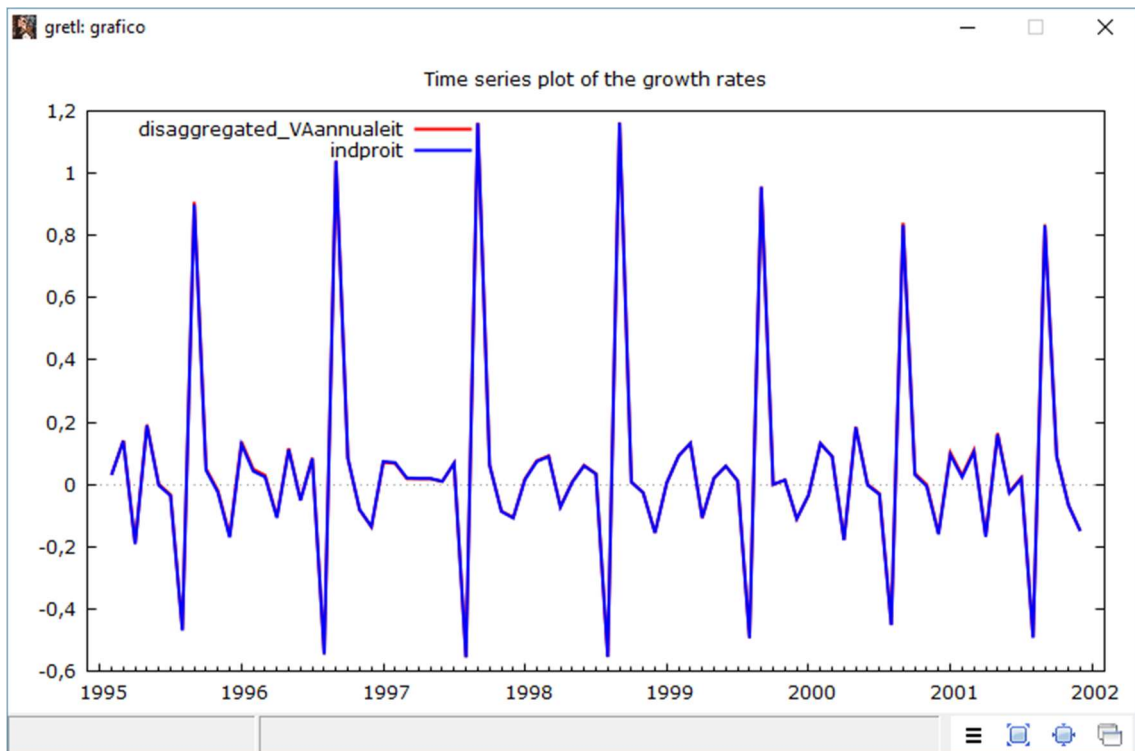


Figura 11: Serie storiche dei tassi di crescita della serie stimata e della serie indicatrice

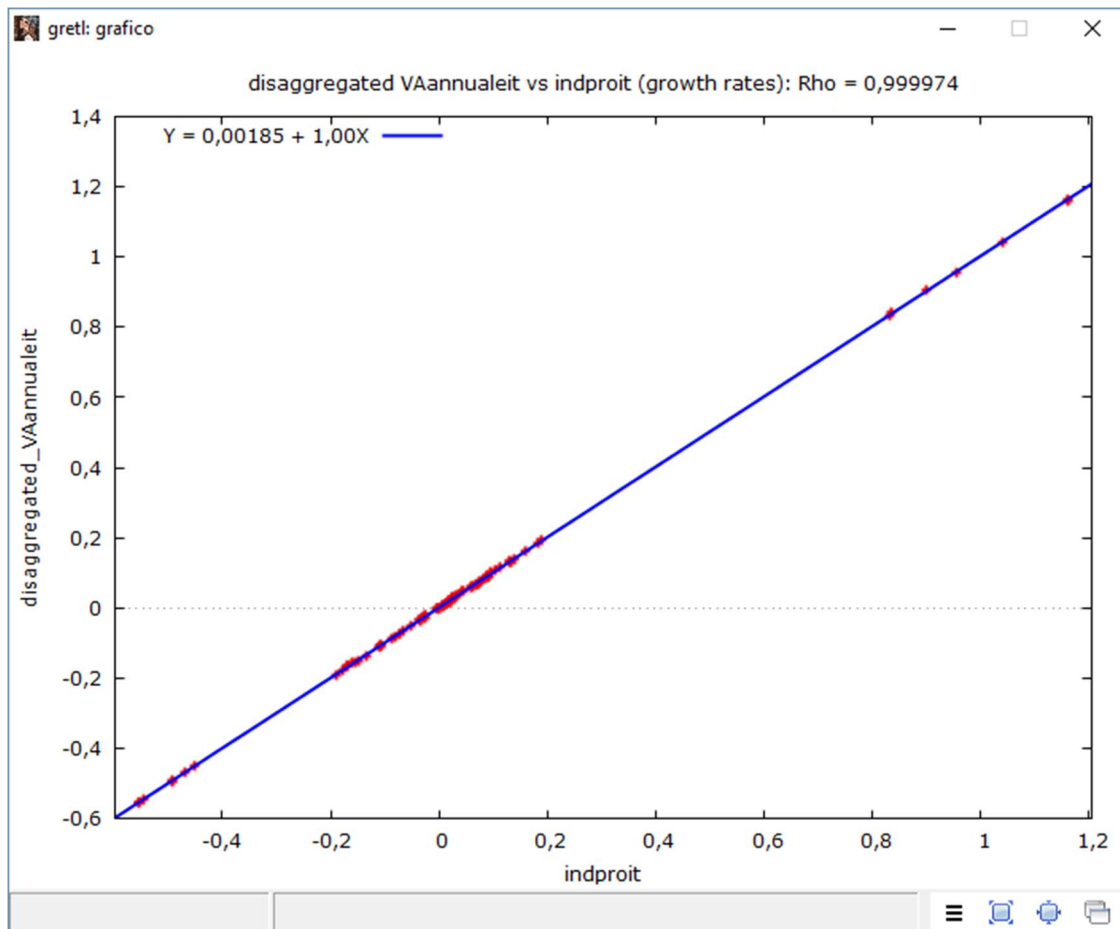


Figura 12: Scatterplot tassi di crescita della serie stimata rispetto a quelli alla serie preliminare

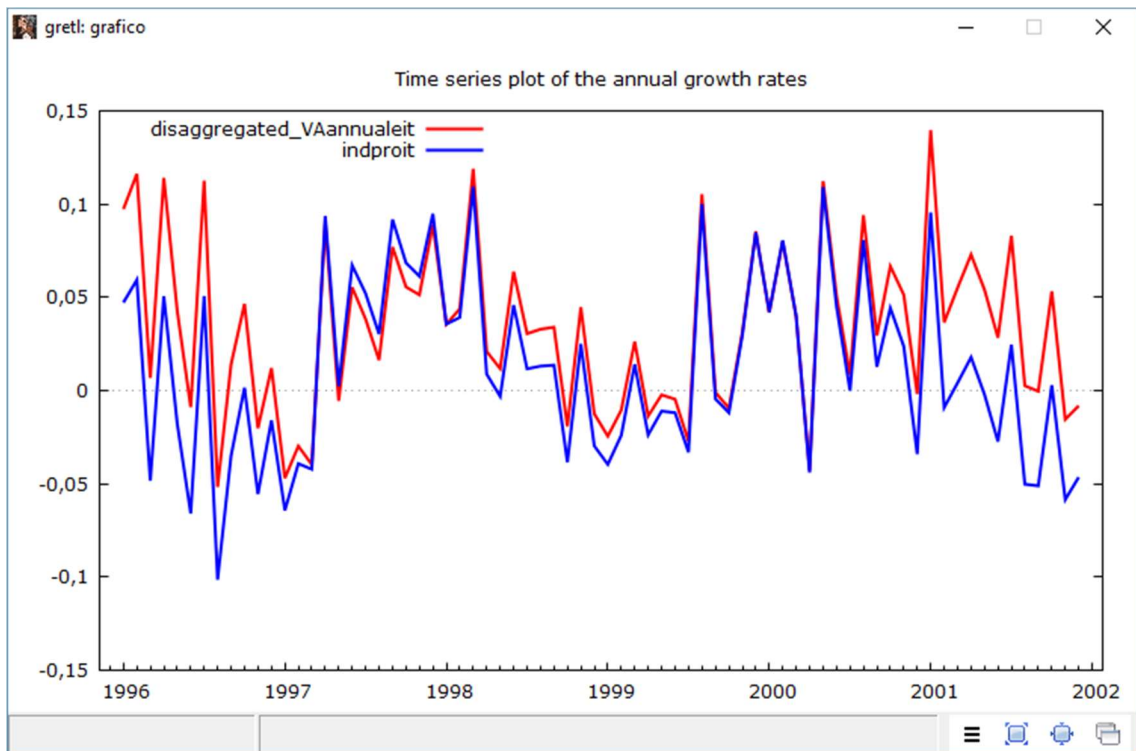


Figura 13: Serie storiche dei tassi di crescita annuali della serie stimata e della serie indicatrice

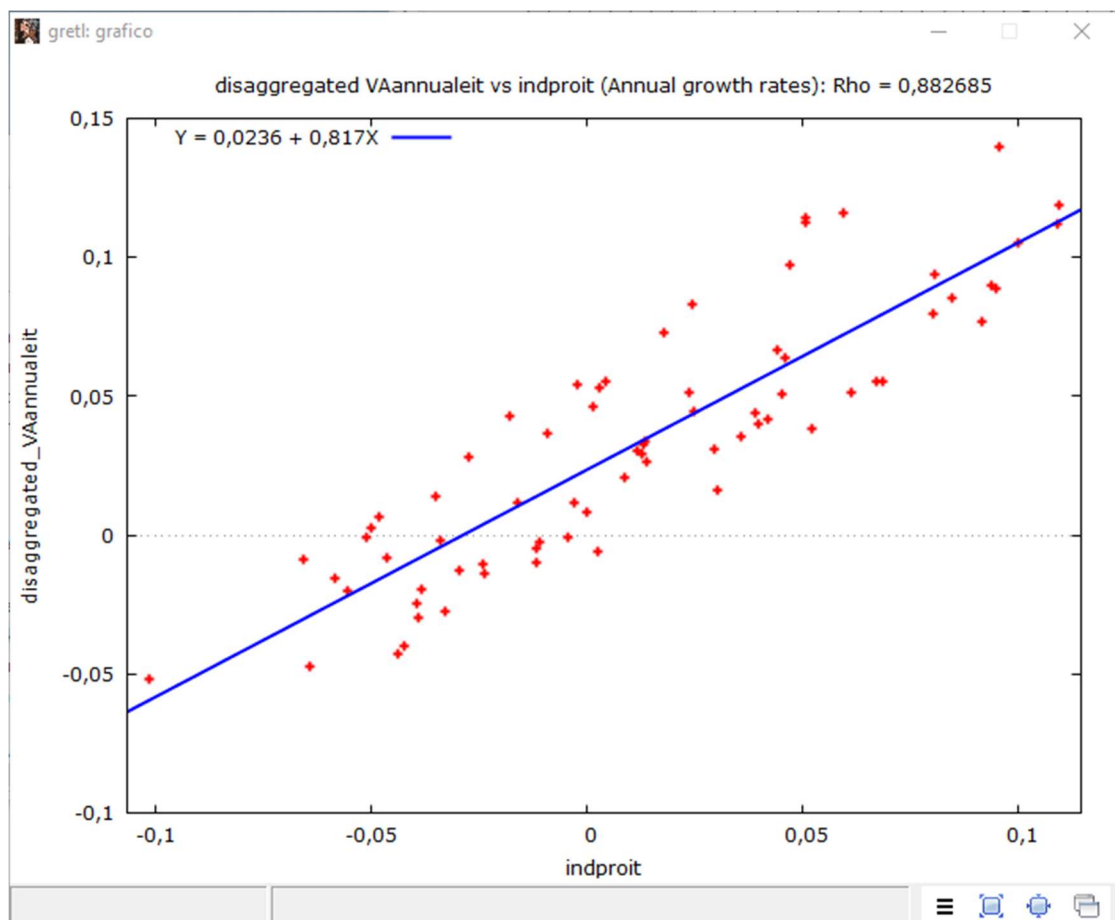


Figura 14: Scatterplot tassi di crescita annuali della serie stimata rispetto a quelli alla serie preliminare

Sono riportati in figura 15 e 16 i grafici della stima di “VAmensileit” selezionando come tecnica di benchmarking Cholette-Dagum con ϕ uguale a 0,94 e Chow-Lin(auto). Nella tabella infine sono riportati i valori delle serie storiche utilizzate e delle stime ottenute.

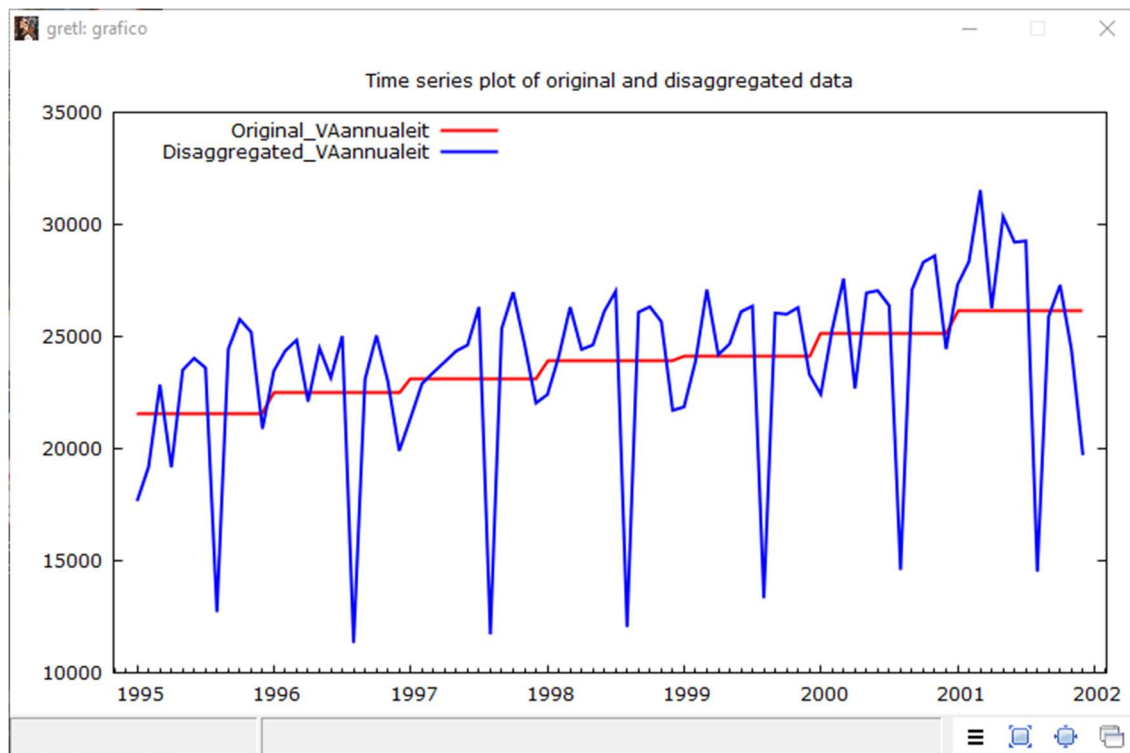


Figura 15: Serie storica stimata di VAannualeit utilizzando Dagum-Cholette come tecnica di benchmarking

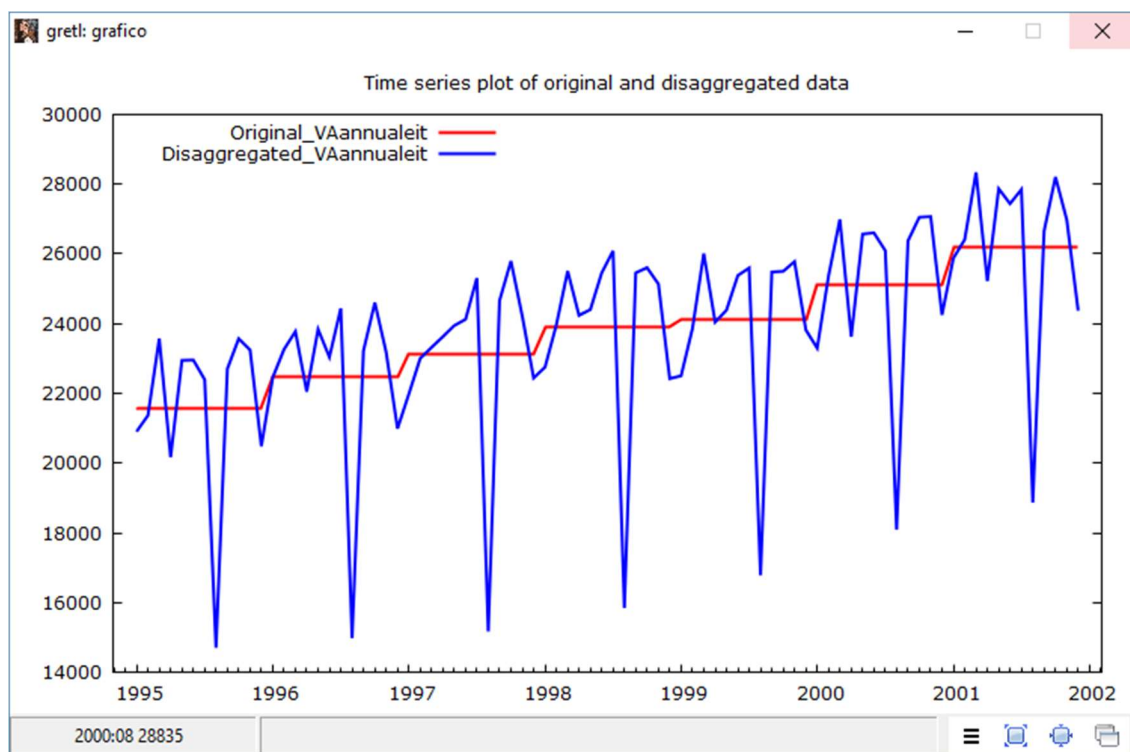


Figura 16: Serie storica stimata di VAannualeit utilizzando Chow-Lin(auto) come tecnica di benchmarking

Tabella 1: Valori della serie di benchmark(VAannualeit), della serie preliminare (indproit) e delle serie stimate utilizzando la tecnica del modified Denton (VAmensileit_MD), di Cholette-Dagum (VAmensileit_CD) e di Chow-Lin (VAmensileit_CL)

obs	VAannualeit	indproit	VAmensileit_MD	VAmensileit_CD	VAmensileit_CL
1995:01	259005,9	114,5	20650,12	17662,90	20904,42
1995:02	259005,9	118,0	21293,92	19180,55	21364,75
1995:03	259005,9	134,5	24300,56	22852,80	23570,00
1995:04	259005,9	108,8	19694,41	19166,24	20168,76
1995:05	259005,9	129,4	23480,61	23506,81	22947,25
1995:06	259005,9	129,1	23498,87	24033,29	22955,94
1995:07	259005,9	124,4	22728,38	23596,44	22392,27
1995:08	259005,9	66,1	12129,55	12711,05	14703,89
1995:09	259005,9	125,5	23137,64	24445,81	22704,16
1995:10	259005,9	131,2	24316,74	25764,54	23566,43
1995:11	259005,9	127,9	23845,73	25196,05	23243,75
1995:12	259005,9	106,2	19929,37	20889,42	20484,26
1996:01	269830,5	119,9	22658,24	23471,74	22452,94
1996:02	269830,5	125,0	23766,81	24363,19	23261,14
1996:03	269830,5	128,0	24464,26	24844,93	23774,54
1996:04	269830,5	114,3	21939,86	22097,36	22048,76
1996:05	269830,5	127,1	24482,13	24486,91	23837,07
1996:06	269830,5	120,6	23290,72	23157,64	23039,96
1996:07	269830,5	130,7	25286,07	25022,00	24438,50
1996:08	269830,5	59,4	11501,98	11338,16	14981,28
1996:09	269830,5	121,1	23460,30	23094,80	23221,62
1996:10	269830,5	131,4	25446,50	25044,24	24601,54
1996:11	269830,5	120,8	23364,25	23010,16	23182,96
1996:12	269830,5	104,5	20169,37	19899,37	20990,19
1997:01	277584,0	112,2	21594,63	21375,79	21978,60
1997:02	277584,0	120,1	23059,21	22899,20	23002,40
1997:03	277584,0	122,6	23492,33	23396,57	23315,79
1997:04	277584,0	125,0	23914,99	23875,94	23624,61
1997:05	277584,0	127,4	24347,23	24354,67	23942,18
1997:06	277584,0	128,7	24579,79	24620,47	24122,01
1997:07	277584,0	137,5	26255,71	26317,95	25309,41
1997:08	277584,0	61,2	11689,88	11714,99	15172,14
1997:09	277584,0	132,2	25265,25	25356,87	24660,71
1997:10	277584,0	140,4	26859,63	26976,64	25794,46
1997:11	277584,0	128,2	24562,99	24662,79	24220,16
1997:12	277584,0	114,4	21962,35	22032,13	22441,53
1998:01	286856,2	116,2	22361,20	22409,95	22749,25
1998:02	286856,2	124,8	24069,69	24107,17	23956,80
1998:03	286856,2	136,0	26283,79	26311,56	25504,83
1998:04	286856,2	126,1	24416,18	24424,80	24237,04
1998:05	286856,2	127,0	24632,38	24626,07	24401,81
1998:06	286856,2	134,6	26146,52	26125,75	25453,10
1998:07	286856,2	139,1	27057,30	27017,21	26085,79
1998:08	286856,2	62,0	12074,17	12044,63	15845,43
1998:09	286856,2	134,0	26124,21	26086,85	25456,12
1998:10	286856,2	135,0	26343,18	26329,04	25605,43
1998:11	286856,2	131,4	25659,53	25664,98	25136,39
1998:12	286856,2	111,0	21688,05	21708,18	22424,21
1999:01	289141,3	111,6	21814,38	21862,63	22503,02
1999:02	289141,3	121,8	23816,63	23899,41	23862,67
1999:03	289141,3	137,9	26972,58	27090,61	26010,39
1999:04	289141,3	123,1	24082,89	24187,99	24045,24
1999:05	289141,3	125,6	24575,54	24672,61	24386,39

1999:06	289141,3	133,0	26025,44	26104,73	25382,45
1999:07	289141,3	134,5	26319,03	26355,79	25595,08
1999:08	289141,3	68,2	13344,46	13330,35	16780,59
1999:09	289141,3	133,4	26099,09	26045,61	25481,31
1999:10	289141,3	133,4	26094,35	25994,61	25501,17
1999:11	289141,3	135,3	26459,25	26290,31	25776,40
1999:12	289141,3	120,4	23537,65	23306,64	23816,59
2000:01	301380,1	116,3	22727,11	22417,69	23297,44
2000:02	301380,1	131,6	25720,35	25314,95	25371,86
2000:03	301380,1	143,4	28046,92	27583,10	26990,09
2000:04	301380,1	117,7	23052,00	22679,57	23624,10
2000:05	301380,1	139,3	27334,36	26947,48	26567,36
2000:06	301380,1	139,0	27344,42	27045,92	26603,98
2000:07	301380,1	134,5	26542,40	26371,67	26091,21
2000:08	301380,1	73,7	14598,42	14589,49	18090,50
2000:09	301380,1	135,1	26868,95	27074,27	26374,09
2000:10	301380,1	139,3	27832,75	28312,67	27049,90
2000:11	301380,1	138,5	27817,72	28600,40	27069,82
2000:12	301380,1	116,3	23494,70	24442,87	24249,75
2001:01	314056,5	127,4	25899,02	27305,90	25874,52
2001:02	314056,5	130,4	26660,66	28331,85	26406,62
2001:03	314056,5	144,0	29593,04	31527,98	28336,50
2001:04	314056,5	119,8	24731,58	26253,83	25218,35
2001:05	314056,5	139,0	28810,99	30338,96	27866,22
2001:06	314056,5	135,2	28120,15	29201,77	27437,12
2001:07	314056,5	137,8	28743,93	29264,58	27846,48
2001:08	314056,5	70,0	14635,48	14514,64	18866,26
2001:09	314056,5	128,2	26858,67	25896,73	26652,57
2001:10	314056,5	139,7	29312,71	27293,02	28205,62
2001:11	314056,5	130,4	27387,76	24416,22	26974,71
2001:12	314056,5	110,9	23302,53	19711,02	24371,54

APPENDICE A

Disaggregazione temporale in Gretl: la funzione **chowlin()**

Come già accennato, nella release attuale di Gretl (2019a), è già presente una funzione dedicata alla disaggregazione temporale cioè `chowlin()`.

Questa funzione permette di espandere i dati di input ad una frequenza più alta, usando la tecnica di interpolazione di Chow-Lin (1971).

Gli argomenti della funzione sono una matrice Y che contiene le serie storiche che si vogliono disaggregare, il valore intero `xfac` che indica l'ordine di aggregazione (i cui unici valori validi sono 3 e 4 per conversioni rispettivamente di tipo trimestrale/mensile e annuale/trimestrale), e un'altra matrice (opzionale) X che contiene eventuali indicatori di riferimento da utilizzare per la disaggregazione che deve avere tante colonne quante la matrice Y e `xfac` volte il numero di righe della matrice Y .

I regressori utilizzati di default per la disaggregazione temporale dalla funzione `chowlin` sono una costante ed un trend quadratico. Se alla funzione viene passata anche la matrice X , le colonne di questa matrice saranno utilizzate come ulteriori regressori.

Una debolezza di questo modo di procedere (oltre a non permettere estrapolazioni, tipi di aggregazioni temporali diverse dalla media dei valori o disaggregazioni da dati annuali a mensili) sta nella scelta di utilizzare sempre come regressore il trend quadratico. Nella maggior parte dei casi questa decisione risulta ingiustificata e finisce solamente per penalizzare la parsimonia del modello.

Una funzione analoga a `chowlin` viene eseguita dal software Gretl quando si sceglie di espandere un dataset o come mostrato in figura 2 al capitolo 4.2.1 ma scegliendo "Interpola i valori di frequenza più alti" oppure dalla console di gretl con il comando `dataset expand interp`.

Prendendo come esempio la serie storica che utilizza Denton (1971) nei suoi esempi illustrativi (senza utilizzare la serie preliminare) possiamo vedere come la procedura utilizzata da Gretl, nel caso di disaggregazione temporale senza indicatori di riferimento, produca dei risultati difficilmente interpretabili (Figura 17).

È stato inoltre appurato che la ricerca del valore ottimale del parametro di autoregressione ρ viene fatta in un intervallo che comprende anche valori negativi, in particolare la serie storica della figura 17 è ottenuta con un valore di ρ pari a circa -0,68. Quando il valore di ρ risulta negativo, la componente erratica del modello di regressione assunto domina i movimenti a breve termine della serie risultante.

Un'alternativa più adatta per attuare questa procedura può essere ottenuta utilizzando come regressori, come consigliato anche da Quilis E. M. (2018), una costante ed una variabile indice lineare.

Per fare questo in gretl, può essere utilizzato il pacchetto benchmarking scegliendo come serie preliminare una variabile indice, "average" come tipo di aggregazione temporale e selezionando come tecnica di Benchmarking "Chow-Lin(auto)" (Figura 18).

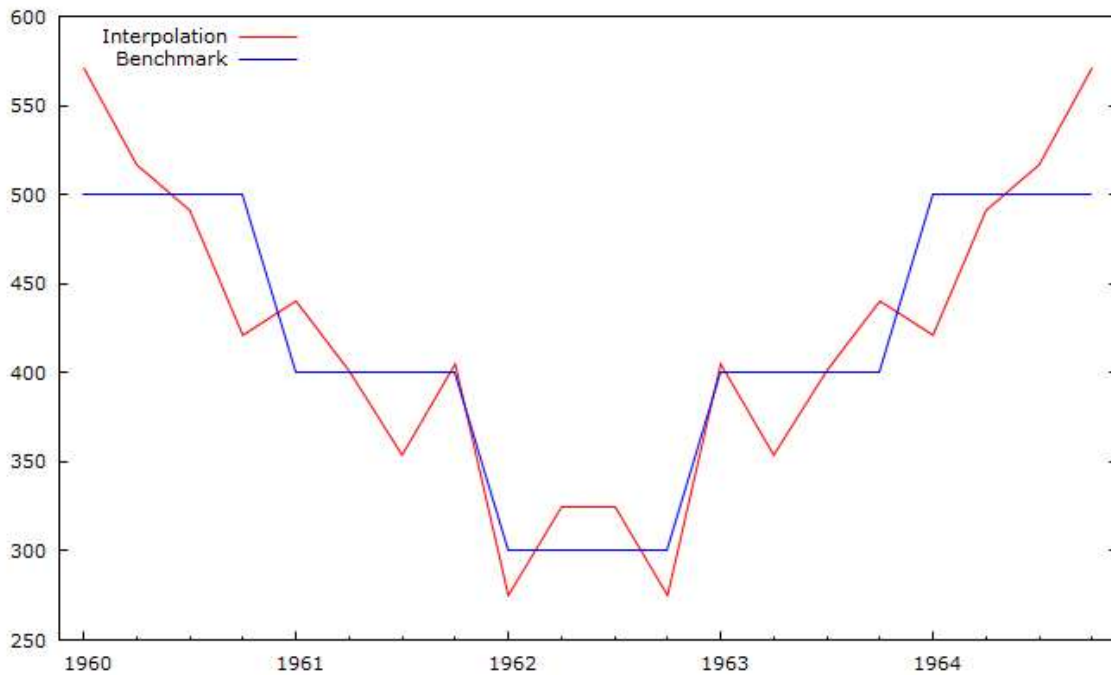


Figura 17: Risultato dell'interpolazione usando la procedura nativa di gretl

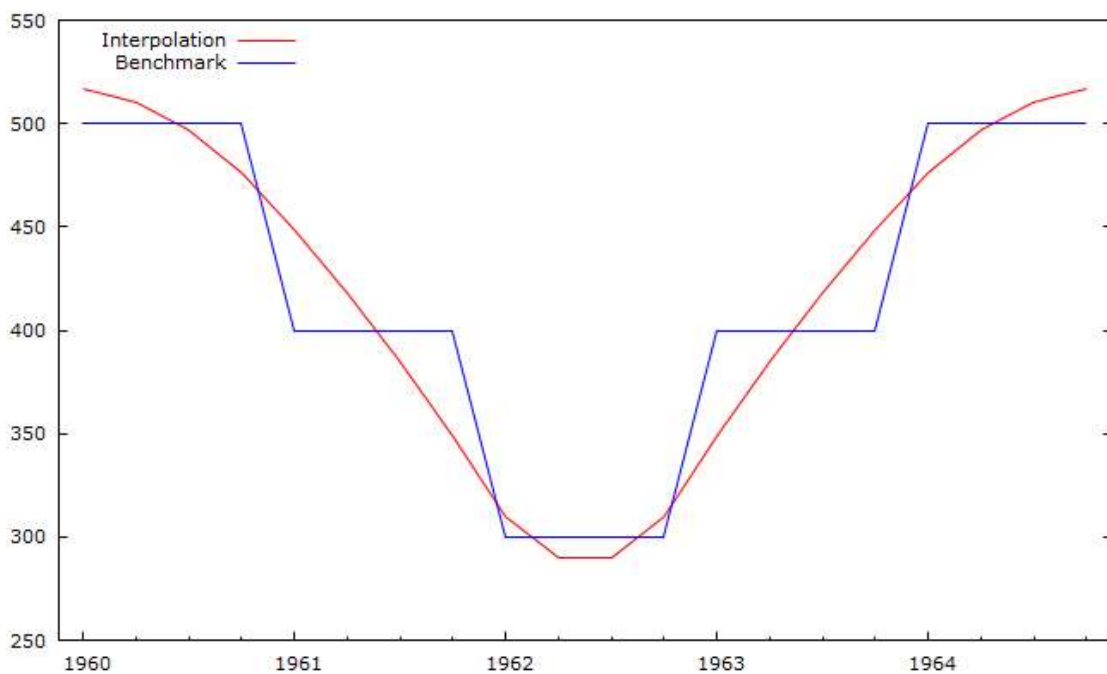


Figura 18: Risultato dell'interpolazione utilizzando il pacchetto Benchmarking

APPENDICE B

Il codice del pacchetto *Benchmarking*

```
# private functions

### funzione dedicata alla conversione di vettori colonna in matrici diagonali
function matrix DiagVector (matrix y)
  scalar n=rows(y)
  matrix X=zeros(n,n)
  loop i=1..n --quiet
    X[i,i]=y[i,1]
  endloop
  return X
end function

### funzione dedicata alla creazione dei grafici relativi ai tassi di crescita
annuali
function void expanded_output (series bench,
                              series high,
                              int k,
                              string slow,
                              string shigh)

  series high_gr
  series bench_gr
  if k==3
    k=12
  endif
  loop i=($t1+k)..$nobs --quiet
    high_gr[i]=(high[i]-high[i-k])/high[i-k]
    bench_gr[i]=(bench[i]-bench[i-k])/bench[i-k]
  endloop
  list stuff = null
  stuff += genseries(sprintf("disaggregated_%s", slow),bench_gr)
  stuff += genseries(sprintf("%s", shigh),high_gr)
  plot stuff
    options --with-lines --time-series --single-yaxis
    literal set title "Time series plot of the annual growth rates"
  end plot --output=display
  scalar c = corr(stuff[1],stuff[2])
  string stitle=sprintf("disaggregated %s vs %s (Annual growth rates): Rho = %g",
slow,shigh,c)
  scatterplot(stuff,stitle)
end function

### funzione utilizzata da expanded_output ed expanded_output2 per la creazione degli
scatterplot
function void scatterplot (list l,
                          string title)

  plot l
    printf "set title %s", title
  end plot --output=display
end function

### funzione dedicata alla creazione dei grafici relativi ai tassi di crescita
function void expanded_output2 (series bench,
                              series high,
                              int k,
                              string slow,
                              string shigh)

  series high_gr
  series bench_gr
  loop i=($t1+1)..$nobs --quiet
```

```

        high_gr[i]=(high[i]-high[i-1])/high[i-1]
        bench_gr[i]=(bench[i]-bench[i-1])/bench[i-1]
    endloop
    list stuff = null
    stuff += genseries(sprintf("disaggregated_%s", slow),bench_gr)
    stuff += genseries(sprintf("%s", shigh),high_gr)
    plot stuff
        options --with-lines --time-series --single-yaxis
        literal set title "Time series plot of the growth rates"
    end plot --output=display
    scalar c = corr(stuff[1],stuff[2])
    string stitle=sprintf("disaggregated %s vs %s (growth rates): Rho = %g",
slow,shigh,c)
    scatterplot(stuff,stitle)
end function

### funzione dedicata all'individuazione del parametro autoregressivo ottimale nel
caso che
### come tecnica di benchmarking venga scelto "Chow-Lin(auto)"
function scalar CL_criterion (scalar n,
                             matrix J,
                             matrix CL,
                             matrix y)
    matrix r=(0.05+(0.94/49)*seq(0,49))'
    matrix val=zeros(50,1)
    loop h=1..50 --quiet
        matrix Aux = zeros(1,n)|I(n)
        Aux = (I(n)-Aux[1:n,])*r[h,1])
        Aux[1,1] = sqrt(1-r[h,1]^2)
        matrix w = inv(Aux'*Aux)
        matrix W = J*w*J'
        matrix Wi = inv(W)
        matrix beta = (CL'*Wi*CL)CL'*Wi*y)
        matrix U = y - CL*beta
        val[h,1]= -U'*Wi*U
    endloop
    scalar rho=r[imaxc(val),1]
    return rho
end function

# public functions

### Funzione "principale" per il funzionamento del pacchetto
function series Benchmark (series low "series of temporally aggregated constraints",
                          series high "Preliminary series to be benchmarked",
                          int k[0:2:0] "Aggregation order" "3: quarterly/monthly",
                          "4: annual/quarterly", "12: annual/monthly",
                          int j[0:3:0] "Type of temporal aggregation" "sum",
                          "average", "sample of the last", "sample of the first",
                          int t[0:12:8] "Benchmarking variants" "Naive
benchmarking", "Pro-rata benchmarking", "Original Denton: additive first
differences", "Original Denton: additive second differences", "Original Denton:
proportional first differences", "Original Denton: proportional second differences",
"Modified Denton: additive first differences", "Modified Denton: additive second
differences", "Modified Denton: proportional first differences", "Modified Denton:
proportional second differences", "Dagum-Cholette", "Chow-Lin (manual)", "Chow-Lin
(auto)",
                          scalar phi[0:0.999:0.84] "Autoregressive parameter (used
only with Dagum-Cholette and Chow-Lin)",
                          bool st[1] "Show text",
                          bool sg[0] "Show graphs")

    # Controllo degli input
    if missing(low[$t1]) || missing(high[$t1])
        funcerr "Series used in the benchmark must start with the current sample"
    endif
    if missing(high[$t2])

```

```

        funcerr "The preliminary series must end with the current sample"
    endif
    if $pd==4 && k!=1
        funcerr "You're asking for a montly disaggregation in a quarterly sample,
check your aggregation order"
    elif $pd==12 && k==1
        funcerr "You're asking for a quarterly disaggregation in a monthly sample,
check your aggregation order"
    endif
    if $obsminor[$t1] !=1
        if k==0 && $obsminor[$t1] != 4 && $obsminor[$t1] != 7 && $obsminor[$t1] != 10
            printf("!be sure that your current sample range starts with the first
month of a quarter")
        elif k==1
            printf("!be sure that your current sample range starts with the first
quarter of a year")
        elif k==2
            printf("!be sure that your current sample range starts with the first
month of a year")
        endif
    endif

    # Inizializzazione delle variabili
    matrix p = high
    matrix y = low
    scalar n = rows(p)
    if k==0
        k=3
        string ao = "3: quarterly/monthly"
    elif k==1
        k=4
        string ao = "4: annual/quarterly"
    elif k==2
        k=12
        string ao = "12: annual/monthly"
    else
        funcerr "Invalid aggregation order"
    endif
    string slow=argname(low)
    string shigh=argname(high)
    if slow == "Benchmark" || shigh == "Benchmark"
        printf("\nWarning!\n Having a variable called Benchmark may generate
errors.")
    endif
    string method
    # Conversione di y da vettore con i valori ripetuti a normale vettore
    scalar m = rows(y)/k
    if m % 1 != 0
        if st
            printf("!all preliminary data are aviable for the last temporally
aggregated constraintlast temporally aggregated constraint will not be used in the
benchmark")
        endif
        m=floor(m)
        matrix boh = I(m)**(ones(1,1)~zeros(1,(k-1)))~zeros(m,(rows(y)-m*k))
    else
        matrix boh = I(m)**(ones(1,1)~zeros(1,(k-1)))
    endif
    y=boh*y
    # Creazione della matrice di aggregazione
    if j==0
        matrix J = I(m)**ones(1,k)
        string at = "sum"
    elif j==1
        matrix J = I(m)**(ones(1,k)/k)
        string at = "average"
    elif j==2

```

```

        matrix J = I(m)**(zeros(1,k-1)~1)
        string at = "sample of the last"
    elif j==3
        matrix J = I(m)**(1~zeros(1,k-1))
        string at = "sample of the first"
    else
        funcerr "invalid j value"
    endif
#Forecasting check
scalar forecast = 0
if n > m*k
    if t==0 || t==1
        printf(": forecasting not supported with naive and pro-rata
benchmarking")
    endif
    forecast = n - m*k
    J = J ~ zeros(m,forecast)
endif

# Applicazione delle tecniche di benchmarking
if t>=2 && t<=5
    matrix D = zeros(1,n)|I(n)
    D = (I(n)-D[1:n,])
    if t==2
        method = "Original Denton (additive first differences)"
        matrix M = D'*D
    elif t==3
        method = "Original Denton: (additive second differences)"
        matrix M = D'*D'*D*D
    elif t==4
        method = "Original Denton (proportional first differences)"
        matrix P = DiagVector(p)
        matrix M = inv(P)*D'*D*inv(P)
    elif t==5
        method = "Original Denton: (proportional second differences)"
        matrix P = DiagVector(p)
        matrix M = inv(P)*D'*D'*D*D*inv(P)
    endif
    matrix S = p+inv(M)*J'*inv(J*inv(M)*J')*(y-J*p)
    series bench = S[,1]
elif t==10
    method = "Dagum-Cholette"
    matrix P = DiagVector(p)
    matrix AR = I(n)-((zeros(1,(n-1)) | (I(n-1)*phi)) ~ zeros(n,1))
    AR[1,1] = sqrt(1-phi^2)
    matrix Ve = P*((AR'*AR))
    syssol1 = (J*Ve*J')y-J*p)
    matrix S = p + Ve*J'*syssol1
    series bench = S[,1]
elif t==11
    method = "Chow-Lin (manual)"
    matrix c1 = ones(n,1)~p
    matrix CL=J*c1
    matrix Aux = zeros(1,n)|I(n)
    Aux = (I(n)-Aux[1:n,]*phi)
    Aux[1,1] = sqrt(1-phi^2)
    matrix w = inv(Aux'*Aux)
    matrix W = J*w*J'
    matrix Wi = inv(W)
    matrix beta = (CL'*Wi*CL)CL'*Wi*y)
    matrix U = y - CL*beta
    matrix L = w*J'*Wi
    matrix u = L*U
    series bench = c1*beta + u
elif t==12
    method = "Chow-Lin (auto)"
    matrix c1 = ones(n,1)~p

```



```

matrix CL=J*c1
/*      matrix u
matrix b = mols(y,CL,&u)
scalar ac1 = corrgm(u,1)[1]
matrix ma = .5
scalar crit = BFGSmax(ma, fchowlin(&ma, ac1)) */
scalar ma=CL_criterion(n,J,CL,y)
matrix Aux = zeros(1,n)|I(n)
Aux = (I(n)-Aux[1:n,]*ma)
Aux[1,1] = sqrt(1-ma^2)
matrix w = inv(Aux'*Aux)
matrix W = J*w*J'
matrix Wi = inv(W)
matrix beta = (CL'*Wi*CL)CL'*Wi*y)
matrix U = y - CL*beta
matrix L = w*J'*Wi
matrix u = L*U
series bench = c1*beta + u
else
if t==0
    method = "Naive benchmarking"
    matrix M=I(n)
elif t==1
    method = "Pro-rata benchmarking"
    matrix M=DiagVector(p)
elif t==6
    method = "Modified Denton (Additive First Differences)"
    matrix D1 = zeros(1,n)|I(n)
    D1 = (I(n)-D1[1:n,])[2:n,]
    matrix M=D1'*D1
elif t==7
    method = "Modified Denton (additive second differences)"
    matrix D1 = zeros(1,n)|I(n)
    D1 = (I(n)-D1[1:n,])[2:n,]
    matrix D2=zeros(1,n-1)|I(n-1)
    D2=(I(n-1)-D2[1:n-1,])[2:n-1,]
    matrix M=D1'*D2'*D2*D1
elif t==8
    method = "Modified Denton (proportional first differences)"
    matrix D1 = zeros(1,n)|I(n)
    D1 = (I(n)-D1[1:n,])[2:n,]
    matrix P = DiagVector(p)
    matrix M=inv(P)*D1'*D1*inv(P)
elif t==9
    method = "Modified Denton (proportional second differences)"
    matrix D1 = zeros(1,n)|I(n)
    D1 = (I(n)-D1[1:n,])[2:n,]
    matrix D2=zeros(1,n-1)|I(n-1)
    D2=(I(n-1)-D2[1:n-1,])[2:n-1,]
    matrix P = DiagVector(p)
    matrix M=inv(P)*D1'*D2'*D2*D1*inv(P)
endif
matrix S=inv(((M~J')|(J~zeros(m,m))))*(M*p|y)
series bench=S[1:n,]
endif

# Produzione dei grafici se richiesti
if sg
    series or_low
    scalar i
    if j==0
        or_low=low/k
    else
        or_low=low
    endif
    list splot=null
    splot+=genseries(sprintf("Original_%s", slow), or_low)

```

```

    splot+=genseries(sprintf("Disaggregated_%s", slow), bench)
    #splot+=genseries(sprintf("Preliminary_series_%s", shigh), high)
    plot splot
        options --time-series --with-lines
        literal set title "Time series plot of original and disaggregated data"
    end plot --output=display
endif

# Produzione dell'output di testo se richiesto
if st
    printf("Series of temporally aggregated constraints: %s", slow)
    printf("Series used as preliminary: %s", shigh)
    printf("Aggregation order: %s", ao)
    printf("Type of aggregation: %s", at)
    printf("Benchmarking method used to disaggregate %s:%s",slow,method)
    if t==10
        printf("Value of the autoregressive parameter used: %g", phi)
    elif t==11
        printf("Value of the innovational parameter used: %g", phi)
    elif t==12
        printf("Estimation of the innovational parameter: %g", ma)
    endif
    printf("Number of low frequency observations: %g", m)
    printf("Number of high frequency observations: %g", n)
endif
if sg
    expanded_output(bench,high,k,slow,shigh)
    expanded_output2(bench,high,k,slow,shigh)
endif
setinfo bench --description=""
return bench
end function

```

Riferimenti Bibliografici

- [1] Baiocchi G. e Distaso W. (2003), **GRETL: Econometric software for the GNU generation**, *Journal of Applied Econometrics*. 18: 105–110
- [2] Cholette P.A. (1984), **Adjusting sub-annual series to yearly benchmarks**, *Survey Methodology*. 10(1): 35-49.
- [3] Cholette P.A. e Dagum E.B. (1994), **Benchmarking time series with autocorrelated survey errors**, *International Statistical Review* 62(3): 365-377.
- [4] Chow G. C. and Lin A. (1971), **Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series**, *The Review of Economics and Statistics* Vol. 53, No. 4, pp. 372-375
- [5] Cottrell A. e Lucchetti R. (2018), **Gretl user's guide**
- [6] Cottrell A. e Lucchetti R. (2018), **A Hansl Primer**
- [7] Cottrell A. e Lucchetti R. (2018), **Gretl Command Reference**
- [8] Denton F.T. (1971), **Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization**, *Journal of the American Statistical Association*, 66(333): 99-102.
- [9] Di Fonzo T. (1987), **La stima indiretta di serie economiche trimestrali**, cleanup editore
- [10] Di Fonzo T. (2003), **Benchmarking di serie storiche economiche. Nota tecnica ed estensioni**, *Dipartimento di scienze statistiche, Università di Padova*, Working paper n. 2003.10
- [11] Dougherty C. (2016), **Introduction to Econometrics**, Oxford University Press
- [12] Eurostat (2018), **ESS guidelines on temporal disaggregation, benchmarking and reconciliation**, Publications Office of European Union
- [13] Chow G. C. e Lin A. (1971), **Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series**, *The Review of Economics and Statistics* Vol. 53, No. 4, pp. 372-375
- [14] Ghysels E. e Qian H. (2016), **Estimating MIDAS Regressions via OLS with Polynomial Parameter Profiling**. <https://ssrn.com/abstract=2837798>
- [15] Helmert, F. R. (1880). **Die mathematischen und physikalischen Theorien der höheren Geodäsie, 1. Teil**. Leipzig.

- [16] International monetary fund. Statistics department (2017), **Quarterly national accounts manual 2017 edition**, *International monetary fund*: 86-125
- [17] Kufel T. (2011), **Ekonometria Rozwiązywanie problemów z wykorzystaniem programu GRETL**, Wydawnictwo Naukowe PWN
- [18] Marini M. (2016), **Nowcasting Annual National Accounts with Quarterly Indicators: An Assessment of Widely Used Benchmarking Methods**, *International monetary fund*, Working Paper n.16/71: 8-9
- [19] Mixon Jr J. W. e Smith R. J. (2006) **Teaching undergraduate econometrics with GRETL**. *Journal of Applied Econometrics*. 21 (7): 1103–1107
- [20] Quilis E. M. (2018), **Temporal disaggregation of time series: a matlab library**
- [21] Rosenblad A. (2008). "**gretl 1.7.3**". *Journal of Statistical Software*. 25 (1): 1–14