



Università Politecnica delle Marche

FACOLTÀ DI ECONOMIA "GIORGIO FUÀ"
Corso di Laurea Magistrale in Scienze Economiche e Finanziarie

Modelli VAR a frequenze miste

Mixed sampling frequency VAR models

Relatore:
Prof. Giulio Palomba

Tesi di Laurea di:
Mario Marchetti

Indice

| | | |
|----------|---|----------|
| 1 | Introduzione | 1 |
| 2 | La letteratura | 4 |
| 2.1 | Il filtro di Kalman | 5 |
| 2.1.1 | La rappresentazione spazio-stato | 5 |
| 2.1.2 | Derivazione del filtro di Kalman | 7 |
| 2.1.3 | Stima di massima verosimiglianza dei parametri . . . | 13 |
| 2.1.4 | Lo smoothing | 13 |
| 2.2 | Il filtro di Hodrick-Prescott | 15 |
| 2.2.1 | Confronto tra Hodrick-Prescott e filtro di Kalman . . | 16 |
| 2.3 | La scomposizione di Beveridge-Nelson | 17 |
| 2.3.1 | Scomposizione di un processo ARMA integrato . . . | 18 |
| 2.3.2 | Un approccio spazio-stato | 24 |
| 2.4 | L'aggregazione temporale | 26 |
| 2.5 | Mixed Data Sampling | 29 |
| 2.5.1 | AR-MIDAS | 33 |
| 2.5.2 | Modello Unrestricted MIDAS | 35 |

INDICE

| | | |
|----------|--|------------|
| 2.5.3 | Variabili esplicative multiple | 37 |
| 2.5.4 | Modelli MIDAS multivariati | 38 |
| 2.6 | Mixed Frequency Vector AutoRegressive | 39 |
| 2.6.1 | Specificazione di Mariano e Murasawa | 40 |
| 2.6.2 | MIDAS Vector AutoRegressive | 51 |
| 2.6.3 | VAR strutturali a frequenze miste | 55 |
| 2.7 | Un confronto tra i diversi metodi | 58 |
| 3 | Il modello econometrico | 61 |
| 3.1 | La specificazione del modello | 64 |
| 3.2 | La stima | 66 |
| 3.2.1 | Misure di bontà previsiva | 72 |
| 3.2.2 | Impulse Response Function | 75 |
| 3.3 | Simulazioni Monte Carlo | 78 |
| 3.4 | Un'applicazione empirica | 81 |
| 3.5 | Disaggregazione temporale e benchmarking | 89 |
| 4 | Conclusioni | 92 |
| A | Codici gretl | 94 |
| | Bibliografia | 107 |

Capitolo 1

Introduzione

Gli economisti spesso si pongono l'obiettivo di spiegare il comportamento di determinate variabili in modo da consentire ai *policy maker* di manovrarle in base alle esigenze della collettività. Molto spesso però le istituzioni di politica economica si trovano a dover prendere decisioni con una frequenza che non necessariamente è quella con cui i dati vengono pubblicati. Anche se una parte non trascurabile delle informazioni economiche è disponibile alla bassa frequenza (cadenza annuale, trimestrale o al massimo mensile), prendere decisioni per tempi più stretti rispetto alla disponibilità dei dati è un approccio usuale.

L'esempio più classico che si possa fare è quello dell'analisi della cosiddetta curva di Phillips. Questa comprende lo studio di variabili come l'inflazione e la disoccupazione che sono pubblicate con cadenza mensile, mentre l'altra variabile considerata è il prodotto interno lordo, che invece è pubblicata con frequenza trimestrale o annuale.

Lavorare con serie storiche economiche campionate a frequenze diver-

se, è dunque una delle situazioni più comuni nella quale si trovano gli economisti.

Inoltre i recenti sviluppi delle tecnologie e il crescente interesse per gli effetti dei co-movimenti tra i mercati finanziari, le economie nazionali e i processi decisionali degli agenti economici forniscono una grande quantità di informazioni disponibili per analisi empiriche.

In questo contesto l'approccio classico di aggregare le variabili ad alta frequenza finché tutti i dati non presentano la stessa cadenza, può avere un costo notevole in termini di informazioni non considerate. In aggiunta questa procedura implica alcune conseguenze in termini di errori di stima e interpretazione dei risultati. Per questo motivo la letteratura econometrica ha analizzato queste conseguenze, sia dal punto di vista univariato che multivariato. Tra gli altri, Marcellino (1999) identifica e riassume quali proprietà delle serie storiche rimangono invariante e quali variano dopo l'aggregazione temporale, misura quindi l'intensità dell'*aggregation bias* in riferimento ad alcune proprietà.

Negli ultimi anni la letteratura ha prestato particolare attenzione allo sviluppo di modelli econometrici in grado di esaminare tutte le informazioni presenti nel dataset a frequenza mista e nel contempo di mitigare la misspecificazione dovuta alla distorsione (bias) di aggregazione. In questa direzione, approcci influenti sono stati proposti da Ghysels, Santa-Clara e Valkanov (2004), che ha dato il via alla letteratura riguardante le regressioni Mixed Data Sampling (MIDAS). A questi, nello stesso filone di letteratura, si sono aggiunti tra gli altri Clements e Galvão (2008) e Forni, Marcellino e Schumacher (2011).

Un altro filone di letteratura invece, si è dedicato allo studio di modelli che potessero considerare dataset campionati a diversa frequenza da un punto di vista multivariato. Una delle tecniche più utilizzate per studiare processi vettoriali a frequenza mista, è quella proposta da Mariano e Murasawa (2003, 2010), che presuppone l'utilizzo del filtro di Kalman.

Ghysels (2016) propone un'idea diversa per trattare i *Mixed Frequency Vector Autoregressive* (MFVAR), che non richiede procedure di filtraggio. Questo approccio si avvicina di più alla letteratura MIDAS rispetto a quello proposto da Mariano e Murasawa, che invece presuppone l'utilizzo di procedure di filtraggio e l'esistenza di shock latenti.

In generale tutte queste metodologie sono caratterizzate dall'idea di sfruttare l'informazione ad alta frequenza al fine capire nel miglior modo possibile il meccanismo generatore dei dati (ad alta frequenza) sottostante la variabile a bassa frequenza.

Nel Capitolo 2, introdurremo parte della letteratura che si è sviluppata con l'intento di implementare modelli econometrici in grado di ridurre le conseguenze dell'aggregazione temporale, con particolare attenzione all'approccio multivariato.

Nel Capitolo 3 metteremo a punto uno dei modelli econometrici presentati, andando a studiare le differenze pratiche tra l'approccio a frequenza mista e l'approccio a bassa frequenza.

Infine, il Capitolo 4 contiene alcune conclusioni derivanti dall'analisi effettuata nei capitoli precedenti.

Inoltre, nell'Appendice A sono riportati i codici per il software gret1 utilizzati per effettuare l'analisi presentata nel Capitolo 3.

Capitolo 2

La letteratura

In questo Capitolo presenteremo una parte della letteratura che si è sviluppata riguardo ai modelli in grado di considerare dataset campionati con frequenze diverse.

L'obiettivo comune di questi modelli è quello di provvedere alla specificazione di sistemi che inglobino più informazione possibile. Per fare questo nel corso degli anni sono nati diversi approcci, ma in generale la scelta sulla specifica di un modello econometrico per dati a frequenza mista si muove tra due classi.

La prima si basa sulla specifica di modelli congiunti per la variabile di interesse e per i previsori (approccio multivariato, MF-VAR). La seconda classe di modelli fa riferimento ad equazioni in cui la variabile a bassa frequenza è regredita su variabili indipendenti a frequenza mista (approccio univariato, regressioni MIDAS).

Il capitolo è strutturato come segue. Nella prima parte presenteremo alcuni strumenti econometrici come il filtro di Kalman, quello di Hodrick-

Prescott e la scomposizione di Beveridge-Nelson. Poi introdurremo alcune metodologie utilizzate per trattare dati a diversa frequenza, partendo dall'aggregazione temporale per arrivare ai processi VAR a frequenza mista, passando per le regressioni MIXED DATA SAMPLING.

2.1 Il filtro di Kalman

In questa sezione introduciamo uno strumento che deriva il suo nome dai contributi pionieristici di Kalman (1960, 1963) il cui obiettivo è quello di esprimere un determinato sistema dinamico in una particolare forma denominata *rappresentazione spazio-stato*.

Il filtro di Kalman (si veda Hamilton (1994), Capitolo 13) è un algoritmo che può essere utilizzato per aggiornare in modo sequenziale una proiezione lineare del sistema dinamico.

Detto algoritmo, è molto utile non solo per effettuare previsioni del futuro ma, in presenza di dataset a frequenza mista, anche per stimare lo "*stato latente*" delle osservazioni mancanti nella serie a bassa frequenza (si veda Harvey, 1990).

2.1.1 La rappresentazione spazio-stato

Dato un vettore y_t di variabili ($n \times 1$) osservato al tempo t e un vettore α_t non osservabile noto come *vettore di stato*, la rappresentazione spazio-stato

delle dinamiche di y_t è data dal seguente sistema di equazioni:

$$\alpha_t = F\alpha_{t-1} + v_t \quad (2.1.1)$$

$$y_t = A'x_t + H'\alpha_t + w_t \quad (2.1.2)$$

Dove l'equazione (2.1.1) è detta *equazione di stato* e la (2.1.2) è detta *equazione delle osservazioni*. Inoltre le matrici F , A' e H' sono matrici di parametri rispettivamente di dimensioni $(r \times r)$, $(n \times k)$ e $(n \times r)$ e x_t è un vettore $(k \times 1)$ di variabili esogene.

Infine i vettori v_t, w_t sono white noise vettoriali rispettivamente di dimensione $(r \times 1)$ e $(n \times 1)$:

$$E(v_t v_\tau') = Q \quad \text{per } t = \tau, \quad 0 \quad \text{altrimenti} \quad (2.1.3)$$

$$E(w_t w_\tau') = R \quad \text{per } t = \tau, \quad 0 \quad \text{altrimenti} \quad (2.1.4)$$

$$E(v_t w_\tau) = 0 \quad \text{per ogni } t \text{ e } \tau \quad (2.1.5)$$

Dove le matrici Q e R sono $(r \times r)$ e $(n \times n)$ e con la (2.1.5) si ipotizza che i disturbi siano sempre tra loro incorrelati. Abbiamo quindi, con le equazioni dalla (2.1.1) alla (2.1.5), il sistema generale in forma spazio-stato. Propriamente, si tratta di un caso particolare di *State Space Form (SSF)* ed indica un modello *invariante nel tempo*, poiché le varianze degli errori e i coefficienti del modello non variano nel tempo.

2.1.2 Derivazione del filtro di Kalman

Si consideri il sistema in forma spazio-stato (2.1.1)-(2.1.5), si presume che l'analista abbia osservato il set informativo:

$$\mathfrak{S}_T = (y'_1, y'_2, \dots, y'_T, x'_1, x'_2, \dots, x'_T)$$

Dunque, gli obiettivi del filtro di Kalman possono essere molteplici: può essere visto come un algoritmo ricorsivo per calcolare previsioni dei minimi quadrati del vettore di stato, oppure si può essere interessati alla stima dei valori dei parametri incogniti presenti nel sistema, sulla base delle osservazioni disponibili fino al periodo t .

Per quanto riguarda il primo dei due impieghi del filtro di Kalman citati poc'anzi, si è interessati alla proiezione lineare di α_{t+1} sul set informativo \mathfrak{S}_t , cioè:

$$\hat{\alpha}_{t+1|t} = \hat{E}(\alpha_{t+1} | \mathfrak{S}_t) \quad (2.1.6)$$

Il filtro di Kalman calcola queste proiezioni in modo ricorsivo e genera $\hat{\alpha}_{1|0}, \hat{\alpha}_{2|1}, \dots, \hat{\alpha}_{T|T-1}$ in successione. Ad ognuna di queste previsioni è associata una matrice di errore quadratico medio (MSE dall'inglese *Mean Squared Error*), rappresentata dalla seguente matrice ($r \times r$):

$$P_{t+1|t} = E[(\alpha_{t+1} - \hat{\alpha}_{t+1|t})(\alpha_{t+1} - \hat{\alpha}_{t+1|t})'] \quad (2.1.7)$$

La procedura ricorsiva inizia con $\hat{\alpha}_{1|0}$ che indica una previsione di α_1 non basata su y_t o x_t . Essa è soltanto il valore atteso non condizionale di α_1 ,

cioè:

$$\hat{\alpha}_{1|0} = E(\alpha_1),$$

con MSE associato:

$$P_{1|0} = E\{[\alpha_1 - E(\alpha_1)][\alpha_1 - E(\alpha_1)]\}$$

Volendo generalizzare tale asserzione, dalla (2.1.1) si può vedere come, se gli autovalori di F giacciono tutti all'interno del cerchio unitario, il processo di α_t è stazionario in covarianza¹. Pertanto, il valore atteso non condizionale di α_t può essere trovato prendendo i valori attesi di entrambi i lati della (2.1.1) con il seguente risultato

$$\begin{aligned} E(\alpha_t) &= F \cdot E(\alpha_{t-1}); \\ (I_r - F)E(\alpha_t) &= 0 \end{aligned}$$

Inoltre, dal momento che la matrice $(I_r - F)$ non è ipotizzata singolare, l'equazione di cui sopra ammette come unica soluzione $E(\alpha_t) = 0$. Possiamo analogamente trovare la varianza non condizionale di α_t moltiplicando la (2.1.1) per la sua trasposta e prendendo i valori attesi:

$$E(\alpha_t \alpha_t') = E[(F\alpha_{t-1} + v_t)(\alpha_{t-1}'F' + v_t')] = F \cdot E(\alpha_{t-1} \alpha_{t-1}') \cdot F' + E(v_t v_t') \quad (2.1.8)$$

¹L'equazione (2.1.1) è quella di un VAR(1) che risulta essere stazionario quando tutti gli autovalori della matrice F sono minori di 1 in modulo $|\lambda| < 1$.

I termini di prodotto incrociato scompaiono poiché²:

$$E(v_t \alpha'_\tau) = 0 \quad \text{per } \tau = t-1, t-2, \dots, 1$$

$$E(v_t y'_\tau) = 0 \quad \text{per } \tau = t-1, t-2, \dots, 1$$

Allora se con $\Sigma = E(\alpha_t \alpha'_t)$ indichiamo la matrice di varianze e covarianze di α_t , l'equazione (2.1.8) implica:

$$\Sigma = F \Sigma F' + Q \quad (2.1.9)$$

La cui soluzione per Σ può essere ottenuta tramite l'operatore *vec*. Dunque se A è una matrice ($m \times n$), allora $\text{vec}(A)$ è un vettore ($mn \times 1$), ottenuto disponendo le colonne di A , una sotto l'altra da destra verso sinistra. Suddetta operazione è detta *vettorizzazione*³ della matrice A . Inoltre è molto utile per la risoluzione dell'equazione (2.1.9) ricordare un risultato collegato all'operatore *vec*, e cioè: *Siano A, B, C matrici le cui dimensioni sono tali che il prodotto $A \cdot B \cdot C$ esista. Allora*

$$\text{vec}(ABC) = (C' \otimes A) \cdot \text{vec}(B) \quad (2.1.10)$$

Dove " \otimes " indica il prodotto di Kronecker.

²Le equazioni che seguono valgono anche per w_t al posto di v_t .

³Se la matrice A è simmetrica di dimensione ($n \times n$) potrebbe essere più utile la "*semi-vettorizzazione*" attraverso l'operatore *vech*. Il vettore $\text{vech}(A)$, di una matrice simmetrica ($n \times n$), è un vettore colonna di dimensioni $(n(n-1)/2 \times 1)$ ottenuto vettorizzando solo la parte triangolare inferiore di A .

Dunque la soluzione per Σ dell'equazione (2.1.9) non è altro che:

$$\text{vec}(\Sigma) = [I_{r^2} - F \otimes F]^{-1} \cdot \text{vec}(Q) \quad (2.1.11)$$

Così, in generale, ammesso che gli autovalori di F siano all'interno del cerchio unitario, le iterazioni del filtro di Kalman possono essere iniziate con $\hat{\alpha}_{1|0} = 0$ e con $P_{1|0}$, che è la matrice ($r \times r$) i cui elementi, espressi come un vettore colonna, sono dati da:

$$\text{vec}(P_{1|0}) = [I_{r^2} - F \otimes F]^{-1} \cdot \text{vec}(Q) \quad (2.1.12)$$

Il passo successivo è quello di calcolare grandezze analoghe per i periodi successivi, nonché la previsione di y_t . Perciò, dati $\hat{\alpha}_{t|t-1}$ e $P_{t|t-1}$, l'obiettivo è quello di calcolare $\hat{\alpha}_{t+1|t}$ e $P_{t+1|t}$, e la previsione del valore di y_t è espressa come segue:

$$\hat{y}_{t|t-1} = \hat{E}(y_t | \mathfrak{S}_{t-1})$$

In aggiunta, in forza della (2.1.2), l'equazione di cui sopra diventa:

$$\hat{y}_{t|t-1} = A'x_t + H' \cdot \hat{E}(\alpha_t | \mathfrak{S}_{t-1}) = A'x_t + H'\hat{\alpha}_{t|t-1} \quad (2.1.13)$$

Dove il *Mean Squared Error* della previsione è dato da:

$$E[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})'] = H'P_{t|t-1}H + R \quad (2.1.14)$$

Successivamente, il valore corrente di α_t viene aggiornato sulla scorta

dell'osservazione disponibile di y_t e si ha:

$$\hat{\alpha}_{t|t} = \hat{E}(\alpha_t | \mathfrak{S}_t) \quad (2.1.15)$$

Per risolvere la (2.1.15) si utilizza la formula di aggiornamento di una proiezione lineare che, nel caso specifico, dice che:

$$\begin{aligned} \hat{\alpha}_{t|t} = & \hat{\alpha}_{t|t-1} + \{E[(\alpha_t - \hat{\alpha}_{t|t-1})(y_t - \hat{y}_{t|t-1})']\} \cdot \\ & \cdot \{E[(y_t - \hat{y}_{t|t-1})(y_t - \hat{y}_{t|t-1})']\}^{-1} \cdot (y_t - \hat{y}_{t|t-1}) \end{aligned}$$

Perciò sostituendo la (2.1.14) e sapendo che $E[(\alpha_t - \hat{\alpha}_{t|t-1})(y_t - \hat{y}_{t|t-1})'] = P_{t|t-1}H$, si ha:

$$\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_tH'\hat{\alpha}_{t|t-1}) \quad (2.1.16)$$

Il MSE associato, ottenuto con la formula di aggiornamento di una proiezione lineare, è espresso come segue:

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1} \quad (2.1.17)$$

A questo punto, l'equazione di stato (2.1.1), viene utilizzata per effettuare la previsione un "passo" in avanti di α_{t+1} :

$$\hat{\alpha}_{t+1|t} = \hat{E}(\alpha_{t+1} | \mathfrak{S}_t) \quad (2.1.18)$$

Sostituendo la (2.1.16) nella (2.1.18) si ottiene la previsione:

$$\hat{\alpha}_{t+1|t} = F\hat{\alpha}_{t|t-1} + \underbrace{FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}}_{K_t}(y_t - A'x_t - H'\hat{\alpha}_{t|t-1}) \quad (2.1.19)$$

La matrice dei coefficienti K_t è nota come la *matrice di guadagno*, ed indica la direzione verso cui va corretta la previsione di α_{t+1} a seconda che le previsioni precedenti abbiano determinato valori inferiori o superiori al valore effettivo di y_t . L'errore quadratico medio di questa previsione si ricava nel modo seguente:

$$P_{t+1|t} = E[(\alpha_{t+1} - \hat{\alpha}_{t+1})(\alpha_{t+1} - \hat{\alpha}_{t+1})'] = FP_{t|t}F' + Q \quad (2.1.20)$$

$$P_{t+1|t} = F[P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1}]F' + Q \quad (2.1.21)$$

Dove la (2.1.21) si ottiene semplicemente sostituendo la (2.1.17) nella (2.1.20).

Le equazioni (2.1.19) e (2.1.21) sono note in letteratura come *filtro di Kalman*.

Riepilogando il filtro di Kalman parte dal valore atteso e dalla varianza non condizionali di α_t . Si procede poi all'iterazione delle equazioni di aggiornamento di Kalman descritte nelle (2.1.19) e (2.1.21). La previsione $\hat{\alpha}_{t+1|t}$ indica la miglior previsione lineare di α_{t+1} basata sul set informativo \mathfrak{S}_t , e la matrice $P_{t+1|t}$ indica il MSE di tale previsione. Infine si effettua la previsione di y_{t+1} con annesso MSE.

È interessante osservare che, per calcolare l'iterazione insita nella (2.1.21) non occorre necessariamente utilizzare la (2.1.19), poiché i valori di $P_{t|t-1}$ e K_t non sono funzioni dei dati, ma esclusivamente dei parametri presenti nel processo. Spesso suddetti parametri sono incogniti e vanno stimati.

2.1.3 Stima di massima verosimiglianza dei parametri

Nella realtà, quando si effettuano analisi tramite il filtro di Kalman, i parametri contenuti nelle matrici F, A', H', Q e R sono ignoti. Pertanto se lo stato iniziale α_1 e le innovazioni v_t, w_t sono gaussiane multivariate, allora la distribuzione di y_t condizionale a $(x_t, \mathfrak{S}_{t-1})$ è gaussiana con valore atteso dato dalla (2.1.13) e varianza data dalla (2.1.14), cioè:

$$\begin{aligned}
 f_{y_t|x_t, \mathfrak{S}_{t-1}}(y_t|x_t, \mathfrak{S}_{t-1}) = & \\
 (2\pi)^{-n/2} [H'P_{t|t-1}H + R]^{-1/2} & \\
 \times \exp \left\{ -\frac{1}{2} (y_t - A'x_t - H'\hat{\alpha}_{t|t-1})' (H'P_{t|t-1}H + R)^{-1} \right. & \quad (2.1.22) \\
 \left. \times (y_t - A'x_t - H'\hat{\alpha}_{t|t-1}) \right\} & \text{ per } t = 1, 2, \dots, T
 \end{aligned}$$

Così, costruendo dalla (2.1.22) la log-verosimiglianza, può essere massimizzata numericamente rispetto ai parametri incogniti. L'espressione di cui sopra rappresenta l'esatta verosimiglianza e non quella condizionale, rendendo la *State Space Form* e il *Kalman Filter* utili strumenti per la stima di modelli in serie storiche, ad esempio per modelli *Moving Average* (MA).

2.1.4 Lo smoothing

In certi casi, l'obiettivo della stima con il filtro di Kalman, può non essere quello di prevedere il valore del vettore di stato α_t , ma piuttosto si è interessati all'interpretazione di tale variabile non osservabile⁴. Di

⁴Hamilton (1994) propone l'esempio contenuto in Stock e Watson (1991) dove si è interessati a conoscere lo stato del ciclo economico ad ogni periodo storico.

conseguenza, è preferibile che la stima comprenda l'intero set informativo e non soltanto l'informazione pregressa (\mathfrak{S}_{t-1}) disponibile ad ogni step della recursione di Kalman. In altre parole l'obiettivo è formare un'inferenza sul valore di α_t basata sul filtraggio (o *smoothing*) dell'insieme completo dei dati a disposizione, cioè:

$$\hat{\alpha}_{t|T} = \hat{E}(\alpha_t | \mathfrak{S}_T) \quad \text{con } T > t. \quad (2.1.23)$$

Perciò, la serie delle stime *smorzate* $\hat{\alpha}_{t|T}$ viene calcolata come segue. In primis si calcola il filtro di Kalman (2.1.19)-(2.1.21) e le previsioni basate sulle osservazioni fino al tempo corrente (2.1.16)-(2.1.17). Le stime $\hat{\alpha}_{t|T}$ vengono poi ottenute con una procedura a ritroso che dal tempo T al tempo 1, corregge le previsioni one-step $\alpha_{t|t-1}$ andando a considerare l'informazione che esse non hanno precedentemente utilizzato (Chirico *et al.*, 2014).

2.2 Il filtro di Hodrick-Prescott

Il filtro di Hodrick-Prescott⁵ (HP, 1997) rappresenta un approccio alternativo alla stima e all'interpretazione di variabili non osservabili. In particolare, il filtro HP, fornisce una stima della variabile non osservabile come soluzione al seguente problema di minimizzazione:

$$\min_{y_t^*} \sum_{t=1}^T (y_t - y_t^*)^2 + \lambda \sum_{t=3}^T [(y_t^* - y_{t-1}^*) - (y_{t-1}^* - y_{t-2}^*)]^2 \quad (2.2.1)$$

Dove y_t è la variabile osservata mentre y_t^* è la variabile non osservata oggetto di stima. Il filtro HP è stato originariamente progettato per scomporre la serie in una componente di lungo periodo chiamata 'trend' e una componente ciclica di breve periodo. Quindi, y_t^* rappresenta il trend sottostante alla variabile y_t , $(y_t - y_t^*)$ le fluttuazioni cicliche e $\Delta^2 y_t^* = (y_t^* - y_{t-1}^*) - (y_{t-1}^* - y_{t-2}^*)$ è la variazione del tasso di crescita del trend sottostante la serie. Detto ciò quindi, il primo termine della (2.2.1) è la somma delle componenti cicliche al quadrato. Il secondo termine è un multiplo λ della somma dei quadrati delle differenze seconde della componente di trend, cioè: $\Delta^2 y_t^*$. Il termine λ è detto parametro di smoothing e influenza le variazioni del tasso di crescita della componente di trend: maggiore è il valore di λ , più *liscia* (in inglese *smooth*) è la serie del trend. Hodrick e Prescott scelgono, per il parametro di smoothing, il valore di $\lambda = 1600$. Alla base di questa scelta c'è l'assunzione che un grande cambiamento nella componente ciclica in un trimestre può essere intorno al 5%, mentre nel

⁵Edward Christian Prescott ha ricevuto il premio Nobel per l'economia nel 2004 insieme a Finn E. Kydland "per il loro contributo alla macroeconomia dinamica attraverso gli studi sulla coerenza temporale della politica economica e le forze trainanti dei cicli economici".

trend intorno al 0.125%, suggerendo una scelta di $\lambda = (5/0.125)^2 = 1600$. È ovvio, quindi, che i valori di λ dipendono anche dalla frequenza temporale delle serie.

2.2.1 Confronto tra Hodrick-Prescott e filtro di Kalman

Boone (2000) propone un confronto tra i due approcci trattati finora. Infatti, è possibile scrivere il filtro di Hodrick-Prescott in forma spazio-stato e imporre alcune restrizioni specifiche sui suoi elementi in modo che possa essere riprodotto tramite il filtro di Kalman. L'equazione di stato è analoga alla prima componente della (2.2.1):

$$y_t = y_t^* + \varepsilon_t \quad \text{con} \quad \varepsilon_t \sim N(0, H) \quad (2.2.2)$$

Mentre l'equazione delle osservazioni è direttamente legata alla seconda parte della (2.2.1), e cioè:

$$\Delta y_t^* = \Delta y_{t-1}^* + v_t \quad \text{dove} \quad v_t \sim N(0, Q) \quad (2.2.3)$$

In altre parole, la variazione del trend segue un random walk. Inoltre, il parametro di smoothing λ è il rapporto tra le due varianze. Spesso però quest'ultimo, piuttosto che essere stimato, viene calibrato. Una scelta comune per i dati macroeconomici, è renderlo una funzione della periodicità: $\lambda = 100 \cdot p^2$ dove p è il numero di sottoperiodi in un anno (Cottrell e Lucchetti, 2019b).

A questo punto si procede con la stima del filtro HP, espresso tramite la

rappresentazione spazio-stato (2.2.2)-(2.2.3), attraverso il filtro di Kalman. Esistono diversi vantaggi rispetto alla possibilità di riprodurre il filtro di Hodrick-Prescott attraverso quello di Kalman. Uno di questi è dovuto al fatto che la procedura di stima utilizzata dal filtro di Kalman prende in considerazione, come già notato in precedenza, la massima verosimiglianza esatta.

Il filtro di Hodrick-Prescott è stato però anche aspramente criticato: su tutti da Hamilton (2018), il quale presenta evidenze contro l'uso di tale procedura. Nello specifico l'autore dimostra come il filtro HP introduce relazioni spurie, tra le serie storiche considerate, che non derivano in alcun modo dal meccanismo generatore dei dati. Inoltre i valori filtrati alla fine del campione sono molto diversi da quelli nel mezzo e sono anch'essi caratterizzati da correlazioni spurie. Un altro problema evidenziato da Hamilton (2018) è dovuto al fatto che, stimando il parametro di smoothing λ , si producono valori per quest'ultimo enormemente in disaccordo con la pratica comune.

2.3 La scomposizione di Beveridge-Nelson

Beveridge e Nelson (1981) propongono una procedura per la scomposizione di una serie storica in una componente permanente e una componente transitoria, consentendo ad entrambe di essere stocastiche. L'assunzione che viene fatta è quella per cui la componente permanente è un *random walk*⁶. Inoltre si assume che la componente transitoria o ciclica è un processo

⁶Un random walk è un processo $I(1)$ la cui differenza prima è un white noise.

stazionario a media zero.

L'idea è, quindi, che la componente ciclica di una serie può essere separata dal trend o componente permanente.

2.3.1 Scomposizione di un processo ARMA integrato

Spesso negli studi empirici vengono analizzate serie storiche che risultano $I(1)$, in quanto la loro differenza prima è stazionaria.

Beveridge e Nelson (1981), nel loro articolo originario, per operare la scomposizione, partono da una serie integrata di ordine uno, y_t , la cui differenza prima è stazionaria, Δy_t . Dunque se Δy_t è stazionaria in covarianza possiamo scrivere la rappresentazione di Wold:

$$\Delta y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots \quad (2.3.1)$$

Dove le ε_t sono disturbi incorrelati con media zero e varianza σ^2 . Inoltre, il valore atteso di y_{t+k} condizionale ai dati disponibili al tempo t è dato da:

$$\begin{aligned} \hat{y}_{t+k} &= E(y_{t+k} | y_t, y_{t-1}, \dots) \\ &= y_t + E(\Delta y_{t+1}, \dots, \Delta y_{t+k} | y_t, y_{t-1}, \dots) \\ &= y_t + \hat{z}_{t+1} + \dots + \hat{z}_{t+k} \quad \text{dove:} \\ \hat{z}_{t+i} &= E(\Delta y_{t+i} | y_t, y_{t-1}, \dots) \end{aligned} \quad (2.3.2)$$

La y_t può quindi essere espressa come la somma delle Δy_t , per di più dalla (2.3.1) si ha che:

$$\hat{z}_{t+i} = \mu + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} + \dots = \sum_{j=1}^{\infty} \theta_j \varepsilon_{t+1-j} \quad (2.3.3)$$

Ora sostituiamo la (2.3.3) nella (2.3.2) e raccogliendo a fattor comune si arriva a:

$$\hat{y}_{t+k} = k\mu + y_t + \left(\sum_{i=1}^k \theta_i \right) \varepsilon_t + \left(\sum_{i=2}^{k+1} \theta_i \right) \varepsilon_{t-1} + \dots \quad (2.3.4)$$

Se consideriamo un orizzonte di previsione potenzialmente infinito o comunque abbastanza lungo abbiamo che:

$$\hat{y}_{t+k} \simeq k\mu + y_t + \left(\sum_{i=1}^{\infty} \theta_i \right) \varepsilon_t + \left(\sum_{i=2}^{\infty} \theta_i \right) \varepsilon_{t-1} + \dots \quad (2.3.5)$$

Poiché Δy_t è un processo stazionario per definizione, si ha che $\sum_{i=1}^{\infty} \theta_i < \infty$, quindi, si può dire che la previsione di \hat{y}_{t+k} è asintoticamente una funzione lineare dell'orizzonte di previsione k , con pendenza μ e con "intercetta" che è a sua volta un processo stocastico. Tale intercetta non è altro che la componente permanente o di trend di y_t . Chiamando questa componente \bar{y}_t si ha:

$$\bar{y}_t = y_t + \left(\sum_{i=1}^{\infty} \theta_i \right) \varepsilon_t + \left(\sum_{i=2}^{\infty} \theta_i \right) \varepsilon_{t-1} + \dots \quad (2.3.6)$$

Per provare che \bar{y}_t è un random walk con drift pari a μ e quindi un processo $I(1)$, dobbiamo solamente dimostrare che la differenza prima di \bar{y}_t è uguale

a μ più un *white noise* incorrelato:

$$\bar{y}_t - \bar{y}_{t-1} = \Delta y_t + \left(\sum_{i=1}^{\infty} \theta_i \right) \varepsilon_t + (\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots) \quad (2.3.7)$$

Se poi sostituiamo la (2.3.1) nella equazione (2.3.7) sopra, si ottiene:

$$T_t = \bar{y}_t - \bar{y}_{t-1} = \mu + \left(\sum_{i=0}^{\infty} \theta_i \right) \varepsilon_t \quad \text{con } \theta_0 = 1 \quad (2.3.8)$$

Per definizione ε_t è un white noise incorrelato di conseguenza la componente permanente T_t è un random walk con drift.

Riassumendo, la previsione di y_t si avvicina ad un percorso lineare man mano che la si estende nel futuro. Inoltre la componente permanente non è altro che il valore che la serie avrebbe se si trovasse su tale percorso. In altre parole, la componente di trend, è la previsione a lungo termine della serie storica calcolata sulla base del tasso medio di variazione della serie stessa. Inoltre, si dimostra che tale componente permanente segue un random walk. Il “trend” di y_t , quindi, può essere interpretato come il valore osservato più tutti i cambiamenti futuri prevedibili. Per questo motivo si può riscrivere nella forma:

$$T_t = y_t + \lim_{k \rightarrow \infty} [(\hat{z}_{t+1} + \dots + \hat{z}_{t+k}) - k\mu] \quad (2.3.9)$$

A questo punto, per ottenere la componente ciclica o transitoria della serie y_t non resta che sottrarre ai valori osservati, quelli trovati dal calcolo della

componente permanente:

$$\begin{aligned}
 C_t &= y_t - T_t \\
 &= \hat{y}_t - \hat{y}_t - \left\{ \lim_{k \rightarrow \infty} [(\hat{z}_{t+1} + \dots + \hat{z}_{t+k}) - k\mu] \right\} \\
 &= - \left(\sum_{i=1}^{\infty} \theta_i \right) \varepsilon_t - \left(\sum_{i=2}^{\infty} \theta_i \right) \varepsilon_{t-1} - \dots
 \end{aligned} \tag{2.3.10}$$

Per chiarire il concetto può essere utile un esempio. Supponiamo che la differenza prima di una qualsiasi serie x_t integrata di ordine uno, segua un processo di tipo MA(1) con media μ :

$$\Delta x_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1} \tag{2.3.11}$$

La componente permanente o di trend la ricaviamo dalla (2.3.8):

$$T_t = \mu + (1 + \theta)\varepsilon_t \tag{2.3.12}$$

Mentre, la componente transitoria è, dalla (2.3.10):

$$C_t = -\theta \varepsilon_t \tag{2.3.13}$$

Ricapitolando, in generale abbiamo che, secondo Beveridge e Nelson (1981),

le componenti T_t e C_t della serie x_t sono rispettivamente:

$$x_t = T_t + C_t \quad (2.3.14)$$

$$T_t = \mu + (1 + \theta)\varepsilon_t \quad (2.3.15)$$

$$C_t = -\theta\varepsilon_t \quad (2.3.16)$$

Dove T_t è un random walk e C_t un processo stazionario.

Se, invece, ragioniamo in termini di polinomi nell'operatore ritardo, si può ricavare una formulazione alternativa ed equivalente della scomposizione di Beveridge-Nelson (BN). Si parte da una proprietà dei polinomi: dato un polinomio qualsiasi $C(z)$ di ordine q , è sempre possibile trovare un polinomio $C^*(z)$ di grado $(q - 1)$ tale per cui:

$$C(z) = C(1) + C^*(z)(1 + z) \quad \text{oppure:} \quad (2.3.17)$$

$$C^*(z) = \frac{C(z) - C(1)}{(1 - z)} \quad (2.3.18)$$

Successivamente, consideriamo un qualsiasi processo $y_t \sim I(1)$, per cui vale $\Delta y_t \sim I(0)$ ed ammette rappresentazione di Wold:

$$\Delta y_t = \mu + C(L)\varepsilon_t \quad (2.3.19)$$

A questo punto applichiamo al polinomio $C(L)$ la proprietà (2.3.17):

$$\Delta y_t = \mu + [C(1) + C^*(L)(1 - L)]\varepsilon_t = C(1)\varepsilon_t + C^*(L)\Delta\varepsilon_t \quad (2.3.20)$$

2.3. LA SCOMPOSIZIONE DI BEVERIDGE-NELSON

Inoltre, tramite la cumulata⁷ di Δy_t , arriviamo a costruire:

$$\text{cum}(\Delta y_t) = y_t = \mu + C(1)\text{cum}(\varepsilon_t) + C^*(L)\varepsilon_t \quad (2.3.21)$$

Dopo questo passaggio, definiamo un processo d_t tale che $\Delta d_t = \varepsilon_t$. Ossia un random walk i cui incrementi sono dati da ε_t , e lo sostituiamo nella (2.3.21), che diventa:

$$y_t = \mu + C(1)d_t + C^*(L)\varepsilon_t \quad (2.3.22)$$

Quindi, si ha:

$$y_t = T_t + C_t \quad \text{con:} \quad (2.3.23)$$

$$T_t = \mu + C(1)d_t \quad (2.3.24)$$

$$C_t = C^*(L)\varepsilon_t \quad (2.3.25)$$

Se ripetiamo l'esempio precedentemente proposto nell'equazione (2.3.11), si osserva che i risultati a cui si giunge sono gli stessi, infatti:

$$T_t = \mu + (1 + \theta)d_t \quad (2.3.26)$$

$$C_t = \frac{1 + \theta L - 1 - \theta}{(1 - L)}\varepsilon_t = \frac{-\theta(1 - L)}{(1 - L)}\varepsilon_t = -\theta\varepsilon_t \quad (2.3.27)$$

Un'interpretazione interessante della grandezza $C(1)$ è quella di misura della persistenza di un dato processo, poiché misura la frazione dello shock che permane nel processo dopo un tempo 'infinito'. È possibile controllare

⁷Operazione opposta rispetto alla differenziazione.

che, applicando la scomposizione qui descritta ad un processo stazionario, $C(1) = 0$, mentre $C(1) \neq 0$ nel caso di processi integrati (Lucchetti, 2015).

2.3.2 Un approccio spazio-stato

Il metodo di scomposizione introdotto da Beveridge e Nelson fornisce un modo conveniente per stimare la componente permanente e transitoria di una serie storica integrata. Dato un modello di previsione per le differenze prime della serie, il trend BN è la previsione di lungo termine del livello della serie (meno il *drift* deterministico), mentre il ciclo BN è il divario tra il livello attuale della serie e le sue previsioni a lungo termine. Nella pratica, il calcolo del trend e del ciclo attraverso la metodologia proposta da Beveridge e Nelson, è spesso complicato dalla presenza di somme infinite nella previsione a lungo termine. Morley (2002) sottolinea, tuttavia, che il calcolo esatto del trend e del ciclo BN è relativamente semplice se il modello di previsione può essere inserito in una forma spazio-stato.

Supponiamo quindi che, la differenza prima della serie oggetto di analisi Δy_t , sia una combinazione lineare degli elementi di un vettore di stato X_t .

$$\Delta y_t = HX_t + \varepsilon_t \quad (2.3.28)$$

Supponiamo inoltre che il vettore di stato X_t evolva secondo la seguente equazione di differenza stocastica del primo ordine:

$$X_t = FX_{t-1} + v_t \quad (2.3.29)$$

2.3. LA SCOMPOSIZIONE DI BEVERIDGE-NELSON

Dove ε_t, v_t sono white noise e gli autovalori di F sono in modulo minori di uno. Quindi è semplice mostrare che la previsione k periodi in avanti della differenza prima di y_t , non è altro che:

$$E(\Delta y_{t+k}|I_t) = HF^k E(X_t|I_t) \quad (2.3.30)$$

Si noti che, poiché il vettore di stato può contenere elementi non osservabili, potrebbe essere necessario ottenere il valore atteso $E(\Delta y_{t+k}|I_t)$ prima del calcolo della (2.3.30). Fortunatamente, il filtro Kalman, che può essere utilizzato per ottenere stime esatte della massima verosimiglianza per i modelli spazio-stato con elementi non osservabili, fornisce questo valore atteso. Pertanto, indicando con $X_{t|t}$ il calcolo del valore atteso del vettore di stato attraverso il filtro di Kalman, la componente di trend BN della serie y è:

$$T_t = y_t + HF(1 - F)^{-1}X_{t|t} \quad (2.3.31)$$

Mentre la componente ciclica, in questo caso è espressa come segue:

$$C_t = -HF(1 - F)^{-1}X_{t|t} \quad (2.3.32)$$

Morley (2002) conclude che, un approccio spazio-stato fornisce un quadro unificato semplice e generale per il calcolo del trend e del ciclo attraverso la scomposizione di Beveridge-Nelson per un'ampia varietà di modelli, inclusi tutti i modelli ARIMA univariati e vettoriali.

Presentati questi strumenti, iniziamo con l'introdurre le tecniche e le metodologie usate in letteratura per affrontare i problemi derivanti dalla

presenza di serie storiche campionate con frequenza diversa.

2.4 L'aggregazione temporale

L'approccio più semplice alla gestione di serie campionate a diversa frequenza, è quello dell'aggregazione temporale. I dati ad alta frequenza vengono aggregati a quelli a bassa frequenza facendo una media oppure prendendo un valore rappresentativo (ad esempio, l'ultimo mese del trimestre).

I metodi di aggregazione standard dipendono soprattutto dalla natura (stock o flussi) delle variabili, infatti di solito si usa la media per le variabili stock, e la somma per le variabili flusso.

Prendere l'ultimo valore disponibile è un'altra opzione sia per le variabili stock che per quelle flusso. L'assunzione sottostante questa alternativa è che l'informazione dei periodi precedenti si riflette tutta sull'ultimo valore rappresentativo.

Gli approcci che abbiamo citato consentono, dunque, di ottenere un dataset bilanciato con le serie storiche campionate con la stessa frequenza.

Marcellino (1999) asserisce che l'aggregazione temporale si presenta quando la frequenza della generazione dei dati è inferiore a quella della raccolta dei dati, quindi non tutte le realizzazioni del processo stocastico originale $\{x_t\}_{t=0}^{\infty}$ sono osservabili. I dati disponibili, alla frequenza più bassa, possono essere pensati come realizzazioni del processo aggregato $\{\mathbf{x}_\tau\}_{\tau=0}^{\infty}$ dove τ è la nuova (bassa) frequenza temporale e gli elementi di \mathbf{x}_τ sono particolari funzioni del processo x_t , il quale ordine di tempo è alla

frequenza più alta, ma non è interamente osservabile. L'esatta relazione tra i due processi dipende dallo schema di aggregazione temporale.

Formalizzando i diversi metodi di aggregazione presentati in precedenza abbiamo che, se x è una variabile stock, il processo di aggregazione più utilizzato è $\{\mathbf{x}_\tau\}_{\tau=0}^\infty = \{x_{tk}\}_{t=0}^\infty$, cioè solo il k -esimo elemento del processo originario viene considerato. Questa situazione è definita da Marcellino (1999) come campionamento "*point-in-time*". Mentre quando gli elementi di x sono flussi, gli elementi del processo aggregato sono definiti come $\{\mathbf{x}_\tau\}_{\tau=0}^\infty = \{\sum_{i=0}^{k-1} x_{tk+i}\}_{t=0}^\infty$. Questo caso è definito campionamento *medio*.

Marcellino (1999) conclude perciò che le conseguenze dell'aggregazione temporale possono essere molteplici all'interno dei modelli in serie storiche. L'autore studia, infatti, gli effetti dell'aggregazione temporale su un insieme di caratteristiche di interesse nell'ambito delle serie storiche, come esogeneità, casualità e cointegrazione.

In primo luogo, Marcellino (1999) replica, con alcune sostanziali differenze, l'esercizio di Lütkepohl (1987). In questo esercizio i due autori derivano, per un processo ARMA, i limiti superiori dell'ordine di ritardo delle componenti AR e MA. In particolare, Marcellino analizza i dati supponendo che il loro processo di generazione possa essere approssimato da un ARMA(p, q). I valori p e q sono scelti da un test F ricorsivo per la significatività del p -esimo e del q -esimo ritardo, partendo da $p = q = 24$. Tale procedura è ripetuta sia per il processo ad alta frequenza sia per il processo aggregato, usando due metodi di aggregazione temporale: il *point-in-time sampling* e l'*average sampling*. Da questa analisi si osserva una sostanziale modifica dell'ordine dei ritardi e dei coefficienti delle componenti AR e

MA, tra il processo ad alta frequenza e quello aggregato alla frequenza più bassa. Inoltre si notano differenze significative anche tra i due metodi di aggregazione utilizzati.

In secondo luogo, Marcellino ribadisce che molte proprietà empiriche non sono invarianti rispetto all'aggregazione. In altre parole, l'aggregazione temporale spesso altera alcune proprietà che sarebbero verificate ad una frequenza disaggregata. Ad esempio, Sims (1971) e Wei (1982) asseriscono che: "la causalità di Granger potrebbe perdersi oppure potrebbe essere creata in modo spurio dopo l'aggregazione". In un processo stocastico, una proprietà è definita "invariante" rispetto all'aggregazione temporale se essa è presente (o meno) sia prima che dopo l'aggregazione. Nella Tabella 2.1 sono riportate le proprietà analizzate nell'articolo di Marcellino (1999), con i relativi autori che per primi ne hanno evidenziato l'irregolarità in corrispondenza dell'aggregazione temporale.

| Proprietà | Autore | Invarianza |
|---------------------------------|-------------------------|------------|
| Radici unitarie | Pierse e Snell (1995) | Sì |
| Cointegrazione | Granger (1990) | Sì |
| Esogeneità | Campos et al. (1990) | No |
| Causalità | Sims (1971), Wei (1982) | No |
| Impulse Response Function (IRF) | Swanson e Granger(1997) | No |
| Decomposizione Trend-ciclo | Lippi e Reichlin (1991) | No |
| Forecasting | Lütkepohl (1987) | No |

Tabella 2.1: Gli effetti dell'aggregazione su alcune proprietà, Marcellino (1999).

Inoltre, anche per le proprietà che presentano invarianza (come integrazione e cointegrazione), la potenza delle procedure di test cala drasticamente in campioni finiti per via del minor numero di osservazioni presenti nel

dataset.

Con l'operazione di aggregazione, perciò, una moltitudine di dati disponibili non vengono considerati, dunque potenziali informazioni utili vengono perse e il modello può risultare mispecificato. Tale situazione può portare a problemi di identificazione, stima e interpretazione ("Aggregation Bias"). Di conseguenza modelli econometrici che considerano direttamente dataset a diversa frequenza possono essere utili per superare le anzidette problematiche.

2.5 Mixed Data Sampling

Uno degli approcci più significativi degli ultimi anni è dovuto a Ghysels et al. (2004) il quale ha dato il via alla letteratura riguardante i modelli *Mixed Data Sampling* (MIDAS). In generale, le regressioni MIDAS affrontano una situazione spesso incontrata nella pratica, in cui l'informazione rilevante è contenuta nei dati ad alta frequenza, mentre la variabile di interesse è campionata ad una frequenza più bassa. Le regressioni MIDAS coinvolgono dunque regressori con frequenze di campionamento diverse.

Per introdurre le regressioni MIDAS, seguendo la notazione utilizzata da Ghysels et al. (2004), supponiamo che y_t sia rilevata ad una certa frequenza fissa, ad esempio annuale, trimestrale, mensile o giornaliera, e prendiamo questo intervallo di tempo (t) come riferimento. Inoltre, sia $x^{(m)}$ campionato m volte più frequentemente, ad esempio con dati annuali e $m = 4$, abbiamo che $x^{(4)}$ è una serie trimestrale. Possiamo quindi scrivere una semplice

regressione lineare MIDAS come segue:

$$y_t = \beta_0 + B(L^{1/m})x_t^{(m)} + \varepsilon_t^{(m)} \quad (2.5.1)$$

Dove:

$$B(L^{1/m}) = \sum_{j=0}^p B(j)L^{j/m} \quad (2.5.2)$$

È un polinomio di ordine p (anche infinito) nell'operatore ritardo $L^{(1/m)}$, e $L^{j/m}x_t^{(m)} = X_{t-j/m}^{(m)}$ e $B(j)$ sono parametri. Nell'esempio sopra riportato si avrebbe che l'equazione (2.5.1) è una proiezione di y_t (annuale) su dati trimestrali $x_t^{(m)}$ utilizzando fino a j_{max} ritardi trimestrali. Tuttavia, anche se assumessimo che l'ordine del polinomio $B(L^{1/m})$ sia finito, il numero dei parametri potrebbe essere piuttosto grande. Ad esempio, per studiare le fluttuazioni giornaliere in un processo negli ultimi sei mesi, avremmo bisogno di stimare $6 \times 22 = 132$ parametri (assumendo di avere 22 giorni di transazioni per mese). Pertanto siamo di fronte ad un *trade-off*. Da un lato la regressione MIDAS sfrutta un set informativo più ampio, dall'altro il costo è la proliferazione dei parametri, poiché un polinomio $B(L^{1/m})$ appropriato potrebbe presentare molti ritardi della variabile esplicativa $x_{t-j/m}^{(m)}$, quindi molti parametri da stimare. C'è bisogno quindi di definire una struttura ai parametri in modo da ottenere risultati ragionevoli e diminuire i tempi di calcolo. Idealmente, vogliamo conservare la maggior parte delle informazioni nella regressione MIDAS, riducendo al contempo il numero di parametri da stimare. Ci sono diversi modi per ridurre lo spazio dei parametri, ad esempio Ghysels et al. (2005) propongono un ap-

proccio che ha le sue radici nella letteratura sui modelli a ritardi distribuiti. Questa parametrizzazione, che è tra le più usate, è chiamata “*Exponential Almon Lag*”. Tale specificazione deriva dal contributo di Almon (1965) alla letteratura dei modelli a ritardi distribuiti.

Si tratta di trasformare il modello espresso nella (2.5.1) come segue:

$$y_t = \beta_0 + B(L^{1/m}; \theta)x_t^{(m)} + \varepsilon_t^{(m)} \quad (2.5.3)$$

Dove, in questo caso il polinomio nell’operatore ritardo è espresso nel seguente modo:

$$B(L^{1/m}; \theta) = \sum_{j=0}^p b(j; \theta)L^{j/m} \quad (2.5.4)$$

Inoltre, θ è un vettore di parametri, la cui stima determina il valore dei coefficienti della regressione MIDAS. Infatti, la parametrizzazione parsimoniosa dei coefficienti dei ritardi opera tramite il termine $b(j; \theta)$ che, in caso si scegliesse di usare la forma funzionale dell’*Exponential Almon Lag*, sarebbe espresso come:

$$b(j; \theta) = \frac{\exp\{\theta_1 j + \dots + \theta_Q j^Q\}}{\sum_{j=1}^K \exp\{\theta_1 j + \dots + \theta_Q j^Q\}} \quad (2.5.5)$$

Nel lavoro sopracitato, Ghysels et al. (2005) utilizzano tale formulazione⁸ con appena due parametri (θ_1, θ_2).

Con l’obiettivo di chiarire meglio il concetto, consideriamo l’esempio classico in cui la variabile dipendente è trimestrale e i regressori invece

⁸Gli stessi autori propongono anche altri schemi di parametrizzazione. Ad esempio quella basata sulla funzione Beta denominata “*Beta Lag*”. Per uno schema lineare privo di parametri da stimare si veda Ghysels, Rubia e Valkanov (2009).

sono campionati mensilmente ($m = 3$). Se scegliamo ad esempio di fissare $p = 12$ avremo un modello che incorpora l'ultimo anno di dati mensili.

Di conseguenza il polinomio nell'operatore ritardo è così espresso:

$$B(L^{1/3}; \theta) = \sum_{j=0}^{12} b(j; \theta) L^{j/3}$$

Ed il modello è specificato come segue:

$$y_t = \beta_0 + B(L^{1/3}; \theta) x_t^{(3)} + \varepsilon_t^{(3)} \quad \text{cioè:}$$

$$y_t = \beta_0 + \beta_1 [b(0; \theta) x_t^{(3)} + b(1; \theta) x_{t-1/3}^{(3)} + b(2; \theta) x_{t-2/3}^{(3)} + b(3; \theta) x_{t-1}^{(3)} + \dots] + \varepsilon_t^{(3)}$$

La parametrizzazione sopra descritta è nota per essere abbastanza flessibile. Per valori diversi dei parametri, i pesi collegati ai ritardi possono assumere varie forme: possono diminuire lentamente o rapidamente, o avere una "forma a gobba". Pertanto, la stima dei parametri dai dati determina automaticamente la forma dei pesi e, di conseguenza, il numero di ritardi da includere nella regressione. Dunque, i pesi dipendono dalla stima di θ .

Abbiamo quindi presentato il modello MIDAS base, il quale può essere stimato, a causa della non linearità dei polinomi ritardo, usando il metodo dei minimi quadrati non lineari⁹ (NLS da *Nonlinear Least Square*) seguendo la letteratura sui modelli a ritardi distribuiti. In questo contesto Ghyesls, Santa-Clara e Valkanov (2004) dimostrano che le regressioni MIDAS portano sempre a stime più efficienti rispetto all'approccio tipico di aggregare tutte le serie alla frequenza più bassa.

⁹Ad esempio, utilizzando gli algoritmi Broyden-Fletcher-Goldfarg-Shanno (BFGS) o Levenberg-Marquardt.

2.5.1 AR-MIDAS

Un'utile generalizzazione dei modelli MIDAS può essere quella di aggiungere nella specificazione un termine autoregressivo. Ghysels, Santa-Clara e Valkanov (2004) mostrano però che l'introduzione di ritardi della variabile dipendente creano perdite di efficienza. Aggiungendo un ritardo della variabile a bassa frequenza y_{t-1} al modello base (x mensile e y trimestrale) si ha:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/3}; \theta) x_t^{(3)} + \varepsilon_t^{(3)}. \quad (2.5.6)$$

Come evidenziato da Clements e Galvão (2008), questa strategia non è adeguata, in quanto esprimendo la (2.5.6) tramite l'inversione del polinomio nell'operatore ritardo della variabile dipendente, si avrebbe:

$$y_t = \beta_0(1 - \lambda)^{-1} + \beta_1(1 - \lambda L)^{-1} B(L^{1/3}; \theta) x_t^{(3)} + (1 - \lambda L) \varepsilon_t^{(3)} \quad (2.5.7)$$

Il polinomio su x_t è un prodotto tra il polinomio in $L^{1/3}$ e il polinomio in L . Questo prodotto genera una risposta stagionale di y_t rispetto a $x^{(3)t}$, indipendentemente dal fatto che x_t mostri un andamento stagionale. Per superare questo problema gli autori propongono l'introduzione della dinamica AR come *fattore comune*:

$$y_t = \beta_0 + \lambda y_{t-1} + \beta_1 B(L^{1/3}; \theta) (1 - \lambda L) x_t^{(3)} + \varepsilon_t^{(3)}. \quad (2.5.8)$$

Si nota che, se il termine λy_{t-1} fosse spostato a sinistra dell'uguaglianza nella (2.5.8), i *fattori comuni* verrebbero "semplificati". Questo meccanismo permette di mantenere non-stagionale la risposta di y_t rispetto a $x_t^{(3)}$. Più in generale dunque possiamo scrivere l'AR-MIDAS come:

$$y_t = \beta_0 + \lambda y_{t-d} + \beta_1 B(L^{1/m}; \theta)(1 - \lambda L)x_{t-h}^{(m)} + \varepsilon_t. \quad (2.5.9)$$

dove omettiamo l'apice sul termine di errore per semplicità di notazione. Per stimare il modello AR-MIDAS, come riportato da Clements e Galvão (2008), la procedura più utilizzata è quella di stimare il modello MIDAS standard, prendere i residui $\hat{\varepsilon}_t$ e stimare un valore iniziale per λ :

$$\hat{\lambda}_0 = \left(\sum_{t=1+h}^T \hat{\varepsilon}_{t-h}^2 \right)^{-1} \cdot \sum_{t=1+h}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-h} \quad (2.5.10)$$

Poi si costruiscono le serie:

$$y_t^* = y_t - \hat{\lambda}_0 y_{t-d} \quad (2.5.11)$$

$$x_{t-h}^{*(m)} = x_{t-h}^{(m)} - \hat{\lambda}_0 x_{t-h-d}^{(m)} \quad (2.5.12)$$

Successivamente si ottiene lo stimatore $\hat{\theta}_1$ applicando i minimi quadrati non lineari (NLS) al modello:

$$y_t^* = \beta_1 + \beta_2 B(L^{1/m}; \theta) x_{t-h}^{*(m)} + \varepsilon_t \quad (2.5.13)$$

Si ottengono quindi nuovi valori di λ ($\hat{\lambda}_1$) dai residui della regressione (2.5.13). Infine, usando come valori iniziali $\hat{\lambda}_1$ e $\hat{\theta}_1$, tramite l'algorit-

mo Broyden–Fletcher–Goldfarg–Shanno (BFGS) si stimano $\hat{\lambda}$ e $\hat{\theta}$ i quali minimizzano la somma dei quadrati dei residui.

2.5.2 Modello Unrestricted MIDAS

Foroni, Marcellino e Schumacher (2011) studiano le caratteristiche di una variante dei MIDAS che non fa ricorso ad alcuna funzione per il polinomio nell'operatore ritardo. In particolare, essi, discutono dei pro e dei contro relativi alle regressioni MIDAS con polinomi ritardo lineari ed illimitati, che non richiedono NLS, ma possono essere stimati attraverso il metodo OLS. Questo approccio viene chiamato dagli stessi U-MIDAS da *Unrestricted* MIDAS.

Una delle ragioni che può motivare l'uso dei modelli U-MIDAS nelle applicazioni macroeconomiche è che la differenza tra le frequenze di campionamento non è così elevata. Ad esempio, molti dei lavori citati, tra cui Clements e Galvão (2008, 2009) e Kuzin et al. (2011), utilizzano dati mensili per prevedere la crescita del PIL trimestrale. In tal caso, il numero di ritardi mensili, necessari per stimare i polinomi nell'operatore ritardo potrebbe non essere elevato.

Tuttavia, quando i dati finanziari entrano in gioco ci troviamo di fronte ad un aumento sensibile dei parametri da stimare, dunque possono essere preferibili restrizioni sulla forma funzionale relativa al polinomio nell'operatore ritardo. Pertanto Foroni, Marcellino e Schumacher (2011) esaminano come i modelli U-MIDAS possono essere specificati tramite un assetto dinamico lineare, e in quali condizioni i parametri del modello ad

alta frequenza sottostante sono identificati.

Supponiamo, quindi, di avere una serie y_t campionata con frequenza t (presa come riferimento), ed N variabili $x_t^{(m)}$ disponibili m volte all'interno del periodo t . Introduciamo, inoltre, il polinomio:

$$\omega(L) = \omega_0 + \omega_1 L + \dots + \omega_{k-1} L^{k-1} \quad (2.5.14)$$

Che caratterizza lo schema di aggregazione temporale. Ad esempio, se si tratta di una variabile flusso lo schema può essere come quello espresso nella (2.5.14), detto campionamento medio o somma, oppure se la variabile è stock si ha $\omega(L) = 1$, cioè *point-in-time*, in linea con quanto detto da Marcellino (1999).

Utilizzando la notazione di Foroni e Marcellino (2013), il modello U-MIDAS basato su un polinomio ritardo lineare è specificato nel seguente modo:

$$c(L^c)\omega(L)y_t = \delta_1(L^{v/m})x_{1t} + \dots + \delta_N(L^{v/m})x_{Nt} + \varepsilon_t \quad (2.5.15)$$

Dove c e v sono valori finiti che indicano l'ordine minimo di ritardo nei seguenti polinomi:

$$c(L^c) = (1 - c_1 L - c_2 L^2 - \dots - c_c L^c)$$

$$\delta_i(L^{v/m}) = (1 + \delta_{i,1} L^{1/m} + \dots + \delta_{i,v} L^{v/m}) \quad \text{con } i = 1, \dots, N$$

Gli autori fanno notare che se gli ordini dei ritardi c e v sono elevati al punto da rendere il termine d'errore incorrelato, allora tutti i parametri

nel modello U-MIDAS (2.5.15) possono essere stimati tramite il metodo OLS (con lo schema di aggregazione $\omega(L)$ noto). Generalizzando, l'ordine dei ritardi v potrebbe variare da una variabile all'altra, e v_i e c possono essere scelti tramite i criteri informativi. Infine, nel caso di U-MIDAS la componente autoregressiva può essere inclusa facilmente senza nessuna restrizione a fattore comune come in Clements e Galvão (2008).

2.5.3 Variabili esplicative multiple

A questo punto viene naturale estendere il modello MIDAS base alla presenza di più variabili esplicative. Per fare questo bisogna generalizzare il modello standard come segue:

$$y_t = \beta_0 + \sum_{i=1}^K \sum_{j=1}^L b_{ij}(L_i^{1/m}; \theta) x_t^{(m_i)} + \varepsilon_t \quad (2.5.16)$$

In questo caso consideriamo due o più variabili esplicative con due o più ritardi associati. In aggiunta, è anche possibile includere variabili esplicative con più frequenze diverse, poiché ogni regressore è modellato con la propria parametrizzazione polinomiale. Ad esempio, la crescita trimestrale del PIL può essere spiegata non solo da indicatori mensili ma, anche da variabili finanziarie settimanali o giornaliere, con le variabili esplicative quindi campionate a due o più frequenze diverse.

2.5.4 Modelli MIDAS multivariati

La regressione (2.5.16) può essere generalizzata ancor di più verso una specificazione multivariata:

$$Y_t = \mathcal{B}_0 + \sum_{i=1}^K \sum_{j=1}^L \mathcal{B}_{ij}(L_i^{1/m}; \theta) x_t^{(m_i)} + \varepsilon_t \quad (2.5.17)$$

Dove Y, X e ε sono processi vettoriali a n dimensioni, \mathcal{B}_0 è un vettore ($n \times 1$) e le \mathcal{B}_{ij} sono matrici ($n \times n$) di polinomi¹⁰. Il problema principale di questi modelli è come gestire l'aumento dei parametri da stimare, in un contesto multivariato. Foroni e Marcellino (2013) indicano che un'approccio potrebbe essere quello di considerare tutti i parametri fuori dalla diagonale controllati da un polinomio, mentre gli elementi sulla diagonale da un secondo polinomio. Le restrizioni sui parametri però vanno valutate volta per volta in base all'applicazione oggetto di studio.

Considerare le regressioni MIDAS da un punto di vista multivariato consente di studiare temi tipici dei contesti multivariati come la causalità di Granger (1969) o la cointegrazione. È proprio per investigare questo tipo di caratteristiche che nella letteratura si sono sviluppati modelli in grado di gestire la presenza di dataset a frequenze miste in un contesto multivariato.

Uno degli approcci più utilizzati in questo ambito è, oltre ai MIDAS, quello relativo ai modelli cosiddetti *Mixed Frequency VAR* (MF-VAR), il cui obiettivo è catturare i movimenti comuni nelle serie considerate e nel contempo utilizzare le informazioni contenute nel dataset *mixed-frequency*.

¹⁰O equivalentemente, polinomi di matrici.

2.6 Mixed Frequency Vector Autoregressive

Uno degli approcci più interessanti in letteratura per trattare le serie a frequenza mista da un punto di vista multivariato, è quello proposto da Zadrozny (1988) per stimare direttamente un modello VARMA(p,q) campionato in frequenze diverse. Tenendo conto delle relazioni flussi-stock tra le variabili, l'autore, deriva la rappresentazione spazio-stato implicita del modello, e ha mostrato come utilizzarla per calcolare la funzione di probabilità gaussiana con l'aiuto del filtro di Kalman. Inoltre, egli propone di trattare tutte le serie come generate alla frequenza più alta, ma alcune di esse non vengono osservate interamente. Le variabili che sono disponibili solo alla frequenza più bassa sono trattate come se fossero generate da un processo latente ad alta frequenza. Nonostante questo tentativo iniziale, i modelli VAR a frequenza mista hanno ricevuto solo recentemente più attenzione, con l'espandersi del concetto di *nowcasting*. Banbura, Giannone e Reichlin (2010) definiscono il *nowcasting* come la previsione del presente, del futuro molto prossimo e del passato molto recente. La maggior parte della letteratura sui Mixed Frequency Vector Autoregressive (MF-VAR) si basa sull'approccio di Zadrozny (1988).

L'MF-VAR è dunque un processo vettoriale autoregressivo specificato in modo da poter gestire dataset campionati con frequenze diverse. Si consideri un processo $y_{1,t}$ di dimensione N_1 a bassa frequenza, ed un processo $y_{2,t}$ di N_2 dimensioni ad alta frequenza, rilevato m volte durante il periodo di riferimento di $y_{1,t}$. Inoltre si consideri il vettore $y_t = [y'_{1,t}, y'_{2,t}]'$, di dimensione ($N = N_1 + N_2$). L'MF-VAR studia l'interazione dinamica

tra le variabili osservate ad alta frequenza $y_{2,t}$, e il processo latente, ad alta frequenza $y_{1,t}^*$, delle variabili osservate alla frequenza più bassa $y_{1,t}$.

Una delle specificazioni più usate in letteratura per quanto riguarda i modelli MF-VAR, è quella proposta da Mariano e Murasava (2003, 2010). Essa presuppone l'utilizzo della rappresentazione spazio-stato del modello e del filtro di Kalman (si veda par. 2.1). Questa specificazione si pone l'obiettivo di studiare sia il processo latente $y_{1,t}^*$ sottostante alla variabile a bassa frequenza $y_{1,t}$, sia il Data Generating Process (DGP) alla frequenza più alta $y_t^* = [y_{1,t}^*, y_{2,t}^*]'$ relativo al vettore $y_t = [y_{1,t}, y_{2,t}]'$.

2.6.1 Specificazione di Mariano e Murasawa

L'idea di Mariano e Murasawa (2003, 2010) è dunque quella di modellare gli MF-VAR attraverso la rappresentazione spazio-stato. Considerando le variabili mensili e trimestrali, gli autori sono interessati all'analisi delle dinamiche tra le variabili mensili osservate e i processi latenti (mensili) sottostanti alle variabili trimestrali.

Seguendo l'esempio da loro presentato andiamo a introdurre la specificazione MF-VAR, che può essere suddivisa in due parti: nella prima si va a specificare la natura autoregressiva del processo ad alta frequenza non completamente osservato (y_t^*); nella seconda parte si definisce la relazione tra le variabili a bassa frequenza (y_t) e la loro controparte ad alta frequenza non osservabile (y_t^*). Partendo dal secondo punto, si va ad investigare la

relazione tra i due vettori:

$$y_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} \quad \text{e} \quad y_t^* = \begin{bmatrix} y_{1,t}^* \\ y_{2,t} \end{bmatrix}.$$

Tale relazione può essere scritta come segue:

$$y_t - \mu = G(L)(y_t^* - \mu^*) \quad \text{con}$$

$$E(y_t) = \mu \quad \text{e} \quad E(y_t^*) = \mu^*$$

Dove $G(L)$ è il polinomio di ordine k che collega le osservazioni a bassa frequenza del vettore $y_{1,t}$, alle realizzazioni ad alta frequenza non osservate $y_{1,t}^*$. Lo schema di aggregazione temporale (si veda par. 2.4) determina la struttura del polinomio $G(L)$. Inoltre, se per semplicità assumiamo che $\mu = \mu^* = 0$, possiamo esprimere la relazione tra y_t e y_t^* nel seguente modo:

$$y_t = G(L)y_t^* \tag{2.6.1}$$

Successivamente, specifichiamo il processo autoregressivo y_t^* non completamente osservabile:

$$y_t^* = \Phi_1 y_{t-1}^* + \dots + \Phi_p y_{t-p}^* + v_t \tag{2.6.2}$$

$$\Phi(L)y_t^* = v_t$$

Con $v_t \sim WN(0, \Sigma)$, e $\Phi(L) = (I - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p)$.

Il modello MF-VAR(p) è quindi ottenuto considerando un sistema

spazio-stato tra il vettore osservabile y_t e il processo ad alta frequenza sottostante y_t^* , dove l'equazione di stato è rappresentata dal VAR(p) specificato nella (2.6.2) e l'equazione delle osservazioni è invece data dalla (2.6.1). Per riportare il modello appena specificato nella forma (2.1.1)-(2.1.2) presentata nel paragrafo 2.1, dobbiamo avere che:

$$\alpha_t = \underbrace{\begin{bmatrix} y_t^* \\ y_{t-1}^* \\ \vdots \\ y_{t-k-1}^* \end{bmatrix}}_{N(k+1) \times 1}, \quad H = \underbrace{\begin{bmatrix} G_0 & G_1 & \dots & G_k \end{bmatrix}}_{N \times N(k+1)} \quad (2.6.3)$$

Dove, come abbiamo detto precedentemente, k è l'ordine del polinomio $G(L)$ e G_i , con $i = 1, \dots, k$, sono le matrici dei pesi relative alle variabili y_{t-i} nella relazione definita dalla (2.6.1).

In precedenza abbiamo detto che l'equazione di stato è determinata dalla dinamica VAR(p) specificata nella (2.6.2), pertanto per riportare tale specificazione nella forma della (2.1.1), scriviamo il processo nella *companion form* dove, in questo caso, si ha:

$$\underbrace{\begin{bmatrix} y_t^* \\ y_{t-1}^* \\ \vdots \\ y_{t-p+1}^* \end{bmatrix}}_{Np \times 1} = \underbrace{\begin{bmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ I_N & 0_{N \times N} & \dots & 0_{N \times N} \\ 0_{N \times N} & \ddots & \dots & \vdots \\ \vdots & \dots & I_N & 0_{N \times N} \end{bmatrix}}_{Np \times Np} \begin{bmatrix} y_{t-1}^* \\ y_{t-2}^* \\ \vdots \\ y_{t-p}^* \end{bmatrix} + v_t \quad (2.6.4)$$

Che è equivalente a scrivere:

$$\begin{aligned} y_t^* &= Fy_{t-1}^* + v_t \quad \text{oppure:} \\ \alpha_t &= F\alpha_{t-1} + v_t \end{aligned}$$

Con $v_t \sim N(0, \Sigma)$. Dove F è la matrice companion del VAR(p). A questo punto si nota che le dimensioni del vettore di stato nelle equazioni (2.6.3) e (2.6.4) e la specificazione delle matrici nel sistema spazio-stato sono connessi a due aspetti: in primis l'ordine dei ritardi che si determina per il processo autoregressivo nell'equazione di stato; in secondo luogo il numero dei ritardi coinvolti nello schema di aggregazione temporale che determina l'ordine del polinomio $G(L)$ nell'equazione delle osservazioni. Seguendo Emili (2017) proponiamo un'esempio in cui l'ordine del VAR è fissato a $p = 2$ e lo schema di aggregazione è di tipo *average sampling*, cioè:

$$y_{1,t} = \frac{1}{3}(y_{1,t}^* + y_{1,t-1}^* + y_{1,t-2}^*) \quad (2.6.5)$$

Di conseguenza, $y_{1,t}$ comprende $m = k + 1 = 3$ realizzazioni di $y_{1,t}^*$. Perciò l'ordine del polinomio $G(L)$ è $k = 2$:

$$G(L) = \begin{bmatrix} \frac{1}{3}I_{N_1} & 0 \\ 0 & I_{N_2} \end{bmatrix} + \begin{bmatrix} \frac{1}{3}I_{N_1} & 0 \\ 0 & 0 \end{bmatrix} L + \begin{bmatrix} \frac{1}{3}I_{N_1} & 0 \\ 0 & 0 \end{bmatrix} L^2 \quad (2.6.6)$$

Ricordiamo che N_1 è il numero delle variabili a bassa frequenza mentre N_2 è il numero di quelle ad alta frequenza, ed $N = N_1 + N_2$. In aggiunta dalla (2.6.3) si evince che la dimensione del vettore di stato α_t è di $3N \times 1$.

Di conseguenza la matrice H non è altro che:

$$H = \begin{bmatrix} G_0 & G_1 & G_2 \\ \frac{1}{3}I_{N_1} & 0_{N_1 \times N_2} & \vdots & \frac{1}{3}I_{N_1} & 0_{N_1 \times N_2} & \vdots & \frac{1}{3}I_{N_1} & 0_{N_1 \times N_2} \\ 0_{N_2 \times N_1} & I_{N_2} & \vdots & 0_{N_2 \times N_1} & 0_{N_2 \times N_2} & \vdots & 0_{N_2 \times N_1} & 0_{N_2 \times N_2} \end{bmatrix} \quad (2.6.7)$$

Allora l'equazione delle osservazioni può essere espressa come segue:

$$y_t = \begin{bmatrix} G_0 & G_1 & G_2 \end{bmatrix} \begin{bmatrix} y_t^* \\ y_{t-1}^* \\ y_{t-2}^* \end{bmatrix} \quad (2.6.8)$$

Dal lato dell'equazione di stato (2.6.2), se l'ordine del VAR è $p = 2$, il vettore α_t è di dimensione $2N \times 1$. Il processo in companion form è:

$$\begin{bmatrix} y_t^* \\ y_{t-1}^* \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 \\ I_N & 0_{N \times N} \end{bmatrix} \begin{bmatrix} y_{t-1}^* \\ y_{t-2}^* \end{bmatrix} + v_t \quad (2.6.9)$$

Le dimensioni del vettore di stato $\alpha_t' = [y_t^*, y_{t-1}^*]'$ sono diverse nelle equazioni (2.6.8) e (2.6.9). Per superare questo problema, Mariano e Murasawa (2010) suggeriscono di indentificare due casi. Il primo si presenta quando $p \leq k + 1$, cioè l'ordine del VAR è strettamente minore dell'ordine del polinomio $G(L)$. Il secondo invece quando al contrario si ha $p > k + 1$.

Nel primo caso ($p \leq k + 1$), il numero degli elementi del vettore di stato α_t è determinato dall'ordine k del polinomio $G(L)$, come mostrato

nella (2.6.3). Le matrici all'interno sistema spazio-stato si modificano come segue:

$$F = \underbrace{\begin{bmatrix} \Phi_1 & \dots & \Phi_p & 0_{N \times N(k+1-p)} \\ & I_{Nk} & & 0_{Nk \times N} \end{bmatrix}}_{N(k+1) \times N(k+1)} \quad (2.6.10)$$

$$H = \underbrace{\begin{bmatrix} G_0 & G_1 & \dots & G_k \end{bmatrix}}_{N \times N(k+1)}$$

Nel secondo caso ($p > k + 1$) invece le dimensioni del vettore di stato e delle matrici nel sistema spazio-stato sono determinate dall'ordine di ritardo p del VAR. In particolare, il vettore di stato ha dimensioni come quelle evidenziate nella (2.6.4) e le matrici sono definite in tal modo:

$$F = \underbrace{\begin{bmatrix} \Phi_1 & \dots & \Phi_{p-1} & \Phi_p \\ & I_{N(p-1)} & & 0_{N(p-1) \times N} \end{bmatrix}}_{Np \times Np} \quad (2.6.11)$$

$$H = \underbrace{\begin{bmatrix} G_0 & \dots & G_k & 0_{N \times N(p-k-1)} \end{bmatrix}}_{N \times Np}$$

Arrivati a questo punto, abbiamo presentato il modello MF-VAR(p) nella forma spazio-stato:

$$\alpha_t = F\alpha_{t-1} + v_t \quad (2.6.12)$$

$$y_t = H\alpha_t + w_t \quad (2.6.13)$$

La suddetta prevede la specificazione di un processo VAR per le serie ad alta frequenza non osservabili completamente (equazione di stato (2.6.12)). Ed inoltre l'individuazione di una relazione tra il vettore delle serie osservate alla frequenza più bassa e il processo ad alta frequenza sottostante a tali variabili (equazione delle osservazioni (2.6.13)).

Il problema successivo da trattare è quello relativo allo studio dei valori inosservati del processo latente sottostante alle serie a bassa frequenza. Se consideriamo l'esempio classico tra variabili mensili e trimestrali, osservate ogni terzo mese del trimestre, la questione è quella di come trattare i valori di $y_{1,t}$ relativi al primo e secondo mese.

Seguendo Mariano e Murasawa (2003,2010), inseriamo elementi nulli al posto delle osservazioni mancanti e riscriviamo l'equazione delle osservazioni come se fossero estrazioni casuali da una variabile casuale $N(0, 1)$ indipendente dai parametri del modello, in modo che il filtro di Kalman "salti" le osservazioni mancanti. Dunque, per ogni t , definiamo la variabile $y_{1,t}^+$ tale per cui:

$$y_{1,t}^+ = \begin{cases} y_{1,t} & \text{se } y_{1,t} \text{ è osservabile} \\ z_t & \text{altrove} \end{cases} \quad (2.6.14)$$

Dove $z_t \sim N(0, I_{N_1})$ e tutte le realizzazioni del processo sono zero. L'equazione delle osservazioni può essere riscritta, più in generale, come:

$$\begin{bmatrix} y_{1,t}^+ \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_{1,t} \\ \mu_2 \end{bmatrix} + \begin{bmatrix} H_{1,t} \\ H_2 \end{bmatrix} \alpha_t + \begin{bmatrix} D_t \\ 0 \end{bmatrix} w_t \quad (2.6.15)$$

Dove:

$$\mu_{1,t} = \begin{cases} \mu_1 & \text{se } y_{1,t} \text{ è osservabile} \\ 0 & \text{altrove} \end{cases} \quad (2.6.16)$$

$$H_{1,t} = \begin{cases} H_1 & \text{se } y_{1,t} \text{ è osservabile} \\ 0 & \text{altrove} \end{cases} \quad (2.6.17)$$

$$D_{1,t} = \begin{cases} 0 & \text{se } y_{1,t} \text{ è osservabile} \\ I_{N_1} & \text{altrove} \end{cases} \quad (2.6.18)$$

Per semplicità di notazione poniamo:

$$y_t^+ = \begin{bmatrix} y_{1,t}^+ \\ y_{2,t} \end{bmatrix}, \quad \mu_t = \begin{bmatrix} \mu_{1,t} \\ \mu_2 \end{bmatrix}, \quad H_t = \begin{bmatrix} H_{1,t} \\ H_2 \end{bmatrix}, \quad D_t = \begin{bmatrix} D_t \\ 0 \end{bmatrix} \quad (2.6.19)$$

Così otteniamo un modello spazio-stato *variabile nel tempo*, cioè:

$$\alpha_t = F\alpha_{t-1} + v_t \quad (2.6.20)$$

$$y_t^+ = \mu_t + H_t\alpha_t + D_t w_t \quad \text{con} \quad (2.6.21)$$

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim N(0, I_{N+N_1})$$

Per stimare il modello sopraindicato, Mariano e Murasawa (2003,2010), propongono di utilizzare l'algoritmo *Expectation-maximization* (EM), il quale è un meccanismo iterativo per trovare la stima di massima verosimiglianza quando il modello dipende da variabili latenti inosservabili. Questa scelta

iniziale deriva dal fatto che la stima di un modello spazio-stato gaussiano lineare con un metodo quasi-Newton potrebbe non riuscire a convergere quando i valori iniziali non sono adeguati.

Assumiamo dunque per semplicità che $\mu_t = \underline{0}$, $\Phi = [\Phi_1, \dots, \Phi_p]$, $\phi = \text{vec}(\Phi')$ e $\theta = [\phi, \text{vec}(\Sigma)]$. Dove Σ è la matrice delle varianze e covarianze dei residui del processo VAR(p). Inoltre si ha che $Y_0^* = 0$, e per ogni t :

$$Y_t^* = [y_1^*, \dots, y_t^*] \quad \text{e per } t \geq 0: \quad (2.6.22)$$

$$S_t = [\alpha_0, \dots, \alpha_t] \quad (2.6.23)$$

Successivamente, introducendo le seguenti matrici di selezione, le cui dimensioni sono determinate dai casi descritti dalle (2.6.10) e (2.6.11):

$$\begin{cases} \mathbf{U} = \begin{bmatrix} I_N & 0_{N \times Nk} \end{bmatrix}, & \mathbf{V} = \begin{bmatrix} I_{Np} & 0_{Np \times N(k-p)} \end{bmatrix} & \text{se } p \leq k \text{ (CASO 1)} \\ \mathbf{U} = \begin{bmatrix} I_N & 0_{N \times N(p-1)} \end{bmatrix}, & \mathbf{V} = \begin{bmatrix} I_{Np} \end{bmatrix} & \text{se } p > k \text{ (CASO 2)} \end{cases}$$

Si può scrivere il Processo VAR(p) nel modo:

$$\begin{aligned} \mathbf{U}\alpha_t &= y_t^* \\ &= \Phi_1 y_{t-1}^* + \dots + \Phi_p y_{t-p}^* + w_t \\ &= \Phi \mathbf{V}\alpha_{t-1} + w_t \\ \mathbf{U}\alpha_t - \Phi \mathbf{V}\alpha_{t-1} &= w_t \end{aligned} \quad (2.6.24)$$

Di conseguenza si è in grado di poter scrivere la funzione di log-verosimiglianza

di θ dato (Y_T^*, S_T) nella maniera:

$$\begin{aligned} l(\theta; Y_T^*, S_T) &= \\ &= -\frac{NT}{2} \ln 2\pi - \frac{T}{2} \ln \det(\Sigma) - \frac{1}{2} \sum_{t=1}^T (U\alpha_t - \Phi V\alpha_{t-1})' \Sigma^{-1} (U - \Phi V\alpha_{t-1}) \end{aligned}$$

A questo punto il primo passo per ottenere le stime EM è quello di eseguire il filtro di Kalman sulla rappresentazione spazio-stato specificata in precedenza.

Come visto nel paragrafo 2.1 si ha bisogno di valori iniziali per effettuare questa operazione. Dunque se la serie α_t è stazionaria i valori iniziali possono essere:

$$\begin{aligned} \hat{\alpha}_{1|0} &= 0 \\ \text{vec}(P_{1|0}) &= (I_{r^2} - F \otimes F)^{-1} \text{vec}(Q) \end{aligned}$$

Dove r è la dimensione del vettore di stato. La seconda equazione implica l'inversione di una matrice potenzialmente di grandi dimensioni, il che può causare problemi computazionali. Quindi Mariano e Murasawa (2003,2010) assumono $\alpha_0 = 0$, che implica:

$$\hat{\alpha}_{1|0} = 0 \quad \text{e:} \quad (2.6.25)$$

$$P_{1|0} = Q \quad (2.6.26)$$

Lo stimatore risultante è asintoticamente equivalente all'esatto stimatore di massima verosimiglianza.

Successivamente il filtro di Kalman calcola le serie $\alpha_{t|t}$, $P_{t|t}$ e le previsioni $\alpha_{t+1|t}$, $P_{t+1|t}$ necessarie per utilizzare l'algoritmo EM.

Infatti il passo successivo è quello di ottenere le stime *smorzate* $\alpha_{t|T}$ e il relativo errore quadratico medio indicato con $P_{t|T}$, attraverso l'algoritmo di *smoothing* (si veda il paragrafo 2.1.4 e Hamilton (1994, par. 13.6)). In seguito si ricavano le stime $\hat{\phi}$ e $\hat{\Sigma}$, cioè del vettore θ , calcolando il valore atteso condizionale della funzione *score* della log-verosimiglianza e ponendolo uguale a zero.

Riassumendo per stimare un modello MF-VAR(p) attraverso l'algoritmo EM, come propongono Mariano e Murasawa (2003,2010), si procede operativamente nel seguente modo:

1. Scegliere un valore iniziale per $\theta^{(0)}$;
2. Eseguire il filtro di Kalman;
3. Expectation Step (E-step): calcolare le stime smorzate $\alpha_{t|T}$ e $P_{t|T}$ attraverso lo smoothing;
4. Maximization Step (M-step): calcolare $[\hat{\phi}, \hat{\Sigma}]$ e usarlo come $\theta^{(1)}$;
5. Iterare fino alla convergenza.

Se l'algoritmo raggiunge la convergenza nel vettore $\hat{\theta}$ ci sono le stime dei parametri del VAR(p) a frequenza mista e la relativa matrice varianza e covarianza degli errori. L'algoritmo di smoothing ci dà $E(y_{1,t}^* | Y_T^+)$, cioè le stime smorzate della serie latente, ad alta frequenza, sottostante quella alla frequenza più bassa.

Gli autori, inoltre, precisano che se invece l'algoritmo *rallenta significativamente in prossimità del massimo* è preferibile passare ad un metodo di ottimizzazione quasi-Newton. Questo significherebbe usare le stime EM

come buoni punti iniziali per una stima quasi-Newton in quanto, come abbiamo detto in precedenza, esso ha problemi di convergenza se gli starting points non sono adeguati.

2.6.2 MIDAS Vector AutoRegressive

Ghysels (2016) presenta un approccio alternativo a quello esposto in precedenza, per l'analisi dei processi VAR a frequenza mista. L'autore considera l'idea di specificare un VAR a frequenza mista senza basarsi su rappresentazioni di processi o shock latenti.

La metodologia più utilizzata è infatti quella che utilizza la rappresentazione spazio-stato che quindi richiede procedure di filtraggio. Questi modelli sono definiti dallo stesso Ghysels: "*parameter driven*". Mentre il nuovo approccio da lui introdotto è detto "*observation driven*", in quanto è formulato esclusivamente con dati osservabili. Tale approccio è più in linea con la letteratura MIDAS (si veda par. 2.5) ed infatti ci riferiremo a quest'ultimo come MIDAS-VAR.

L'idea generale di Ghysels è quella di specificare un processo VAR a bassa frequenza per il vettore x_t ottenuto "*impilando*" le m osservazioni delle variabili ad alta frequenza, disponibili durante il periodo a bassa frequenza, con le serie già disponibili alla frequenza più bassa. Il risultato è un vettore di dimensione $N_S \times 1$ dove $N_S = N_1 + mN_2$ in cui N_1 è il numero di serie a bassa frequenza, N_2 il numero di quelle ad alta frequenza ed m è il numero di osservazioni ad alta frequenza presenti nell'intervallo a bassa frequenza.

In altre parole, nel suo articolo Ghysels (2016), scompone ogni variabile ad alta frequenza in m variabili ad una frequenza più bassa. Ad esempio, se si sta studiando l'interazione tra una variabile mensile e una trimestrale, quella mensile sarà scomposta in tre variabili trimestrali. Successivamente l'autore modella le variabili risultanti con quelle originariamente disponibili alla frequenza più bassa.

Se poniamo dunque che $x_{H,t}^{(m)}$ è la variabile ad alta frequenza e $x_{L,t}$ è quella a bassa frequenza si ha che il MIDAS-VAR(p) è espresso come:

$$\begin{bmatrix} x_{H,t}^{(1)} \\ \vdots \\ x_{H,t}^{(m)} \\ x_{L,t} \end{bmatrix} = A_0 + \sum_{j=1}^p A_j \begin{bmatrix} x_{H,t-1}^{(1)} \\ \vdots \\ x_{H,t-1}^{(m)} \\ x_{L,t-1} \end{bmatrix} + \varepsilon_t \quad (2.6.27)$$

Dove t è l'indice di riferimento al periodo di bassa frequenza e le A_j sono matrici di dimensione $(N_1 + mN_2) \times (N_1 + mN_2)$. La generica matrice A_j può essere espressa nel modo seguente:

$$A_j = \begin{bmatrix} a_j^{1,1} & \dots & a_j^{1,m} & a_j^{1,L} \\ \vdots & & \dots & \vdots \\ a_j^{m,1} & \dots & a_j^{m,m} & a_j^{m,L} \\ a_j^{L,1} & \dots & a_j^{L,m} & a_j^{L,L} \end{bmatrix} \quad (2.6.28)$$

Si ottiene, dunque, un VAR alla frequenza più bassa che però tiene conto delle informazioni disponibili nelle serie a maggiore frequenza.

Il vantaggio di questo approccio è che si possono utilizzare tutti gli

strumenti propri dei processi VAR standard. Ad esempio, possono essere stimati tramite il metodo OLS, si può applicare la scomposizione di Cholesky e calcolare la funzione di risposta d'impulso strutturale.

Uno dei problemi principali di questa specificazione è la numerosità dei parametri da stimare, il cui numero dipende da m (sono esattamente: $N_1 \times m \cdot N_2 \times p$). Nel caso in cui si considerano variabili macroeconomiche, dove m non è *troppo grande*, il numero dei parametri da stimare può essere ancora gestibile con tecniche di stima equivalenti a quelle utilizzate nei VAR standard. Ma se in gioco ci sono variabili finanziarie (ad esempio con frequenza giornaliera) si può incorrere in un incremento eccessivo dei parametri da stimare, con una conseguente perdita di efficienza delle tecniche di stima.

Per risolvere questo problema, Ghysels (2016), propone di utilizzare le parametrizzazioni viste nei modelli MIDAS. Così si può raggiungere il grado di parsimonia spesso necessario nelle applicazioni empiriche. Tale soluzione però, aggiunge l'autore, può portare ad una perdita di informazioni che invece sarebbero disponibili nei modelli senza restrizioni.

Con l'intento di fare ulteriore chiarezza, introduciamo un esempio relativo ad una specificazione MIDAS-VAR(p). Consideriamo due serie: $x_{H,t}^{(m)}$ mensile e $x_{L,t}$ trimestrale, di conseguenza si ha $m = 3$. Dunque il modello MIDAS-VAR(1) à la Ghysels (2016) per il vettore x_t , è specificato come

segue:

$$\begin{bmatrix} x_{H,t}^{(1)} \\ x_{H,t}^{(2)} \\ x_{H,t}^{(3)} \\ x_{L,t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} x_{H,t-1}^{(1)} \\ x_{H,t-1}^{(2)} \\ x_{H,t-1}^{(3)} \\ x_{L,t-1} \end{bmatrix} + \varepsilon_t \quad (2.6.29)$$

si nota che da un sistema bivariato con $N_1 = N_2 = 1$ si passa ad un VAR a quattro dimensioni ($N_1 + mN_2 = 4$).

Se esaminiamo l'ultima equazione del modello (2.6.29), si riconosce che è molto simile ad una regressione U-MIDAS (par. 2.5.2), in cui si utilizzano i dati ad alta frequenza per spiegare le serie a bassa frequenza:

$$x_{L,t} = a_{41}x_{H,t-1}^{(1)} + a_{42}x_{H,t-1}^{(2)} + a_{43}x_{H,t-1}^{(3)} + a_{44}x_{L,t-1} + \varepsilon_{4,t} \quad (2.6.30)$$

Al contrario la prima equazione, come la seconda e la terza, misura anche l'impatto della serie a bassa frequenza su quella a frequenza più alta:

$$x_{H,t}^{(1)} = a_{11}x_{H,t-1}^{(1)} + a_{12}x_{H,t-1}^{(2)} + a_{13}x_{H,t-1}^{(3)} + a_{14}x_{L,t-1} + \varepsilon_{1,t} \quad (2.6.31)$$

Bacchiocchi, Bastianin, Missale e Rossi (2018) fanno notare come l'equazione (2.6.31) è direttamente collegata con il modello *Reverse Unrestricted* MIDAS (RU-MIDAS) presentato da Forni Guérin e Marcellino (2018) che, similmente ai modelli U-MIDAS, può essere stimato attraverso il metodo OLS.

In conclusione, se il vettore impilato è $x_t = [x_{H,t}^{(1)}, x_{H,t}^{(2)}, x_{H,t}^{(3)}, x_{L,t}]'$, un

MIDAS-VAR(p) può essere specificato come segue:

$$x_t = A_1 x_{t-1} + \dots + A_p x_{t-p} + \varepsilon_t \quad (2.6.32)$$

In cui il termine d'errore ε_t è un white noise con $E(\varepsilon_t \varepsilon_t') = \Sigma$. Questa notazione ci permette di arrivare ad una forma più compatta per esprimere il modello:

$$A(L)x_t = \varepsilon_t \quad (2.6.33)$$

Dove L è l'operatore ritardo a bassa frequenza e $A(L) = I_{N_S} - \sum_{i=1}^p A_i L^i$. Successivamente, una volta stimato il modello, l'attenzione si sposta sull'interpretazione e sull'utilizzo di quest'ultimo. Ghysels (2016) propone di analizzare la forma strutturale del processo MIDAS-VAR, sia per quanto riguarda la previsione delle serie, sia per la *policy analysis*.

2.6.3 VAR strutturali a frequenze miste

In un primo momento, la previsione dal modello MIDAS-VAR sembra semplice, poichè esso è essenzialmente un VAR standard. Tuttavia l'unità di tempo di riferimento è quella a bassa frequenza, mentre si può voler prevedere alla frequenza più alta. Ad esempio, aggiornare la previsione per x_t ogni mese del trimestre. Ciò richiede l'utilizzo della forma strutturale del processo.

Per chiarire il perché di questa necessità, partiamo dall'esempio insito nell'equazione (2.6.29), in cui il nostro obiettivo è quello di effettuare una previsione per il terzo mese della variabile ad alta frequenza $x_{H,t}^{(3)}$. È chiaro

che per ottenere una previsione che sia più precisa possibile è auspicabile l'utilizzo di tutta l'informazione disponibile al tempo t . Dall'equazione (2.6.31) si evince però che, in una specificazione di quel tipo, non vengono utilizzati il primo ed il secondo mese del trimestre t i quali sono i ritardi ad alta frequenza della variabile che stiamo cercando di prevedere.

Dunque, per considerare anche i ritardi ad alta frequenza per la previsione attraverso il MIDAS-VAR, si utilizza la rappresentazione strutturale in quanto essa descrive le correlazioni contemporanee tra le variabili presenti nel modello. Pertanto, definiamo C come la matrice triangolare inferiore ottenuta tramite la scomposizione di Cholesky operata sulla matrice delle varianze e covarianze, che nell'esempio considerato sarebbe:

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ c_{21} & 1 & 0 & 0 \\ c_{31} & c_{32} & 1 & 0 \\ c_{41} & c_{42} & c_{43} & 1 \end{bmatrix} \quad (2.6.34)$$

Con $\Sigma = CC'$.

Moltiplicando la matrice (2.6.34) all'equazione del MIDAS-VAR(1) espressa nella (2.6.29), otteniamo la rappresentazione strutturale a cui spesso ci si riferisce come MIDAS-SVAR:

$$Cx_t = CA_{t-1} + u_t \quad (2.6.35)$$

O, equivalentemente:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ c_{21} & 1 & 0 & 0 \\ c_{31} & c_{32} & 1 & 0 \\ c_{41} & c_{42} & c_{43} & 1 \end{bmatrix} \begin{bmatrix} x_{H,t}^{(1)} \\ x_{H,t}^{(2)} \\ x_{H,t}^{(3)} \\ x_{L,t} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{bmatrix} \begin{bmatrix} x_{H,t-1}^{(1)} \\ x_{H,t-1}^{(2)} \\ x_{H,t-1}^{(3)} \\ x_{L,t-1} \end{bmatrix} + u_t \quad (2.6.36)$$

Dove $u_t = C\varepsilon_t$ è il vettore degli errori strutturali e $\alpha_{ij} = [CA]_{ij}$. Ad esempio, se si osserva la terza riga della l'equazione di cui sopra l'equazione per $x_{H,t}^{(3)}$ diventa:

$$x_{H,t}^{(3)} = -c_{31}x_{H,t}^{(1)} - c_{32}x_{H,t}^{(2)} + \alpha_{31}x_{H,t-1}^{(1)} + \alpha_{32}x_{H,t-1}^{(2)} + \alpha_{33}x_{H,t-1}^{(3)} + \alpha_{34}x_{L,t-1} + u_{3,t} \quad (2.6.37)$$

Che coinvolge i ritardi ad alta frequenza.

Per quanto riguarda l'analisi della politica attraverso la funzione di risposta d'impulso, c'è poco da aggiungere. Il modello utilizzato è pressoché lo stesso. Richiamiamo brevemente la teoria alla base della funzione di risposta d'impulso per applicarla al nostro caso specifico, considerato finora. Ad esempio, sappiamo che per un modello VAR(1) le matrici dei coefficienti a media mobile (MA) contengono le risposte d'impulso del sistema e l'i-esimo coefficiente della rappresentazione MA è uguale ad A^i . Inoltre poiché la matrice Σ può essere scomposta come CC' grazie alla scomposizione di Cholesky, la matrice che descrive la risposta d'impulso strutturale i periodi dopo lo shock è data da:

$$\Theta_i = A^i C \quad (2.6.38)$$

Per concludere, Foroni e Marcellino (2014,2016) hanno studiato vari aspetti

dei VAR strutturali a frequenze miste, dall'identificazione all'intensità dell'aggregation bias. Gli autori (2014) hanno utilizzato la specificazione MF-VAR di Mariano e Murasawa (2010) per stimare un modello *dynamic stochastic general equilibrium* (DSGE) con dati a frequenza mista. Uno studio più generale sugli MF-SVAR è quello presentato dagli stessi Foroni e Marcellino (2016), dove considerano il semplice caso di un VAR(1) ed analizzano la relazione tra rappresentazione ad alta e bassa frequenza del sistema, facendo riferimento a due schemi di aggregazione: *point in time* e *sum sampling*. Facendo questo hanno dimostrato che impiegando una ragionevole procedura di stima, l'analisi dei dati a frequenza mista allevia il già citato aggregation bias.

2.7 Un confronto tra i diversi metodi

Finora abbiamo visto alcuni dei metodi che sono stati proposti in letteratura per trattare i dati a diversa frequenza. In generale c'è accordo nel dire che lo sfruttamento dei dati a frequenza mista è importante per il *nowcasting* e per le previsioni a breve termine. Ora proviamo a riassumere i vantaggi e gli svantaggi dei diversi metodi, confrontando le loro caratteristiche principali.

Al momento l'aggregazione temporale è ancora la tecnica predominante nelle applicazioni empiriche: tutti i dati sono campionati alla stessa (bassa) frequenza. Con questa procedura però potenziali informazioni utili sono scartate.

Uno degli approcci per trattare dati a frequenza mista presentati in

questo elaborato fa riferimento ai modelli Mixed Data Sampling (MIDAS). Inizialmente le regressioni MIDAS sono state applicate a dati finanziari, quindi il problema principale con cui gli studiosi si sono dovuti confrontare, è stato quello dell'aumento dei parametri da stimare che poteva portare a problemi computazionali. Ghysels et al. (2004) presentano una soluzione al potenziale problema della proliferazione dei parametri da stimare, attraverso una parsimoniosa specificazione del polinomio ritardo.

Recentemente le MIDAS sono state implementate per investigare e prevedere variabili macroeconomiche. In tali contesti è possibile anche non imporre alcuna restrizione, in quanto il numero dei parametri da stimare può rimanere limitato.

Un secondo filone di ricerca si è basato sull'utilizzo della rappresentazione spazio-stato per interagire con dati a frequenza diversa attraverso dei sistemi VAR (MF-VAR). L'idea sottostante l'assetto spazio-stato è quella di trattare la variabile a bassa frequenza come una a frequenza più alta con osservazioni mancanti. L'uso del filtro di Kalman permette di ottenere delle stime di queste osservazioni mancanti. Gli svantaggi principali dell'MF-VAR derivano dalla sua complessità computazionale, la quale incrementa con il numero delle variabili coinvolte nel modello. Di conseguenza spesso solo modelli con un ristretto numero di variabili possono essere implementati.

Ghysels (2016) in un recente articolo¹¹ propone un approccio ai VAR a frequenze miste che non fa riferimento alla rappresentazione spazio-

¹¹Curioso il titolo scelto da Ghysels per questo articolo il quale è: *“Macroeconomics and reality of mixed frequency data”*. È chiaro il riferimento al celebratissimo articolo di Sims (1980) che ha dato il via alla letteratura sui processi VAR.

stato. Tale specificazione riduce sicuramente il carico computazionale del modello in quanto non richiede procedure di filtraggio. Inoltre, in linea con la letteratura MIDAS, permette l'utilizzo di polinomi ritardo parsimoniosi in grado di ridurre il numero di parametri da stimare.

Riassumendo, c'è consenso nella letteratura sul fatto che l'inserimento dei dati a diversa frequenza aggiunga informazioni rispetto alla soluzione tipica dell'aggregazione temporale. Non è chiaro però quali dei metodi studiati sia il "migliore". La sensazione è che la risposta sia: dipende da quali variabili si considerano e quale è il fine dell'analisi. Ad esempio, se l'obiettivo è la previsione di variabili macroeconomiche non c'è bisogno di specificazioni particolarmente parsimoniose. Il contrario accade invece se si coinvolgono serie finanziarie, dove il trade-off tra complessità del modello e sfruttamento di un set informativo esteso, è favorevole al primo.

Capitolo 3

Il modello econometrico

In questo capitolo si cercherà di stabilire, attraverso un'applicazione empirica, quali siano i vantaggi che si possono ottenere nell'analisi delle serie storiche campionate a diversa frequenza. Nel capitolo precedente abbiamo detto che la metodologia più utilizzata, quando si ha a che fare con dati campionati con frequenza diversa, è quella dell'aggregazione temporale. Questo approccio però, come sottolinea Marcellino (1999), può comportare la perdita di informazioni potenzialmente rilevanti. Per questo motivo, in letteratura si è cercato di implementare modelli econometrici che potessero incorporare più informazione possibile.

Tra gli altri Mariano e Murasawa (2003, 2010) e Ghysels (2016) hanno affrontato il problema dal punto di vista multivariato, andando a studiare i cosiddetti VAR a frequenza mista, sotto diversi aspetti. In particolare, il secondo articolo presenta un approccio che abbiamo definito "*observation driven*", in quanto è formulato esclusivamente su dati osservabili senza utilizzare procedure di filtraggio o shock latenti.

Questo lavoro si concentra proprio sul MIDAS-VAR proposto da Ghysels (2016); dunque l'obiettivo è quello di valutare quali siano i pro e i contro di questo approccio facendo un esercizio di stima e confrontandolo con l'approccio classico di aggregazione delle serie alla frequenza più bassa.

Per comprendere la portata del problema molti autori, tra cui Foroni e Marcellino (2016) e anche lo stesso Ghysels (2016), utilizzano la funzione di risposta d'impulso (Impulse Response Function, IRF). Marcellino (1999) ha annoverato quest'ultima tra le caratteristiche di un processo di serie storiche che sembrano essere influenzate dall'aggregazione temporale.

Per un'analisi preliminare consideriamo quindi di lavorare con due variabili disponibili alla frequenza mensile, ad esempio produzione industriale e inflazione¹. Il campione utilizzato per questa analisi parte dal primo mese del 1990 e arriva a Luglio 2018; si potrebbe stimare in un primo luogo un VAR(1) mensile e in seguito uno trimestrale, ottenuto in questo caso attraverso *l'average sampling*. Abbiamo effettuato questo breve esercizio per vedere, anche graficamente, quali possano essere le conseguenze dell'aggregazione temporale su uno dei più importanti strumenti a disposizione dell'econometrico: l'analisi strutturale e l'IRF.

Come mostrato nella Figura 3.1, le risposte sono abbastanza diverse. In particolare, l'entità della risposta istantanea sembra variare in maniera significativa. Dunque questo primo veloce esercizio, ci può dare un'idea di quanta informazione si può perdere o distorcere quando non si sfrutta in pieno il dataset a disposizione.

¹I dati utilizzati in questo esercizio sono stati scaricati dal database disponibile nel software econometrico open source gretl.

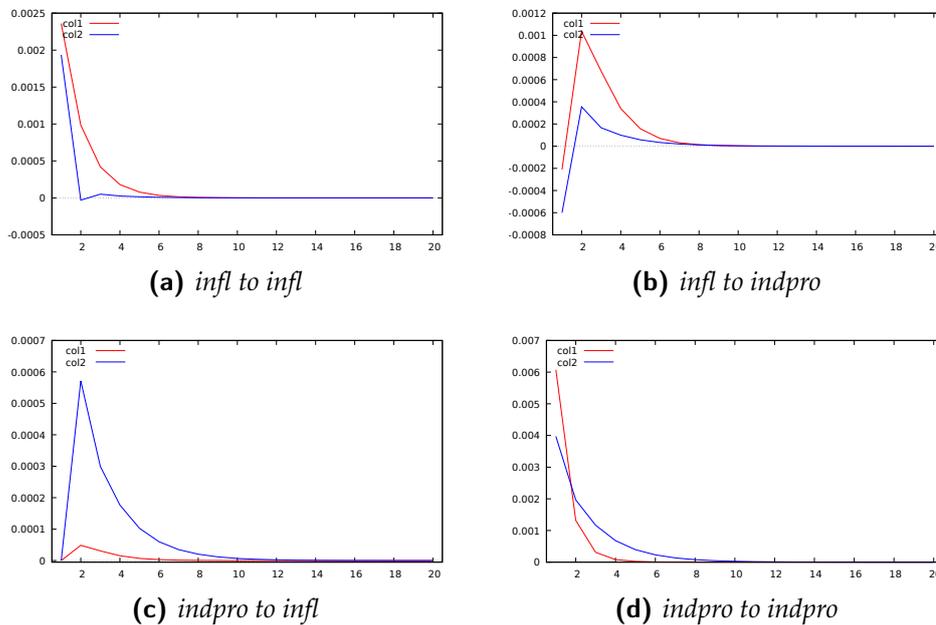


Figura 3.1: IRF mensile (linea rossa) e IRF trimestrale (linea blu)

Il proposito di questa breve introduzione è quello di provare a far comprendere il motivo dello sviluppo di modelli econometrici per dati campionati a diversa frequenza, ed inoltre strutturare la base di partenza del presente lavoro.

Il capitolo si sviluppa come segue. Nella prima parte effettueremo un'analisi su delle serie generate artificialmente. Successivamente ripeteremo lo studio generando le serie più volte all'interno di una simulazione Monte Carlo. Fatto ciò, si procederà con un'applicazione empirica su dati USA. In conclusione si presenterà un approccio alternativo alla gestione di dataset a frequenze miste.

Inoltre, occorre dire che lo sviluppo delle analisi appena presentate è svolto attraverso l'ausilio del pacchetto econometrico open source `gret1` (Cottrell e Lucchetti, 2019b). I codici elaborati sono riportati nell'Appendice

A.

3.1 La specificazione del modello

Il modello che si vuole andare ad analizzare empiricamente è dunque il cosiddetto MIDAS-VAR proposto da Ghysels (2016) (si veda par. 2.6.2). Nel suo articolo Ghysels effettua i seguenti esercizi. In primo luogo, dopo aver scelto un opportuno DGP, simula un processo VAR bivariato in presenza di una serie ad alta e una a bassa frequenza, con $m = 3$. Successivamente esegue un'applicazione empirica del MIDAS-VAR in cui le serie considerate sono: l'inflazione, il tasso di disoccupazione, la produzione industriale e il PIL. Tutte le serie sono a frequenza mensile ad eccezione del PIL che invece è campionato trimestralmente. In questo capitolo si procederà in modo analogo.

Il progetto di simulazione che si vuole implementare è strettamente correlato a quello adottato nella letteratura sugli MFVAR (si veda Ghysels *et al.* (2016)), in quanto si considerano due serie a bassa frequenza (trimestrali) ed una ad alta frequenza (mensile), in modo tale da aver il seguente DGP:

$$\begin{bmatrix} x_t^{(1)} \\ x_t^{(2)} \\ x_t^{(3)} \\ y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} x_{t-1}^{(1)} \\ x_{t-1}^{(2)} \\ x_{t-1}^{(3)} \\ y_{t-1} \\ z_{t-1} \end{bmatrix} + \varepsilon_t \quad (3.1.1)$$

3.1. LA SPECIFICAZIONE DEL MODELLO

Dove y_t e z_t sono le variabili trimestrali e $x_t^{(m)}$ è la variabile mensile con $m = 3$. Dunque da un VAR trivariato si passa ad un MIDAS-VAR a $K = 1 \cdot 3 + 2 = 5$ dimensioni, in quanto la serie mensile viene “scomposta” in 3 serie trimestrali. Il numero di osservazioni nel campione è fissato a $T = 100$ per la variabile a bassa frequenza e quindi a 300 per la variabile ad alta frequenza, in modo tale da avere un grandezza realistica del campione².

Operativamente si assume, in un primo momento, che il DGP è un processo MIDAS-VAR(1) con i seguenti coefficienti³:

$$A = \begin{bmatrix} 0 & 0.1 & 0.4 & 0 & 0 \\ 0.1 & -0.1 & 0.2 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 \\ 0 & -0.9 & 0.9 & 0.2 & 0 \\ 0 & 0 & 0 & 0.9 & 0.6 \end{bmatrix} \quad (3.1.2)$$

Tale selezione dei parametri è stata effettuata in modo tale che:

- x_t Granger-causa y_t , poiché la quarta riga della matrice A presenta parametri diversi da zero e di segno opposto per quanto riguarda il secondo e terzo mese di x_t ;
- y_t Granger-causa z_t , poiché l’elemento posto alla riga cinque e alla colonna quattro è diverso da zero;
- le serie siano stazionarie in covarianza (autovalori della matrice dei parametri tutti all’interno del cerchio unitario).

²Cento trimestri sono esattamente 25 anni.

³Per la scelta dei coefficienti si è preso spunto dal codice MatLab scritto da Kaiji Motegi, scaricabile da http://www2.kobe-u.ac.jp/~motegi/Matlab_Codes.html.

Così facendo, possiamo studiare quanto si guadagna in termini di previsione delle variabili a bassa frequenza, se si utilizzano anche regressori ad alta frequenza. Inoltre è possibile effettuare anche una previsione della variabile ad alta frequenza.

Il passo successivo da compiere è quello della stima del MIDAS-VAR(1) dalle serie simulate. Come abbiamo detto nel paragrafo 2.6.2 questa operazione non è complicata in quanto il VAR à la Ghysels (2016) gode di tutte le proprietà dei VAR tradizionali, quindi può essere stimato con l'OLS.

3.2 La stima

Si procede, quindi, alla stima del MIDAS-VAR(1) avendo nel dataset le serie simulate, attraverso il software econometrico gret1. L'output restituisce la stima di tante equazioni quante sono le dimensioni del MIDAS-VAR, ed è il seguente:

Tabella 3.1: MIDAS-VAR

MIDAS-VAR system, lag order 1
 OLS estimates, observations 2-100 ($T = 99$)

Equation 1: x1

| | Coefficient | Std. Error | t-ratio | p-value |
|--------------------|-------------|-------------------------|----------|---------|
| const | -0.113046 | 0.110760 | -1.021 | 0.3101 |
| x1 _{t-1} | 0.00794473 | 0.101666 | 0.07815 | 0.9379 |
| x2 _{t-1} | 0.0380542 | 0.105438 | 0.3609 | 0.7190 |
| x3 _{t-1} | 0.413363 | 0.114898 | 3.598 | 0.0005 |
| y _{t-1} | 0.0957280 | 0.0684013 | 1.400 | 0.1650 |
| z _{t-1} | 0.0723038 | 0.0507976 | 1.423 | 0.1580 |
| Mean dependent var | -0.092352 | S.D. dependent var | 1.145860 | |
| Sum squared resid | 108.0936 | S.E. of regression | 1.078099 | |
| R ² | 0.159939 | Adjusted R ² | 0.114775 | |
| F(5,93) | 3.541252 | P-value(F) | 0.005637 | |
| $\hat{\rho}$ | -0.043390 | Durbin-Watson | 2.036118 | |

F-tests of zero restrictions

| | | |
|----------------|-------------------------|----------|
| All lags of x1 | $F(1, 93) = 0.00610665$ | [0.9379] |
| All lags of x2 | $F(1, 93) = 0.13026$ | [0.7190] |
| All lags of x3 | $F(1, 93) = 12.9431$ | [0.0005] |
| All lags of y | $F(1, 93) = 1.95862$ | [0.1650] |
| All lags of z | $F(1, 93) = 2.02598$ | [0.1580] |

Equation 2: x2

| | Coefficient | Std. Error | t-ratio | p-value |
|--------------------|-------------|-------------------------|-----------|---------|
| const | 0.166600 | 0.102754 | 1.621 | 0.1083 |
| x1 _{t-1} | 0.0374250 | 0.0943177 | 0.3968 | 0.6924 |
| x2 _{t-1} | -0.0605918 | 0.0978166 | -0.6194 | 0.5371 |
| x3 _{t-1} | 0.0698288 | 0.106593 | 0.6551 | 0.5140 |
| y _{t-1} | 0.0213429 | 0.0634571 | 0.3363 | 0.7374 |
| z _{t-1} | -0.0232686 | 0.0471258 | -0.4938 | 0.6226 |
| Mean dependent var | 0.159729 | S.D. dependent var | 0.981831 | |
| Sum squared resid | 93.03170 | S.E. of regression | 1.000170 | |
| R ² | 0.015238 | Adjusted R ² | -0.037706 | |
| F(5, 93) | 0.287821 | P-value(F) | 0.918660 | |
| $\hat{\rho}$ | 0.059964 | Durbin-Watson | 1.852862 | |

F-tests of zero restrictions

| | | |
|----------------|-----------------------|----------|
| All lags of x1 | $F(1, 93) = 0.157448$ | [0.6924] |
| All lags of x2 | $F(1, 93) = 0.383709$ | [0.5371] |
| All lags of x3 | $F(1, 93) = 0.429153$ | [0.5140] |
| All lags of y | $F(1, 93) = 0.113122$ | [0.7374] |
| All lags of z | $F(1, 93) = 0.243794$ | [0.6226] |

Equation 3: x3

| | Coefficient | Std. Error | t-ratio | p-value |
|--------------------|-------------|-------------------------|-----------|---------|
| const | 0.0470860 | 0.101034 | 0.4660 | 0.6423 |
| x1 _{t-1} | -0.0140834 | 0.0927387 | -0.1519 | 0.8796 |
| x2 _{t-1} | 0.0664800 | 0.0961790 | 0.6912 | 0.4912 |
| x3 _{t-1} | 0.114265 | 0.104808 | 1.090 | 0.2784 |
| y _{t-1} | -0.0956175 | 0.0623947 | -1.532 | 0.1288 |
| z _{t-1} | 0.0133634 | 0.0463368 | 0.2884 | 0.7737 |
| Mean dependent var | 0.078786 | S.D. dependent var | 0.978750 | |
| Sum squared resid | 89.94280 | S.E. of regression | 0.983426 | |
| R ² | 0.041932 | Adjusted R ² | -0.009577 | |
| F(5, 93) | 0.814069 | P-value(F) | 0.542610 | |
| $\hat{\rho}$ | 0.021113 | Durbin-Watson | 1.944321 | |

F-tests of zero restrictions

| | | |
|----------------|------------------------|----------|
| All lags of x1 | $F(1, 93) = 0.0230617$ | [0.8796] |
| All lags of x2 | $F(1, 93) = 0.477773$ | [0.4912] |
| All lags of x3 | $F(1, 93) = 1.18859$ | [0.2784] |
| All lags of y | $F(1, 93) = 2.34844$ | [0.1288] |
| All lags of z | $F(1, 93) = 0.0831731$ | [0.7737] |

CAPITOLO 3. IL MODELLO ECONOMETRICO

Equation 4: y

| | Coefficient | Std. Error | t-ratio | p-value |
|--------------------|-------------|--------------------|----------|---------|
| const | -0.0689383 | 0.109726 | -0.6283 | 0.5314 |
| $x1_{t-1}$ | -0.00191465 | 0.100717 | -0.01901 | 0.9849 |
| $x2_{t-1}$ | -0.848161 | 0.104454 | -8.120 | 0.0000 |
| $x3_{t-1}$ | 0.723719 | 0.113825 | 6.358 | 0.0000 |
| y_{t-1} | 0.252890 | 0.0677628 | 3.732 | 0.0003 |
| z_{t-1} | 0.0202307 | 0.0503234 | 0.4020 | 0.6886 |
| Mean dependent var | -0.155997 | S.D. dependent var | 1.605853 | |
| Sum squared resid | 106.0849 | S.E. of regression | 1.068035 | |
| R^2 | 0.580226 | Adjusted R^2 | 0.557657 | |
| $F(5,93)$ | 25.70952 | P-value(F) | 3.31e-16 | |
| $\hat{\rho}$ | 0.005811 | Durbin-Watson | 1.971801 | |

F-tests of zero restrictions

| | | |
|----------------|-------------------------|----------|
| All lags of x1 | $F(1,93) = 0.000361385$ | [0.9849] |
| All lags of x2 | $F(1,93) = 65.934$ | [0.0000] |
| All lags of x3 | $F(1,93) = 40.4261$ | [0.0000] |
| All lags of y | $F(1,93) = 13.9277$ | [0.0003] |
| All lags of z | $F(1,93) = 0.161615$ | [0.6886] |

Equation 5: z

| | Coefficient | Std. Error | t-ratio | p-value |
|--------------------|-------------|--------------------|----------|---------|
| const | 0.0616234 | 0.106680 | 0.5776 | 0.5649 |
| $x1_{t-1}$ | -0.0228262 | 0.0979211 | -0.2331 | 0.8162 |
| $x2_{t-1}$ | 0.0693008 | 0.101554 | 0.6824 | 0.4967 |
| $x3_{t-1}$ | 0.0238425 | 0.110665 | 0.2154 | 0.8299 |
| y_{t-1} | 0.853377 | 0.0658814 | 12.95 | 0.0000 |
| z_{t-1} | 0.592134 | 0.0489262 | 12.10 | 0.0000 |
| Mean dependent var | -0.087500 | S.D. dependent var | 2.227670 | |
| Sum squared resid | 100.2760 | S.E. of regression | 1.038381 | |
| R^2 | 0.793809 | Adjusted R^2 | 0.782724 | |
| $F(5,93)$ | 71.60778 | P-value(F) | 2.30e-30 | |
| $\hat{\rho}$ | 0.017692 | Durbin-Watson | 1.954342 | |

F-tests of zero restrictions

| | | |
|----------------|-----------------------|----------|
| All lags of x1 | $F(1,93) = 0.0543394$ | [0.8162] |
| All lags of x2 | $F(1,93) = 0.465678$ | [0.4967] |
| All lags of x3 | $F(1,93) = 0.0464176$ | [0.8299] |
| All lags of y | $F(1,93) = 167.786$ | [0.0000] |
| All lags of z | $F(1,93) = 146.473$ | [0.0000] |

Dalla prima equazione, cioè quella relativa a $x_t^{(1)}$, si evince che l'unico coefficiente significativo è quello relativo al ritardo del terzo mese del

trimestre. Inoltre osservando anche la seconda e la terza equazione, si nota come non siano presenti parametri significativi. Questo è in linea con quanto previsto in base al DGP scelto. In tale situazione, la previsione ad alta frequenza è poco robusta.

Più interessanti sono le ultime due equazioni del MIDAS-VAR, le quali riguardano rispettivamente y_t e z_t . Per quanto riguarda la prima, dall'output di stima si deduce che il primo e il secondo mese del trimestre della variabile x_t sono significativi quindi, come ci si aspettava, x_t Granger-causa y_t , si veda Ghysels, Hill e Motegi (2016). Anche il ritardo della variabile dipendente risulta statisticamente diverso da zero. L' R^2 di questa equazione si attesta su valori intorno al 58%, mentre l' R^2 aggiustato è leggermente più basso (55%). Non si segnalano problemi di altro genere.

In riferimento all'equazione per z_t , è interessante evidenziare come i parametri relativi ad $x_t^{(m)}$ non siano assolutamente significativi, mentre è statisticamente diverso da zero quello relativo al primo *lag* alla variabile dipendente e quello relativo alla variabile y_{t-1} . Pertanto, per quanto concerne questa analisi, si può dire che x_t non Granger-causa z_t . Successivamente si procede con la stima del VAR(1) tradizionale, ottenuto aggregando le serie ad alta frequenza.

La tecnica di aggregazione temporale varia a seconda del tipo di variabili che si sta studiando; siano esse stock o flussi. Supponiamo in questo esercizio di avere a che fare con variabili stock.

I metodi di aggregazione più utilizzati in letteratura, per le variabili cosiddette stock, sono l'*average sampling* e il *point-in-time sampling*. Nel primo metodo si effettua una media dei mesi all'interno del trimestre,

mentre nel secondo si sceglie come valore rappresentativo del trimestre l'ultimo disponibile (per dettagli si veda par 2.4).

In questo esempio useremo come metodo di aggregazione temporale l'average sampling o campionamento medio. I motivi di questa scelta sono dovuti al fatto che il campionamento medio si presta in modo migliore all'obiettivo di far capire quali possano essere le conseguenze del non considerare l'approccio a frequenze miste, ed inoltre gli analisti tendono ad utilizzare in maniera più frequente questo metodo.

La stima del VAR(1) tradizionale, trimestrale e trivariato restituisce il seguente output:

Tabella 3.2: VAR tradizionale

Traditional VAR system (average sampling), lag order 1
 OLS estimates, observations 2-100 ($T = 99$)

Equation 1: x

| | Coefficient | Std. Error | t-ratio | p-value |
|--------------------|-------------|--------------------|----------|---------|
| const | 0.0412120 | 0.0631772 | 0.6523 | 0.5158 |
| x_{t-1} | 0.207197 | 0.0997656 | 2.077 | 0.0405 |
| y_{t-1} | -0.00327611 | 0.0393123 | -0.08334 | 0.9338 |
| z_{t-1} | 0.0130250 | 0.0287390 | 0.4532 | 0.6514 |
| Mean dependent var | 0.048721 | S.D. dependent var | 0.630388 | |
| Sum squared resid | 37.05228 | S.E. of regression | 0.624519 | |
| R^2 | 0.048578 | Adjusted R^2 | 0.018533 | |
| $F(3,95)$ | 1.616830 | P-value(F) | 0.190609 | |
| $\hat{\rho}$ | 0.046482 | Durbin-Watson | 1.842229 | |

F-tests of zero restrictions

| | | |
|---------------|------------------------|----------|
| All lags of x | $F(1,95) = 4.31326$ | [0.0405] |
| All lags of y | $F(1,95) = 0.00694483$ | [0.9338] |
| All lags of z | $F(1,95) = 0.205405$ | [0.6514] |

Equation 2: y

| | Coefficient | Std. Error | t-ratio | p-value |
|--------------------|-------------|--------------------|----------|---------|
| const | -0.124470 | 0.161376 | -0.7713 | 0.4424 |
| x_{t-1} | -0.0660444 | 0.254835 | -0.2592 | 0.7961 |
| y_{t-1} | 0.203272 | 0.100417 | 2.024 | 0.0457 |
| z_{t-1} | 0.0214399 | 0.0734091 | 0.2921 | 0.7709 |
| Mean dependent var | -0.155997 | S.D. dependent var | 1.605853 | |
| Sum squared resid | 241.7534 | S.E. of regression | 1.595234 | |
| R^2 | 0.043390 | Adjusted R^2 | 0.013181 | |
| $F(3, 95)$ | 1.436329 | P-value(F) | 0.237070 | |
| $\hat{\rho}$ | -0.008949 | Durbin-Watson | 1.943719 | |

F-tests of zero restrictions

| | | |
|---------------|------------------------|----------|
| All lags of x | $F(1, 95) = 0.0671665$ | [0.7961] |
| All lags of y | $F(1, 95) = 4.0977$ | [0.0457] |
| All lags of z | $F(1, 95) = 0.085299$ | [0.7709] |

Equation 3: z

| | Coefficient | Std. Error | t-ratio | p-value |
|--------------------|-------------|--------------------|----------|---------|
| const | 0.0724961 | 0.104164 | 0.6960 | 0.4881 |
| x_{t-1} | 0.0528203 | 0.164489 | 0.3211 | 0.7488 |
| y_{t-1} | 0.851805 | 0.0648164 | 13.14 | 0.0000 |
| z_{t-1} | 0.586927 | 0.0473836 | 12.39 | 0.0000 |
| Mean dependent var | -0.087500 | S.D. dependent var | 2.227670 | |
| Sum squared resid | 100.7231 | S.E. of regression | 1.029681 | |
| R^2 | 0.792890 | Adjusted R^2 | 0.786350 | |
| $F(3, 95)$ | 121.2311 | P-value(F) | 2.31e-32 | |
| $\hat{\rho}$ | 0.013081 | Durbin-Watson | 1.959659 | |

F-tests of zero restrictions

| | | |
|---------------|-----------------------|----------|
| All lags of x | $F(1, 95) = 0.103116$ | [0.7488] |
| All lags of y | $F(1, 95) = 172.707$ | [0.0000] |
| All lags of z | $F(1, 95) = 153.43$ | [0.0000] |

Dai risultati delle stime del modello, si osserva come nella prima equazione relativa alla variabile x_t trimestrale non vi sia particolare differenza rispetto alla controparte a frequenza mista. Infatti il p-value del coefficiente relativo a x_{t-1} è di poco al di sotto del livello di confidenza convenzionale del 5%, mentre gli altri coefficienti presentano un p-value ampiamente superiori a tale livello. In aggiunta l' R^2 è molto simile alle tre equazioni riguardanti $x_t^{(m)}$ presenti nel MIDAS-VAR.

Più interessanti, per quanto riguarda l'analisi qualitativa, sono le equazioni relative alle due variabili disponibili, già in origine, alla frequenza più bassa.

Nella seconda equazione, infatti, si osserva che x_t non Granger-causa y_t . Tale risultato va in contrasto con quanto risulta dall'analisi precedente. Questo perché l'impatto positivo di $x_t^{(3)}$ e l'impatto negativo di $x_t^{(2)}$ su y_t , risultanti dalla quarta equazione nella Tabella 3.1, si compensano a vicenda dopo l'aggregazione. Ciò evidenzia un vantaggio, in questo caso, dell'approccio a frequenze miste. Infine, analizzando l'ultima equazione relativa a z_t si evince che non c'è alcuna differenza rilevante con l'approccio a frequenza mista. Le variabili statisticamente significative sono quelle relative al primo ritardo di y_t e z_t .

Dopo questa prima analisi qualitativa, si può dire che per quanto riguarda il set informativo simulato attraverso un apposito DGP, si hanno dei vantaggi nell'utilizzo di modelli econometrici a frequenza mista poiché diminuisce il rischio di distorcere l'informazione contenuta nel dataset completo.

A questo punto, non resta che valutare la bontà relativa dei due modelli VAR e MIDAS-VAR. Per fare ciò in letteratura esistono vari indicatori di bontà di previsione e test utilizzati nel contesto delle serie storiche.

3.2.1 Misure di bontà previsiva

In questa sede faremo riferimento ad un indicatore, il **Root Mean Square Error** (RMSE), e a due test cioè quelli di Diebold e Mariano (1995) e di

Giacomini e White (2006).

Per quanto riguarda l'errore quadratico medio, il confronto è effettuato sulle previsioni (*in sample*) relative alle variabili y_t e z_t , ed i risultati sono i seguenti:

| | |
|---------------------------------|--------|
| Forecast (in sample) RMSEs: | |
| MIDAS-VAR, forecasting y: | 1.0352 |
| MIDAS-VAR, forecasting z: | 1.0064 |
| Traditional VAR, forecasting y: | 1.5627 |
| Traditional VAR, forecasting z: | 1.0087 |

Si evidenzia un abbassamento significativo dell'errore quadratico medio soltanto per la previsione di y_t . È quello che ci si aspettava, infatti il modello a frequenze miste aggiunge informazione soprattutto nella previsione di y_t in quanto si nota una significatività dei parametri relativi alle serie ad alta frequenza $x_t^{(m)}$, nell'equazione di y_t .

I test di Diebold-Mariano (DM) e di Giacomini-White (GW) confermano le considerazioni appena fatte. Questi test cercano di stabilire quale tra i due modelli alternativi abbia la migliore capacità previsiva circa l'andamento nel tempo di una serie storica. In entrambi i test, se si accetta l'ipotesi nulla, i modelli posti a confronto hanno la stessa accuratezza previsiva. Al contrario se l'ipotesi nulla non viene accettata, allora uno dei due modelli è migliore per quanto riguarda la previsione. Nel test DM se la statistica test è negativa vuol dire che la previsione da preferire è quella ricavata con il primo modello stimato (in questo caso il MIDAS-VAR) mentre il contrario se la statistica test è positiva. Per quanto riguarda il test GW invece, l'informazione su quale modello sia il "migliore", ce la dà il segno

| | Diebold-Mariano | | Giacomini-White | |
|-------|-----------------|-----------------|-----------------|-----------------|
| | U-shaped | V-shaped | U-shaped | V-shaped |
| y_t | -4.2616 (0.000) | -4.5731 (0.000) | 15.4781 (0.000) | 17.4109 (0.000) |
| z_t | -0.3489 (0.727) | -1.040 (0.298) | 0.1228 (0.726) | 1.0816 (0.298) |

Tabella 3.3: Statistica test e (p-value) dei test di Diebold e Mariano (1995) e Giacomini e White (2006).

della *loss function*. Se quest'ultima è positiva è il secondo modello stimato quello con più accuratezza previsiva, il primo se è negativa.

Dalla tabella 3.3 si può vedere come, in relazione alla previsione di y_t , il modello migliore è rappresentato dal MIDAS-VAR. Questo si deduce dal fatto che l'ipotesi nulla di medesima accuratezza previsiva deve essere rifiutata per entrambi i test. In aggiunta la statistica test del test DM è negativa quindi il modello con la migliore capacità previsiva è, in questo caso, il MIDAS-VAR.

Il test GW anch'esso rifiuta l'ipotesi nulla e, poichè il software gret1 restituisce in questo caso una funzione di perdita negativa, si può dire che i due test concordano su quale sia il modello con più accuratezza previsiva. Infine si segnala che tutti e due i test concordano sia se si definisce una funzione di perdita *U-shaped* sia *V-shaped*.

In merito alla previsione per z_t , invece, i test risultano contrari rispetto alle considerazioni appena fatte per y_t , ma l'interpretazione è in linea con quella dedotta dall'analisi del RMSE. Il test DM, accetta l'ipotesi nulla di stessa accuratezza previsiva, in entrambe le versioni. Questo significa che, i due modelli posti a confronto, sono statisticamente equivalenti per quanto riguarda la previsione di z_t .

Analogo è l'esito del test GW, che accetta l'ipotesi nulla in entrambe le varianti utilizzate. Quindi, secondo i test, non c'è un vantaggio significativo nell'utilizzo dell'approccio a frequenze miste per la previsione di z_t . In altre parole, per quanto riguarda la previsione di z_t , i due modelli hanno pressoché la stessa accuratezza nelle previsioni.

In conclusione, con la specificazione dei parametri alla base del DGP, ci si aspettava una migliore previsione per y_t del MIDAS-VAR rispetto al VAR tradizionale. L'indicatore di bontà previsiva e i test hanno confermato le analisi preliminari.

3.2.2 Impulse Response Function

Un altro importante strumento a disposizione per valutare l'affidabilità e la bontà del modello, è rappresentato dall'analisi strutturale e dalla cosiddetta funzione di risposta d'impulso (IRF, dall'inglese Impulse Response Function). In econometria, le funzioni di risposta d'impulso vengono utilizzate per descrivere come l'economia reagisce nel tempo agli impulsi esogeni, che gli economisti chiamano solitamente shock. Le funzioni di risposta d'impulso descrivono la reazione di variabili macroeconomiche endogene come produzione, consumo, investimento e occupazione al momento dello shock e in momenti successivi. Quindi confrontare le funzioni di risposta d'impulso ottenute, prima con l'approccio a frequenza mista poi con l'aggregazione temporale, può essere utile per valutare quale dei due modelli sia più appropriato per l'analisi che si vuole effettuare.

In ognuno dei due modelli stimati, applichiamo la scomposizione di

Cholesky per rendere gli shock ortogonali. Calcoliamo le funzioni di risposta all'impulso per scoprire cosa succede al sistema fino a dieci trimestri avanti, dopo uno shock esogeno di x_t .

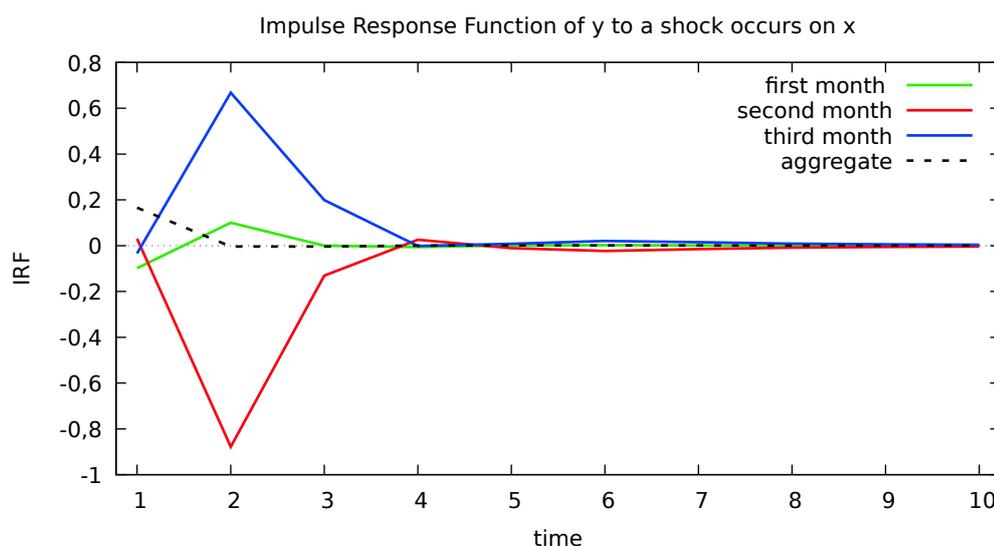


Figura 3.2: IRF di y_t rispetto ad uno shock del primo (linea verde), secondo (linea rossa) e terzo mese (linea blu) di x_t , in confronto con la risposta di y_t ad uno shock della variabile aggregata (linea tratteggiata).

Nella figura 3.2 si mettono a confronto le risposte di impulso di y_t ad uno shock esogeno del primo, secondo e terzo mese di $x_t^{(m)}$, con le risposte di impulso di y_t ad uno shock della variabile x_t ottenuta aggregando le osservazioni ad alta frequenza.

Si osserva come la risposta d'impulso negativa di y_t rispetto ad uno shock al secondo mese del trimestre di x_t viene compensata da quella positiva rispetto ad uno al terzo mese, di conseguenza la risposta d'impulso aggregata non riesce a cogliere tali dinamiche.

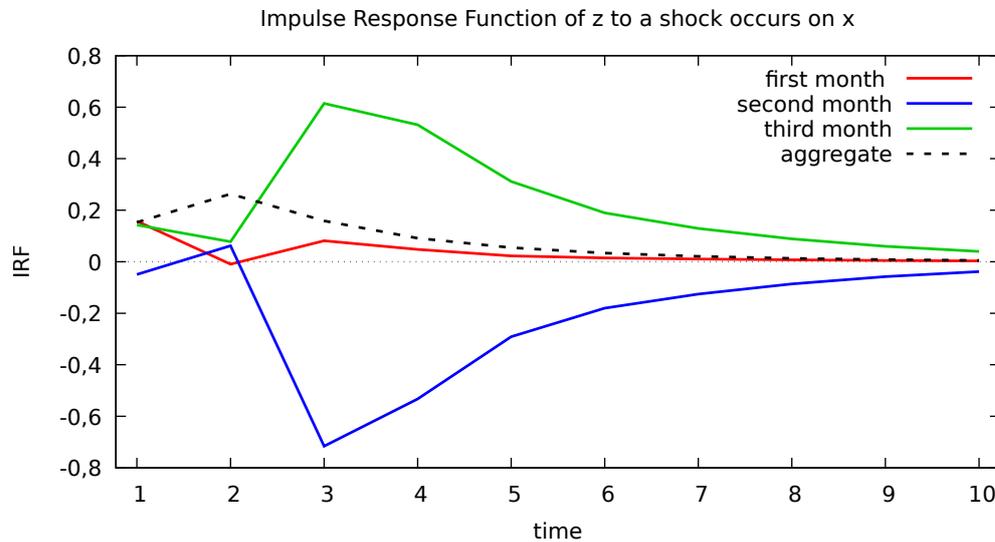


Figura 3.3: IRF di z_t rispetto ad uno shock del primo (linea rossa), secondo (linea blu) e terzo mese (linea verde) di x_t , in confronto con la risposta di z_t ad uno shock della variabile aggregata (linea tratteggiata).

Altri spunti di riflessione li offre la figura 3.3, che mette a confronto le risposte d'impulso di z_t .

Nel capitolo precedente si è detto che la variabile x_t non granger-causa la variabile z_t , e ciò sembra verificato dalla risposta d'impulso aggregata. Al contrario, invece, il primo e il secondo mese del trimestre della variabile $x_t^{(m)}$ sembrano avere un un impatto significativo su z_t . Questo è un tipico esempio di "catena causale" o "causal chain", ovvero $x_t^{(3)}$ causa z_t tramite y_t . Per comprendere questo passaggio basta osservare la matrice:

$$A^2 = \begin{bmatrix} 0.01 & -0.01 & 0.06 & 0 & 0 \\ -0.01 & 0.02 & 0.04 & 0 & 0 \\ 0 & 0 & 0.01 & 0 & 0 \\ -0.09 & -0.09 & 0.09 & 0.04 & 0 \\ 0 & -0.81 & 0.81 & 0.72 & 0.36 \end{bmatrix} \quad (3.2.1)$$

In cui la parte triangolare inferiore non è più zero. In base al modello a

bassa frequenza, non è possibile osservare l'effetto di x_t né su y_t né su z_t . Si rileva, pertanto, un vantaggio nell'approccio *mixed frequency*.

Tutte le considerazioni appena fatte però hanno un limite: si riferiscono al caso specifico considerato, cioè al processo di simulazione stabilito e al metodo di aggregazione utilizzato.

Attraverso l'implementazione di simulazioni Monte Carlo si possono generare più variabili considerando diverse combinazioni di parametri, tenendo conto anche della componente casuale nella generazione delle serie stesse.

L'obiettivo delle varie combinazioni di parametri è quello di coprire una vasta gamma di DGP con diversi gradi di persistenza e correlazione tra le variabili ad alta frequenza e quelle a bassa frequenza.

3.3 Simulazioni Monte Carlo

Per implementare la simulazione abbiamo selezionato altre due configurazioni di coefficienti alla base del DGP, oltre a quella relativa all'equazione (3.1.2). Le matrici sono le seguenti:

$$A1 = \begin{bmatrix} 0.3 & 0.1 & 0.2 & 0 & 0 \\ 0.1 & -0.4 & -0.15 & 0 & 0 \\ 0 & 0.1 & -0.3 & 0 & 0 \\ 0.2 & -0.5 & 0.68 & 0.3 & 0 \\ 0 & 0 & 0 & 0.75 & 0.5 \end{bmatrix} \quad A2 = \begin{bmatrix} 0.3 & 0.1 & 0.2 & 0 & 0 \\ 0.1 & -0.4 & -0.15 & 0 & 0 \\ 0 & 0.1 & -0.3 & 0 & 0 \\ 0.2 & 0.5 & 0.68 & 0.3 & 0 \\ 0.5 & -0.35 & -0.2 & 0.7 & 0.4 \end{bmatrix}$$

In questo modo si hanno più combinazioni di correlazione tra le variabili ad alta frequenza (mensili) e quella a bassa frequenza (trimestrali).

Per valutare gli effetti dell'aggregazione e i benefici derivanti dall'uso di dati a frequenza mista, generiamo i dati R volte, per ogni combinazione di parametri, fissando $R = 1000$. Successivamente per ogni ripetizione si stimano i due modelli *mixed* e *low frequency*⁴ e contestualmente si calcolano le previsioni nel campione. Per ogni ripetizione si determina infine l'errore quadratico medio (RMSE) della previsione.

Si analizza quindi se la media sulle ripetizioni dell'errore di previsione del modello a frequenza mista è minore rispetto all'errore commesso utilizzando il modello a bassa frequenza. Inoltre si effettuano i test di Diebold-Mariano e Giacomini-White⁵ registrando quante volte, all'interno della simulazione, quest'ultimi stabiliscono che la previsione più accurata è quella effettuata attraverso il modello VAR a frequenze miste.

I risultati della simulazione Monte Carlo sono riassunti nelle tabelle che seguono:

| RMSE | A | | A1 | | A2 | |
|-----------|--------|--------|--------|--------|--------|--------|
| | yt | zt | yt | zt | yt | zt |
| MIDAS-VAR | 0.9841 | 0.9763 | 0.9740 | 0.9739 | 0.9769 | 0.9775 |
| VAR | 1.3243 | 0.9865 | 1.1433 | 0.9844 | 1.1322 | 1.1709 |

Tabella 3.4: Media del *forecast RMSE* sulle $R = 1000$ simulazioni dei due modelli *mixed* e *low frequency*, per ognuna delle combinazioni dei parametri. Il VAR tradizionale a bassa frequenza è ottenuto attraverso il metodo di aggregazione point-in-time.

Dalla tabella 3.4 si deduce che in media l'RMSE della previsione nel campione, è più basso per il MIDAS-VAR rispetto al VAR tradizionale (point-in-time sampling) in ognuna delle impostazioni dei parametri. Per

⁴L'analisi è effettuata considerando due metodi di aggregazione: point in time e average sampling.

⁵Entrambi i test sono svolti attraverso la specificazione di una funzione di perdita U-shaped.

le prime due combinazioni di parametri si ha un risultato analogo a quello ottenuto nell'analisi illustrata nel capitolo precedente, cioè si evidenzia un significativo abbassamento dell'errore di previsione per la variabile y_t , mentre per z_t si registra una differenza minima. Nell'ultima specificazione (A2) invece si ha un vantaggio più rilevante per la previsione di entrambe le variabili.

| | A | | A1 | | A2 | |
|----|-------|-------|-------|-------|-------|-------|
| | y_t | z_t | y_t | z_t | y_t | z_t |
| DM | 1000 | 2 | 939 | 2 | 904 | 979 |
| GW | 1000 | 2 | 933 | 2 | 901 | 979 |

Tabella 3.5: Numero di volte, all'interno della simulazione Monte Carlo, che i test di Diebold-Mariano e Giacomini-White rifiutano l'ipotesi nulla di stessa capacità previsiva a favore del MIDAS-VAR. Il confronto è tra quest'ultimo e un VAR tradizionale stimato aggregando le serie ad alta frequenza tramite il point-in-time sampling.

I risultati dei test riportati nella tabella 3.5 confermano che, per la previsione di y_t , è statisticamente migliore il MIDAS-VAR per la quasi totalità dei casi. Mentre i due modelli sono equivalenti per la previsione di z_t , ad eccezione dell'ultima specificazione dei parametri alla base del DGP.

| RMSE | A | | A1 | | A2 | |
|-----------|--------|--------|--------|--------|--------|--------|
| | y_t | z_t | y_t | z_t | y_t | z_t |
| MIDAS-VAR | 0.9841 | 0.9763 | 0.9740 | 0.9739 | 0.9769 | 0.9775 |
| VAR | 1.5842 | 0.9869 | 1.3095 | 0.9847 | 1.0517 | 1.1879 |

Tabella 3.6: Media del *forecast RMSE* sulle $R = 1000$ simulazioni, dei due modelli mixed e low frequency, per ognuna delle combinazioni dei parametri. Il VAR tradizionale a bassa frequenza è ottenuto attraverso il metodo di aggregazione medio (average sampling).

La scelta del metodo di aggregazione non cambia di molto i risultati dell'analisi, come si evince dalle tabelle 3.6 e 3.7. L'unica differenza che balza all'occhio è quella relativa alle serie generate dalla combinazione di

parametri A2, dove per la previsione di y_t si abbassa la percentuale di casi in cui è migliore il modello MIDAS-VAR (poco sopra al 40%). Questo può essere dovuto al fatto che effettuando una media del trimestre si utilizza comunque più informazione rispetto all'utilizzo di un solo valore, quindi il VAR tradizionale "recupera" capacità previsiva.

| | A | | A1 | | A2 | |
|----|-------|-------|-------|-------|-------|-------|
| | y_t | z_t | y_t | z_t | y_t | z_t |
| DM | 1000 | 2 | 1000 | 5 | 441 | 987 |
| GW | 1000 | 2 | 1000 | 4 | 429 | 985 |

Tabella 3.7: Numero di volte, all'interno della simulazione Monte Carlo, che i test di Diebold-Mariano e Giacomini-White rifiutano l'ipotesi nulla di stessa capacità previsiva a favore del MIDAS-VAR. Il confronto è tra quest'ultimo e un VAR tradizionale stimato aggregando le serie ad alta frequenza tramite l'average sampling.

Riassumendo i risultati basati sulla simulazione, si può dire che l'utilizzo dei processi VAR a frequenza mista offre una soluzione più efficiente per la previsione in determinate circostanze. Ad esempio, quando da un'analisi preliminare risulta che le serie ad alta frequenza sono statisticamente rilevanti per spiegare le altre variabili presenti nel sistema, allora il modello più efficiente per quanto riguarda la previsione è quello che utilizza più informazione possibile. Infine occorre dire che il metodo di aggregazione utilizzato non inficia considerevolmente la natura dei risultati.

3.4 Un'applicazione empirica

A questo punto, forniamo un'applicazione empirica per illustrare ulteriormente l'analisi teorica condotta finora e, nel contempo, analizzare l'importanza delle questioni di aggregazione temporale nella pratica. Ana-

logamente a Ghysels (2016), condurremo il classico esempio della curva di Phillips. Nello specifico, consideriamo le seguenti due serie ad alta frequenza (mensili):

- Tasso di inflazione (serie `infl` o anche π_t , figura 3.4), individuato nel tasso di variazione dell'indice dei prezzi al consumo, cioè della serie "CPI, Consumer Price Index for All Urban Consumers: All Items" (serie `cpiaucs1`). La variazione è calcolata applicando la differenza logaritmica ad alta frequenza⁶;
- Tasso di disoccupazione, individuato dalla serie "Civilian Unemployment Rate" (serie `u2rate` oppure u_t , figura 3.5).

E la seguente serie a bassa frequenza (trimestrale):

- PIL reale, definito dalla serie "Real Gross Domestic Product, Bil. of Chn. 2012 \$, SAAR" (serie `gdpc1`).

A quest'ultima applichiamo il filtro di Hodrick-Prescott che è un metodo di smoothing delle serie storiche il cui obiettivo è rimuovere la componente ciclica dal dato grezzo. Si ottiene così una rappresentazione della serie, che chiameremo y_t o "output gap", nella quale sono evidenziate solamente le fluttuazioni di breve periodo (figura 3.6).

Il campione che consideriamo per questa analisi, va dal primo trimestre del 1970 al primo trimestre del 2016, per un totale di 185 trimestri. I 12 trimestri che vanno dal secondo trimestre del 2016 al primo del 2019 saranno utilizzati per effettuare una previsione *out of sample*.

⁶Si veda il comando `hldiff` in Cottrell e Lucchetti (2019a)

3.4. UN'APPLICAZIONE EMPIRICA

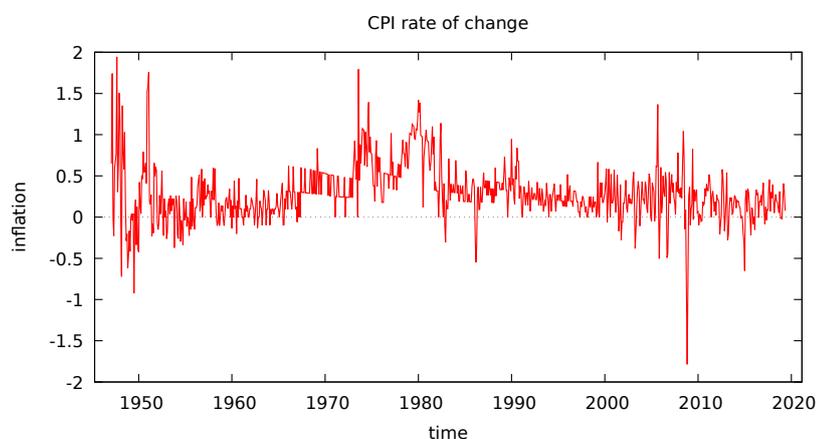


Figura 3.4: Inflazione mensile, negli Stati Uniti, calcolata come tasso di variazione dell'indice dei prezzi al consumo.

Le serie descritte sono ricavate dal database della Federal Reserve Bank of St.Louis, disponibile direttamente all'interno del pacchetto econometrico gretl.

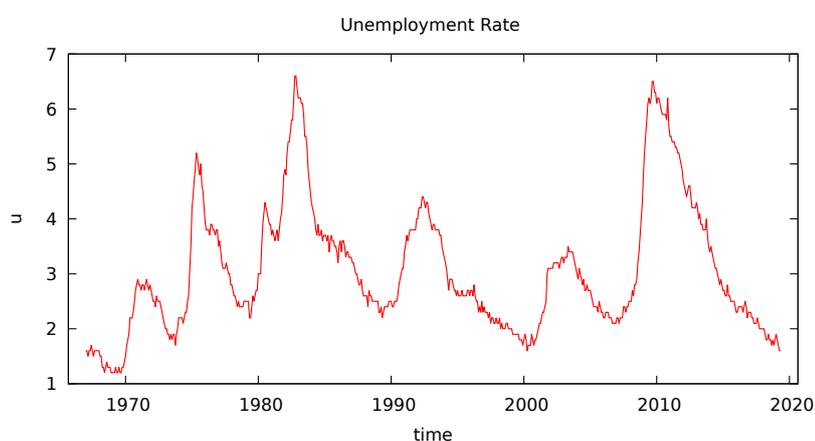


Figura 3.5: Tasso di disoccupazione negli Stati Uniti, serie mensile.

Dalle figure 3.4 e 3.6 si deduce che le due serie sono, con molta probabilità, stazionarie in covarianza. Lo stesso non si può dire per la serie del tasso di disoccupazione, rappresentata nella figura 3.5, per la quale c'è

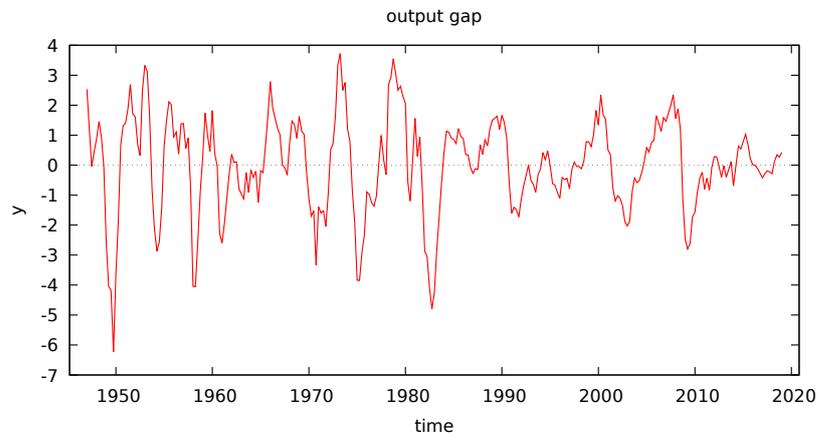


Figura 3.6: Componente ciclica del PIL reale negli Stati Uniti, serie trimestrale.

bisogno di un test di radice unitaria per stabilire se essa è integrata oppure no. Nella tabella 3.8 sono riportati i test augmented Dickey–Fuller (ADF), per le variabili precedentemente descritte.

| serie | | const | const + t | const + $t + t^2$ |
|---------|-----------------|-------------|-------------|-------------------|
| y_t | statistica test | -7.12844 | -7.12359 | -7.13324 |
| | (p-value) | (1.482e-10) | (1.083e-09) | (1.433e-10) |
| π_t | statistica test | -5.46404 | -5.44537 | -6.38362 |
| | (p-value) | (2.086e-06) | (2.197e-05) | (4.768e-07) |
| u_t | statistica test | -3.63037 | -3.56979 | -3.68698 |
| | (p-value) | (0.005234) | (0.0324) | (0.07254) |

Tabella 3.8: Test ADF con costante, trend e trend quadratico.

Tutti i p-value dei test sono inferiori al 5%, ad eccezione dell'ultimo test per la serie u_t . Possiamo affermare che le serie sono stazionarie.

A questo punto si procede con la stima del modello MIDAS-VAR à la Ghysels (2016) di $K = 2 \times 3 + 1 = 7$ dimensioni per x_t , dove:

$$x_t = \begin{bmatrix} \pi_t^{(1)} \\ \pi_t^{(2)} \\ \pi_t^{(3)} \\ u_t^{(1)} \\ u_t^{(2)} \\ u_t^{(3)} \\ y_t \end{bmatrix} \quad (3.4.1)$$

La scelta dell'ordine di ritardo è effettuata tramite i criteri informativi. Dunque decidiamo di scegliere il modello con il criterio di Akaike minore e fissiamo $p = 2$.

Il passo successivo è quello di calcolare le previsioni fuori dal campione per y_t , e l'errore quadratico medio (RMSE) a loro associato, attraverso il MIDAS-VAR(2) appena stimato.

La medesima procedura viene applicata al VAR(2) tradizionale ottenuto aggregando le serie ad alta frequenza (π_t e u_t) attraverso il campionamento medio (average sampling).

L'output ottenuto da gret1 è il seguente:

```
Forecast RMSEs (out of sample):
MIDAS-VAR(2), forecasting y:      0.3146
Traditional VAR(2), forecasting y: 0.5259
```

Si ottiene, quindi, una diminuzione dell'errore di previsione per y_t (output gap), attraverso l'utilizzo del processo VAR a frequenza mista. Inoltre, i test di Diebold e Mariano (1995) e Giacomini e White (2006), avvalorano la tesi introdotta dall'analisi degli RMSE.

| | Diebold-Mariano | | Giacomini-White | |
|---------|-----------------|------------------|-----------------|-----------------|
| | U-shaped | V-shaped | U-shaped | V-shaped |
| y_t | -2.8479 (0.004) | -3.8533 (0.0001) | 6.4427 (0.011) | 8.2237 (0.0041) |
| LF sign | / | / | -0.178 | -0.204 |

Tabella 3.9: Statistica test e (p-value) dei test di Diebold e Mariano (1995) e Giacomini e White (2006).

Infatti la tabella 3.9 mostra come l'ipotesi nulla di equivalente precisione nella previsione, è rifiutata. I p-value dei test sono tutti al di sotto del 5%, la statistica test del DM è negativa e il segno della funzione di perdita del GW è negativo. Ciò implica che, secondo i test svolti in questa sede, il modello "migliore" per la previsione della variabile alla frequenza trimestrale (y_t), è il MIDAS-VAR(2).

In sintesi, anche l'analisi empirica conferma le osservazioni emerse dall'esame sulle serie simulate. L'approccio a frequenze miste aggiunge informazione rilevante rispetto all'approccio a bassa frequenza, e tale informazione in più si sostanzia in una maggiore precisione della previsione.

Interessante è anche analizzare la funzione di risposta d'impulso ottenuta attraverso il modello strutturale a frequenze miste e confrontarla con quella ottenuta dal modello a bassa frequenza. Gli esempi numerici delle sezioni precedenti suggeriscono che potremmo aspettarci funzioni di risposta d'impulso piuttosto diverse da quelle dei tradizionali modelli VAR rispetto a quelli a frequenza mista. Inoltre, gli esempi numerici nella sezione precedente ci hanno detto che gli shock che si verificano in diversi mesi avranno impatti diversi sulla variabile oggetto di analisi.

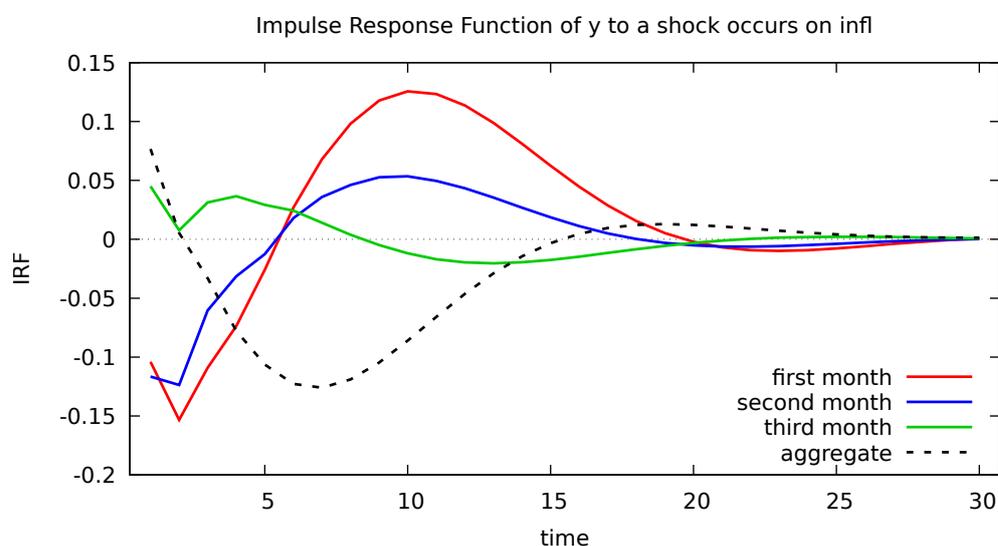


Figura 3.7: IRF di y_t rispetto ad uno shock del primo (linea rossa), secondo (linea blu) e terzo mese (linea verde) di π_t , a confronto con la risposta di y_t ad uno shock della variabile π_t aggregata (linea tratteggiata).

Infatti, la risposta d'impulso dell'output gap rispetto ad uno shock esogeno dell'inflazione (figura 3.7), mostra andamenti piuttosto differenti tra i diversi mesi del trimestre e la soluzione aggregata. In particolare, uno shock dell'inflazione al primo e al secondo mese dell'inflazione, sembra comportare una risposta di entità maggiore rispetto a quella della serie aggregata, che sembra invece più *smussata*. Per lo shock avvenuto al primo o al secondo mese, la risposta istantanea è negativa. In seguito si ha un aggiustamento verso l'alto che porta la funzione a valori positivi, prima di convergere verso lo zero.

Al contrario, il percorso che segue la funzione risposta d'impulso di y_t rispetto ad uno shock esogeno della disoccupazione (figura 3.8), è lo stesso nei mesi del trimestre e nella soluzione aggregata. Ciò che cambia è la magnitudo della risposta istantanea, che risulta maggiore in valore

assoluto per la serie a bassa frequenza rispetto a quelle ad alta frequenza.

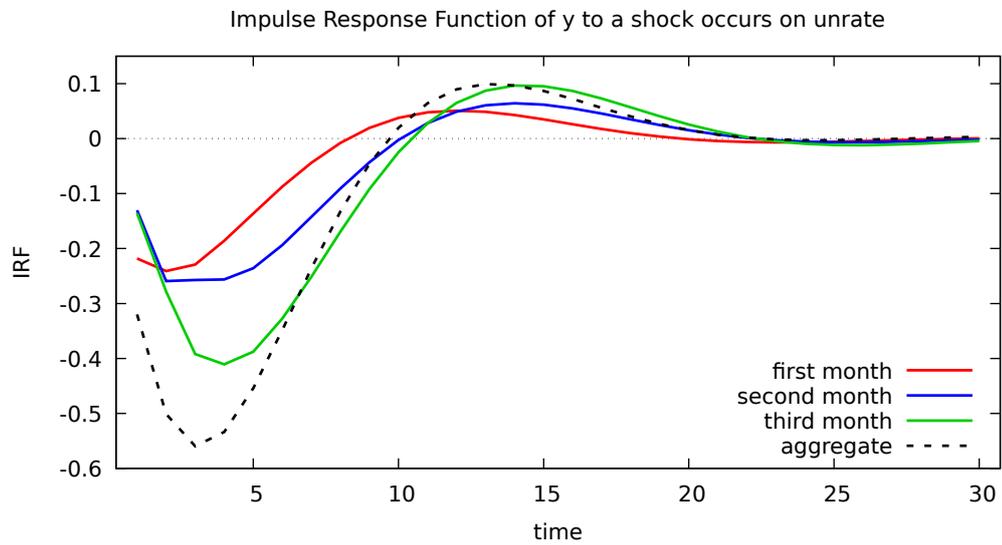


Figura 3.8: IRF di y_t rispetto ad uno shock del primo (linea rossa), secondo (linea blu) e terzo mese (linea verde) di u_t , a confronto con la risposta di y_t ad uno shock della variabile u_t aggregata (linea tratteggiata).

In conclusione, l'analisi empirica ci conferma che, attraverso l'aggregazione temporale, si scartano o distorcono informazioni potenzialmente rilevanti. Ne è un chiaro esempio quello della funzione di risposta d'impulso, la quale mostra comportamenti piuttosto differenti rispetto all'approccio a frequenze miste e quello alla più bassa frequenza.

3.5 Disaggregazione temporale e benchmarking

Fino ad ora, si è parlato del problema dell'aggregazione temporale nei modelli di serie storiche, e di come esso si possa risolvere o mitigare attraverso l'utilizzo di approcci a frequenze miste.

In quest'ultimo paragrafo, invece, presentiamo un approccio alternativo o addirittura contrario chiamato appunto disaggregazione temporale o benchmarking temporale. Questo approccio permette di ripartire (o disaggregare), ad esempio, una serie trimestrale (detto benchmark) in una serie mensile, attraverso delle metodologie che fanno riferimento a delle serie indicatrici mensili.

Dunque l'obiettivo è quello di ottenere una serie, originariamente non osservabile, ad alta frequenza che sia coerente con il benchmark a bassa frequenza osservabile e con gli indicatori di riferimento ad alta frequenza. Si tratta, in altre parole, di interpolare la serie di riferimento ad alta frequenza con quella a bassa frequenza con l'intento di stimare gli m valori all'interno del periodo a bassa frequenza.

Maronilli (2019) ha sviluppato un pacchetto di funzioni per il software `gret1` che offre la possibilità di implementare le seguenti tecniche di benchmarking:

- Naive benchmarking;
- Pro-rata benchmarking;
- Original Denton (Denton, 1971) ;
- Modified Denton (Cholette, 1983);
- Dagum-Cholette (Cholette e Dagum, 1994);

- Chow-Lin (Chow e Lin, 1971).

Ricalcando l'applicazione empirica esposta nel paragrafo precedente, stimiamo il PIL reale mensile, usando ad esempio la serie mensile della disoccupazione come indicatore di riferimento. La tecnica di benchmarking utilizzata è il modified Denton con funzione obiettivo basata sulle prime differenze proporzionali (per dettagli si veda Di Fonzo, 2003). Quello che si ottiene è una serie mensile Y_t del PIL reale.

A questo punto, si applica il filtro di Hodrick-Prescott alla serie mensile in modo da ottenere la serie dell'output gap mensile stimata.

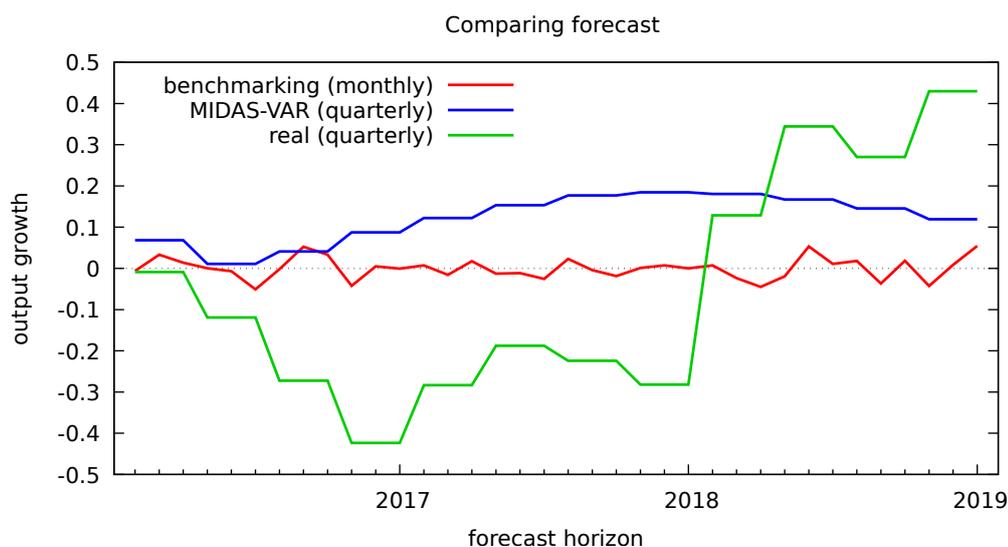


Figura 3.9: Confronto della previsione dell'output gap nel periodo che va dal secondo mese del 2016 al primo del 2019, ottenuta attraverso il benchmarking (linea rossa), il MIDAS-VAR(2) (linea blu) coi i dati veri (linea verde).

Nella figura 3.9 si raffrontano i risultati della previsione ottenuta con la tecnica di benchmarking scelta, con quella ottenuta tramite il MIDAS-VAR e i dati veri.

3.5. DISAGGREGAZIONE TEMPORALE E BENCHMARKING

In conclusione, il breve esercizio appena eseguito, ha l'obiettivo di presentare un approccio alternativo nella gestione di dataset campionati a frequenze miste.

È ovvio che in letteratura di strategie per trattare dataset di questo tipo ne esistono delle più disparate, ma il benchmarking e la disaggregazione temporale sono risultati particolarmente interessanti per due motivi. In primis perché, recentemente, Maronilli (2019) ha sviluppato un pacchetto che permette l'implementazione di tali tecniche nel software libero gretl. In secondo luogo poiché la tecnica di disaggregazione temporale si presenta, concettualmente, esattamente contraria alla prassi comune di aggregazione delle serie alla più bassa frequenza.

Capitolo 4

Conclusioni

La recente attenzione per i dati a frequenza mista, ha portato gli studiosi a fornire modelli in grado di gestire set di dati composti da variabili campionate a frequenza diversa. Uno degli strumenti econometrici di questo tipo è rappresentato dal cosiddetto VAR a frequenze miste. Tale classe di modelli si basa sullo studio congiunto della variabile di interesse e dei previsori, tramite un approccio multivariato.

Da questo punto di vista i contributi più rilevanti sono stati proposti da Mariano e Murasawa (2003, 2010) e Ghysels (2016). Nel terzo capitolo del presente studio si è cercato di replicare, con alcune differenze, il recente lavoro di Ghysels, con l'obiettivo di mostrare i pro e i contro di un'impostazione econometrica a frequenze miste.

I risultati delle simulazioni e dell'analisi empirica illustrano come i modelli *mixed frequency* offrono un concreto vantaggio nella previsione, in determinate condizioni. Infatti, se la variabile disponibile alla frequenza

più alta risulta essere un buon previsore per la variabile a bassa frequenza¹, allora l'informazione in più che sfrutta un modello a frequenze miste si sostanzia in un minor errore di previsione.

Inoltre, lo stesso Ghysels (2016), ma anche Forni e Marcellino (2014, 2016), indagano le conseguenze dell'aggregazione temporale sull'analisi strutturale dei modelli VAR, ed in particolare si concentrano sulla funzione di risposta d'impulso.

Soprattutto Forni e Marcellino (2014) concludono che: *“ l'approccio comune nella letteratura empirica di aggregare i dati alla frequenza più bassa disponibile, è inefficiente e può distorcere l'identificazione degli shock strutturali e il loro meccanismo di propagazione”*.

Quindi, anche nella nostra analisi, abbiamo confrontato le risposte d'impulso ad alta frequenza con quella aggregata, notando una considerevole differenza. In particolare, la risposta d'impulso della variabile oggetto di studio ad uno shock esogeno della variabile ad alta frequenza, sembra avere portata differente se lo shock avviene nel primo, nel secondo o nel terzo mese del trimestre.

In conclusione è doveroso dire che la portata di tali risultati ottenuti non ha la presunzione di essere generale; per risultati più completi si rimanda a studi più elaborati e complessi. Nel presente lavoro ci si limita a concludere che, per le determinate condizioni analizzate, l'approccio a frequenze miste offre dei vantaggi per quanto riguarda la previsione delle serie storiche. Il modello VAR a frequenze miste aggiunge, inoltre, informazioni potenzialmente rilevanti per quanto riguarda l'analisi strutturale.

¹Da un'analisi preliminare a bassa frequenza o da una semplice correlazione.

Appendice A

Codici gretl

Codice 1: MIDAS-VAR su serie simulate.

```
set echo off
set messages off
clear
set seed 1234

# We analyze three variables x, y, z.
# x is a monthly variable while y and z are quarterly variables.
# The ratio of sampling frequencies, m, is equal to 3.
# construct a 5 x 1 mixed frequency vector X = [x1, x2, x3, y, z]'.

## Step 1: Initial setting
# general setting
T = 100          # sample size is 100 quarters, a realistic size.
m = 3           # ratio of sampling frequencies (month vs. quarter)
K_H = 1         # one high frequency variable x
K_L = 2         # two low frequency variables y, z
K = K_L + m*K_H # dimension of MF-VAR
n = K_H + K_L   # number of series
p = 1           # VAR lag length included.
h = 20         # forecast horizon
par = 1        # 1, 2 or 3 select the parameter setting
aggr = 1       # 1 or 2 select aggregation method
                # 1 = average sampling ; 2 = point in time sampling
```

```

# true lag order is 1.

nulldata T --preserve

# Let's generate data from this DGP and fit MF-VAR(1) to see what happens.
# Assume DGP is MF-VAR(1) with coefficient:
if par == 1
  matrix A = {0, 0.1 , 0.4 , 0 , 0;\
    0.1 ,-0.1 , 0.2 , 0 , 0;\
    0 , 0 , 0.1 , 0 , 0;\
    0.0 , -0.9 , 0.9 , 0.2 , 0;\
    0 , 0 , 0 , 0.9 , 0.6}
elif par == 2
  matrix A = { 0.3, 0.1 , 0.2 , 0 , 0;\
    0.1 , -0.4 , -0.15 , 0 , 0;\
    0 , 0.1 , -0.3 ,0, 0;\
    0.2 , -0.5 , 0.68 , 0.3 , 0;\
    0 , 0 , 0 , 0.75 , 0.5}
elif par == 3
  matrix A = { 0.3, 0.1 , 0.2 , 0 , 0;\
    0.1 , -0.4 , -0.15 , 0 , 0;\
    0 , 0.1 , -0.3 ,0, 0;\
    0.2 , 0.5 , 0.68 , 0.3 , 0;\
    0.5 , -0.35 , -0.2 , 0.7 , 0.4}
endif

## Step 2: Mixed frequency analysis
# generate VAR(1) process

matrix Data = varsimul(A, mnormal(T, K), zeros(1, K))
Data = Data[2:,] # cut initial value
# Convert matrix in series
loop i = 1..m --quiet
  series x$i = Data[,i]
endloop
series y = Data[,4]
series z = Data[,5]
#MF-VAR(1) Ghysels(2016)
#smpl ; -h
MFVAR <- var p x1 x2 x3 y z --impulse-responses --silent
# Cholesky decomposition
matrix B = cholesky($sigma)

```

APPENDICE A. CODICI GRETL

```
# store IRF in a matrix
matrix V = $vma

# store IRF of y on x1 x2 x3 shock
matrix IRF_mixed = V * (B ** I(K))
IRF_xtoy = IRF_mixed[,4] ~ IRF_mixed[,9] ~ IRF_mixed[,14]
IRF_xtoz = IRF_mixed[,5] ~ IRF_mixed[,10] ~ IRF_mixed[,15]
#MFVAR.show
COMP = $compan
error = (A - COMP)
fcast --quiet
FC = $fcast

# Low frequency analysis
# for comparison, aggregate x into quarterly frequency

if aggr == 1
    list X = x1 x2 x3
    series x = mean(X) # average sampling
elif aggr == 2
    series x = x3 # point-in-time sampling (end)
endif

tradVAR <- var p x y z --impulse-responses --silent
# Cholesky decomposition
matrix B1 = cholesky($sigma)
# store IRF in a matrix
matrix V1 = $vma
#store IRF of y to x shock
matrix IRF_trad = V1 * (B1 ** I(n))
IRF_xtoy ~= IRF_trad[,2]
IRF_xtoz ~= IRF_trad[,3]
#tradVAR.show
fcast --quiet
FC1 = $fcast

##forecast range
smpl 2 T
print "Forecast RMSEs:"
printf " MFVAR, forecasting y:  %.4f\n", fcstats(y, FC[,4])[2]
printf " MFVAR, forecasting z:  %.4f\n", fcstats(z, FC[,5])[2]
printf " Traditional VAR, forecasting y:  %.4f\n", fcstats(y, FC1[,2])[2]
```

```

printf " Traditional VAR, forecasting z: %.4f\n", fcstats(z, FC1[,3])[2]
# Forecast evaluation test
lossfun = 1 # 1="U-shape", 2="V-shape"
verb = 1 # 1 = print result 0 = silent
cond = 0 # for GW: 0=un-, 1=conditional

# Deibold-Mariano Test
include FEP.gfn --force
# test for y
bundle b = null
b.y = y
b.f1 = FC[,4] # FC 1 series
b.f2 = FC1[,2] # FC 2 series
b.loss = lossfun # 1="U-shape", 2="V-shape"
b.verb = verb
#b.fhor = 1 # for small-sample corrected critical values (default=1)
applyFCtests(&b, "DM")
print b

# test for z
bundle b1 = null
b1.y = z
b1.f1 = FC[,5] # FC 1 series
b1.f2 = FC1[,3] # FC 2 series
b1.loss = lossfun # 1="U-shape", 2="V-shape"
b1.verb = verb
#b1.fhor = 1 # for small-sample corrected critical values (default=1)
applyFCtests(&b1, "DM")
print b1

# Giacomini-White test
# test for y
bundle b = null
b.y = y
b.f1 = FC[,4] # FC 1 series
b.f2 = FC1[,2] # FC 2 series
b.loss = lossfun # 1="U-shape" (default), 2="V-shape"
b.verb = verb
b.cond = cond # 0=un-, 1=conditional
#list b.z = # conditioning variables
#b.fhor = 1 # forecast horizon (default=1)
applyFCtests(&b, "GW")
print b

```

```
# test for z
bundle b1 = null
b1.y = z
b1.f1 = FC[,5] # FC 1 series
b1.f2 = FC1[,3] # FC 2 series
b1.loss = lossfun # 1="U-shape" (default), 2="V-shape"
b1.verb = verb
b1.cond = cond # 0=un-, 1=conditional
#list b1.z = # conditioning variables
#b1.fhor = 1 # forecast horizon (default=1)
applyFCtests(&b1, "GW")
print b1
```

Codice 2: Simulazione Monte Carlo.

```
set echo off
set messages off
set verbose off
set seed 1234567
clear
include FEP.gfn --force

## Step 1: Initial setting
# general setting
T = 100          # sample size is 100 quarters, a realistic size.
m = 3           # ratio of sampling frequencies (month/quarter)
K_H = 1         # one high frequency variable x
K_L = 2         # two low frequency variables y, z
n = K_H + K_L   # number of series
K = K_L + m*K_H # dimension of MF-VAR
p = 1           # VAR lag length included.
MC = 1000      # number of Monte Carlo iteration
aggr = 1       # select aggregation method
                #1 = average 2 = point in time
par = 1        # select parameter setting (1,2,3,4) vs.
# true lag order is 1.
nulldata T --preserve

# Let's generate data from this DGP and
# fit MF-VAR(1) to see what happens.
if par == 1
    matrix A = {0, 0.1 , 0.4 , 0 , 0;\
                0.1 ,-0.1 , 0.2 , 0 , 0;\
```

```

    0 , 0 , 0.1 , 0 , 0;\
    0.0 , -0.9 , 0.9 , 0.2 , 0;\
    0 , 0 , 0 , 0.9 , 0.6}
elif par == 2
    matrix A = { 0.3, 0.1 , 0.2 , 0 , 0;\
    0.1 , -0.4 , -0.15 , 0 , 0;\
    0 , 0.1 , -0.3 ,0, 0;\
    0.2 , -0.5 , 0.68 , 0.3 , 0;\
    0 , 0 , 0 , 0.75 , 0.5}
elif par == 3
    matrix A = { 0.3, 0.1 , 0.2 , 0 , 0;\
    0.1 , -0.4 , -0.15 , 0 , 0;\
    0 , 0.1 , -0.3 ,0, 0;\
    0.2 , 0.5 , 0.68 , 0.3 , 0;\
    0.5 , -0.35 , -0.2 , 0.7 , 0.4}
elif par == 4
    matrix A = {0.4 , 0.1 , -0.2 , 0.12 , 0.21;\
    0.1 , 0.3 , -0.14 , 0.09 , -0.2 ;\
    0.08 , -0.2 , -0.37 , -0.1 , -0.24;\
    0.1 , 0.2 , 0.73 , 0.4 , -0.3;\
    -0.05 , -0.2 , - 0.7 , 0.2 , 0.5}

endif

# pre-allocation
matrix fcast_y_MFVAR = zeros(MC,1)
matrix fcast_z_MFVAR = zeros(MC,1)
matrix fcast_y_tradVAR = zeros(MC,1)
matrix fcast_z_tradVAR = zeros(MC,1)
matrix count_DM_y = zeros(MC,1)
matrix count_DM_z = zeros(MC,1)
matrix count_GW_y = zeros(MC,1)
matrix count_GW_z = zeros(MC,1)

## Step 2: Mixed frequency analysis
loop i = 1..MC --quiet

# generate VAR(1) process

matrix Data = varsimul(A, mnormal(T, K), zeros(1, K))
Data = Data[2:,:]
# Convert matrix in series

```

APPENDICE A. CODICI GRETL

```
loop j = 1..m --quiet
    series x$j = Data[,j]
endloop
series y = Data[,4]
series z = Data[,5]

#MF-VAR(1) à la Ghysels(2016)
var p x1 x2 x3 y z --silent
fcast --quiet
matrix FC = $fcast
matrix B = cholesky($sigma)

# Low frequency analysis
if aggr == 1
    list X = x1 x2 x3
    series x = mean(X) # average sampling
elif aggr == 2
    series x = x3 # point-in-time sampling (end)
endif

var p x y z --silent
fcast --quiet
matrix FC1 = $fcast

# Forecast range
smpl 2 T
fcast_y_MFVAR[i] = fcstats(y, FC[,4])[2]
fcast_z_MFVAR[i] = fcstats(z, FC[,5])[2]
fcast_y_tradVAR[i] = fcstats(y, FC1[,2])[2]
fcast_z_tradVAR[i] = fcstats(z, FC1[,3])[2]

# Forecast evaluation test
lossfun = 1 # 1="U-shape", 2="V-shape"
verb = 0 # 1 = print result 0 = silent
cond = 0 # for GW: 0=un-, 1=conditional

# Deibold-Mariano Test
# test for y
bundle b = null
b.y = y
b.f1 = FC[,4] # FC 1 series
b.f2 = FC1[,2] # FC 2 series
```

```

b.loss = lossfun # 1="U-shape", 2="V-shape"
b.verb = verb
#b.fhor = 1 # for small-sample corrected critical values (default=1)
applyFCtests(&b, "DM")

# test for z
bundle b1 = null
b1.y = z
b1.f1 = FC[,5] # FC 1 series
b1.f2 = FC1[,3] # FC 2 series
b1.loss = lossfun # 1="U-shape", 2="V-shape"
b1.verb = verb
#b1.fhor = 1 # for small-sample corrected critical values (default=1)
applyFCtests(&b1, "DM")

if b.DMpvalssc < 0.05 && b.DMstat < 0
  count_DM_y[i] = 1
else
  count_DM_y[i]= 0
endif

if b1.DMpvalssc < 0.05 && b1.DMstat < 0
  count_DM_z[i] = 1
else
  count_DM_z[i]= 0
endif

# Giacomini-White test
# test for y
bundle b = null
b.y = y
b.f1 = FC[,4] # FC 1 series
b.f2 = FC1[,2] # FC 2 series
b.loss = lossfun # 1="U-shape" (default), 2="V-shape"
b.verb = verb
b.cond = cond # 0=un-, 1=conditional
#list b.z = # conditioning variables
#b.fhor = 1 # forecast horizon (default=1)
applyFCtests(&b, "GW")

# test for z
bundle b1 = null
b1.y = z

```

APPENDICE A. CODICI GRETL

```
b1.f1 = FC[,5] # FC 1 series
b1.f2 = FC1[,3] # FC 2 series
b1.loss = lossfun # 1="U-shape" (default), 2="V-shape"
b1.verb = verb
b1.cond = cond # 0=un-, 1=conditional
#list b1.z = # conditioning variables
#b1.fhor = 1 # forecast horizon (default=1)
applyFCtests(&b1, "GW")
if b.GWpval < 0.05 && b.sign < 0
    count_GW_y[i] = 1
else
    count_GW_y[i]= 0
endif

if b1.GWpval < 0.05 && b1.sign < 0
    count_GW_z[i] = 1
else
    count_GW_z[i]= 0
endif
endloop

scalar mRMSE_mixed_y = meanc(fcast_y_MFVAR)
scalar mRMSE_mixed_z = meanc(fcast_z_MFVAR)
scalar mRMSE_trad_y = meanc(fcast_y_tradVAR)
scalar mRMSE_trad_z = meanc(fcast_z_tradVAR)
scalar std_mixed_y = var(fcast_y_MFVAR)
scalar std_mixed_z = var(fcast_z_MFVAR)
scalar std_trad_y = var(fcast_y_tradVAR)
scalar std_trad_z = var(fcast_z_tradVAR)
scalar DMwin_MIDASVAR_y = sum(count_DM_y)
scalar DMwin_MIDASVAR_z = sum(count_DM_z)
scalar GWwin_MIDASVAR_y = sum(count_GW_y)
scalar GWwin_MIDASVAR_z = sum(count_GW_z)

## summury of the result
printf "Parameter setting = A%.0f\n",par
if aggr == 1
    print "Aggregation method: average sampling "
elif aggr == 2
    print "Aggregation method: point-in-time sampling"
endif
print "Forecast RMSEs (in sample): MEAN      STD.ERR"
```

```

printf "      MFVAR, forecasting y:  %.8f , %8f\n", mRMSE_mixed_y , std_mixed_y
printf "      MFVAR, forecasting z:  %.8f , %8f\n", mRMSE_mixed_z , std_mixed_z
printf "      Traditional VAR, forecasting y:  %.8f , %8f\n", mRMSE_trad_y , std_trad_y
printf "      Traditional VAR, forecasting z:  %.8f , %8f\n", mRMSE_trad_z , std_trad_z
printf "      DM test prefer MIDASVAR for forecast y: %.0f\n ", DMwin_MIDASVAR_y
printf "      DM test prefer MIDASVAR for forecast z: %.0f\n ", DMwin_MIDASVAR_z
printf "      GW test prefer MIDASVAR for forecast y: %.0f\n ", GWwin_MIDASVAR_y
printf "      GW test prefer MIDASVAR for forecast z: %.0f\n ", GWwin_MIDASVAR_z

```

Codice 3: Analisi empirica.

```

set echo off
set messages off
set horizon 30
clear
include DiebMar.gfn
include FEP.gfn --force
include tbench.gfn

# load data
open fedstl.bin --quiet
data gdpc1
data cpiaucsl --compact=spread # create midas list for cpi
data u2rate --compact=spread # create midas list for u
# data fedfunds --compact=spread

# some useful stuff
K_H = 2 # number of high frequency series
K_L = 1 # number of low frequency series
n = K_H + K_L # number of series
m = 3 # ratio of sampling frequency (month vs quarter)
K = K_L + m*K_H # dimension of MFVAR
aggr = 1 # set aggregation method 1=averare, 2=point-in-time
p = 2 # lag order, from information criteria use 2 or 3
bench = 0 # 1 do the benchmark analysis, else do not

# to collect forecasts
matrix FC = {}
matrix FC1 = {}

# form list of high-frequency log differences
list dX = cpiaucsl_m1 cpiaucsl_m2 cpiaucsl_m3
list infl = hflldiff(dX,100)

```

APPENDICE A. CODICI GRETL

```
list u = u2rate_m1 u2rate_m2 u2rate_m3
series y = 100*hpfilt(ln(gdpc1))
rename 10 infl_m3
rename 9 infl_m2
rename 8 infl_m1
# low frequency aggregated series
if aggr == 1 #average sampling
    series infla = mean(infl)
    series unrate = mean(u)
elif aggr == 2 # point in time sampling last value
    series infla = infl[3]
    series unrate = u[3]
endif

# estimation sample
smpl 1970:1 2016:1

#MFVAR à la Ghysels (2016) estimated by OLS
var 12 infl u y --lagselect
MFVAR <- var p infl u y --silent
fcast --out-of-sample --quiet
matrix FC = $fcast

# structural analysis: IRF (high)
matrix B = cholesky($sigma) # Cholesky decomposition of the covariance matrix
matrix V = $vma
matrix IRF_mixed = V * (B ** I(K))
IRF_infltoy = IRF_mixed[,7] ~ IRF_mixed[,14] ~ IRF_mixed[,21]
IRF_utoy = IRF_mixed[,28] ~ IRF_mixed[,35] ~ IRF_mixed[,42]

# low frequency analysis: traditional VAR
var 12 infla unrate y --lagselect
tradVAR <- var p infla unrate y --silent
fcast --out-of-sample --quiet
matrix FC1 = $fcast

# structural analysis: IRF (low)
matrix B1 = cholesky($sigma)
matrix V1 = $vma
matrix IRF_trad = V1 * (B1 ** I(n))
IRF_infltoy ~= IRF_trad[,3]
IRF_utoy ~= IRF_trad[,6]
```

```

##forecast range
smpl 2016:2 2019:1
print "Forecast RMSEs (out of sample):"
printf " MFVAR, forecasting y:  %.4f\n", fcstats(y, FC[,7])[2]
printf " Traditional VAR, forecasting y:  %.4f\n", fcstats(y, FC1[,3])[2]

# Forecast evaluation test
lossfun = 1 # 1="U-shape", 2="V-shape"
verb = 1 # 1 = print result 0 = silent
cond = 0 # for GW: 0=un-, 1=conditional
fhor = 1 # forecast horizon, for small-sample corrected critical values (default=1)

#Deibold-Mariano test
# test for y
bundle b = null
b.y = y
b.f1 = FC[,7] # FC 1 series
b.f2 = FC1[,3] # FC 2 series
b.loss = lossfun # 1="U-shape", 2="V-shape"
b.verb = verb
b.fhor = fhor
applyFCtests(&b, "DM")
print b

# Giacomini-White test
# test for y
bundle b1 = null
b1.y = y
b1.f1 = FC[,7] # FC 1 series
b1.f2 = FC1[,3] # FC 2 series
b1.loss = lossfun # 1="U-shape" (default), 2="V-shape"
b1.verb = verb
b1.cond = cond # 0=un-, 1=conditional
#list b1.z = # conditioning variables
b1.fhor = fhor
applyFCtests(&b1, "GW")
print b1

# plot different IRFs
# gnuplot 1 2 3 4 --time-series --with-lines --matrix=IRF_infltoy --output=display
# gnuplot 1 2 3 4 --time-series --with-lines --matrix=IRF_utoy --output=display

if bench == 1 # benchmarking and MIDAS

```

```
series fymix = FC[,7]
dataset expand
matrix cy = {y}
matrix cmy = {fymix}
clear --dataset
open fedstl.bin
data u2rate
data gdpc1
Y=tbench(gdpc1,u2rate,0,0,8,0.84,1,0)
# forecast range
smpl 2016:02 2019:01
series foutgr = hpfilt(ln(Y))
series foutgr1 = cmy
series foutgr2 = cy
gnuplot foutgr foutgr1 foutgr2 --time-series --with-lines --output=display
endif
```

Bibliografia

ALMON, S. (1965). *The distributed lag between capital appropriations and expenditures*. *Econometrica: Journal of the Econometric Society*, pp. 178–196.

BACCHIOCCHI, E., BASTIANIN, A., MISSALE, A. E ROSSI, E. (2018). *Structural analysis with mixed-frequency data: A midas-svar model of us capital flows*. Rap. tecn., arXiv.org.

BANBURA, M., GIANNONE, D. E REICHLIN, L. (2010). *Nowcasting*. Rap. tecn., CEPR Discussion Papers.

BEVERIDGE, S. E NELSON, C. R. (1981). *A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'*. *Journal of Monetary economics*, 7(2), pp. 151–174.

BOONE, L. (2000). *Comparing semi-structural methods to estimate unobserved variables: The hpmv and kalman filters approaches*. OECD Economics Department Working Papers. No.240.

BIBLIOGRAFIA

- CAMPOS, J., ERICSSON, N. R. E HENDRY, D. F. (1990). *An analogue model of phase-averaging procedures*. *Journal of Econometrics*, 43(3), pp. 275–292.
- CHIRICO, P. *et al.* (2014). *Modelli strutturali e filtri di kalman per serie storiche univariate-teoria ed applicazioni con gretl*. Rap. tecn., University of Turin.
- CHOLETTE, P.-A. (1983). *Adjusting sub-annual series to yearly benchmarks*. Statistics Canada, Methodology Branch, Time Series Research and Analysis
- CHOLETTE, P. A. E DAGUM, E. B. (1994). *Benchmarking time series with autocorrelated survey errors*. *International Statistical Review / Revue Internationale de Statistique*, pp. 365–377.
- CHOW, G. C. E LIN, A.-L. (1971). *Best linear unbiased interpolation, distribution, and extrapolation of time series by related series*. Princeton University.
- CLEMENTS, M. P. E GALVÃO, A. B. (2008). *Macroeconomic forecasting with mixed-frequency data*. *Journal of Business & Economic Statistics*, 26(4), pp. 546–554.
- CLEMENTS, M. P. E GALVÃO, A. B. (2009). *Forecasting us output growth using leading indicators: An appraisal using midas models*. *Journal of Applied Econometrics*, 24(7), pp. 1187–1206.
- COTTRELL, A. E LUCCHETTI, R. (2019a). *Gretl Function Package Guide*. gretl documentation.

- COTTRELL, A. E LUCCHETTI, R. (2019b). *Gretl User's Guide*. gretl documentation.
- 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), pp. 99–102.
- DI FONZO, T. (2003). *Benchmarking di serie storiche economiche. nota tecnica ed estensioni*.
- DIEBOLD, F. X. E MARIANO, R. S. (1995). *Comparing predictive accuracy*. *Journal of Business and Economic Statistics*, 13(3), pp. 253–263.
- EMILI, S. (2017). *Identification and Estimation of Structural Var Models with Mixed Frequency Data: a Moment-Based Approach*. Tesi di dottorato, Alma Mater Studiorum.
- FORONI, C., GUÉRIN, P. E MARCELLINO, M. (2018). *Using low frequency information for predicting high frequency variables*. *International Journal of Forecasting*, 34(4), pp. 774–787.
- FORONI, C. E MARCELLINO, M. (2013). *A survey of econometric methods for mixed-frequency data*. SSRN Electronic Journal.
- FORONI, C. E MARCELLINO, M. (2014). *Mixed-frequency structural models: identification, estimation, and policy analysis*. *Journal of Applied Econometrics*, 29(7), pp. 1118–1144.

BIBLIOGRAFIA

- FORONI, C. E MARCELLINO, M. (2016). *Mixed frequency structural vector auto-regressive models*. Journal of the Royal Statistical Society: Series A (Statistics in Society), 179(2), pp. 403–425.
- FORONI, C., MARCELLINO, M. E SCHUMACHER, C. (2011). *U-midas: Midas regressions with unrestricted lag polynomials*. Discussion Paper Series 1: Economic Studies 35, Deutsche Bundesbank.
- GHYSELS, E. (2016). *Macroeconomics and the reality of mixed frequency data*. Journal of Econometrics, 193(2), pp. 294 – 314.
- GHYSELS, E., HILL, J. B. E MOTEGI, K. (2016). *Testing for granger causality with mixed frequency data*. Journal of Econometrics, 192(1), pp. 207–230.
- GHYSELS, E., RUBIA, A. S. E VALKANOV, R. (2009). *Multi-period forecasts of volatility: Direct, iterated, and mixed-data approaches*. Rap. tecn., University of North Carolina.
- GHYSELS, E., SANTA-CLARA, P. E VALKANOV, R. (2004). *The midas touch: Mixed data sampling regression models*. Série Scientifique.
- GHYSELS, E., SANTA-CLARA, P. E VALKANOV, R. (2005). *There is a risk-return trade-off after all*. Journal of Financial Economics, 76(3), pp. 509–548.
- GIACOMINI, R. E WHITE, H. (2006). *Tests of conditional predictive ability*. Econometrica, 74(6), pp. 1545–1578.
- GRANGER, C. W. J. (1969). *Investigating causal relations by econometric models and cross-spectral methods*. Econometrica: Journal of the Econometric Society, pp. 424–438.

- GRANGER, C. W. J. (1990). *Aggregation of time-series variables: A survey*. In *Disaggregation in Econometric Modelling* (curato da BARKER, T. E PESARAN, M. H.), pp. 17–34. Routledge, London.
- HAMILTON, J. D. (1994). *Econometria delle serie storiche*. Monduzzi.
- HAMILTON, J. D. (2018). *Why you should never use the hodrick-prescott filter*. *Review of Economics and Statistics*, 100(5), pp. 831–843.
- HARVEY, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.
- HODRICK, R. J. E PRESCOTT, E. C. (1997). *Postwar us business cycles: an empirical investigation*. *Journal of Money, credit, and Banking*, pp. 1–16.
- KALMAN, R. E. (1960). *A new approach to linear filtering and prediction problems*. *Journal of basic Engineering*, 82(1), pp. 35–45.
- KALMAN, R. E. (1963). *New methods in wiener filtering theory*. In *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability*, edited by JL Bogdanoff and F. Kozin, John Wiley & Sons, New York.
- KUZIN, V., MARCELLINO, M. E SCHUMACHER, C. (2011). *Midas vs. mixed-frequency var: Nowcasting gdp in the euro area*. *International Journal of Forecasting*, 27(2), pp. 529–542.
- LIPPI, M. E REICHLIN, L. (1991). *Trend-cycle decompositions and measures of persistence: does time aggregation matter?* *The Economic Journal*, 101(405), pp. 314–323.

BIBLIOGRAFIA

- LUCCHETTI, R. (2015). *Appunti di analisi delle serie storiche*.
- LÜTKEPOHL, H. (1987). *Forecasting aggregated vector ARMA processes*, vol. 284. Springer Science & Business Media.
- MARCELLINO, M. (1999). *Some consequences of temporal aggregation in empirical analysis*. *Journal of Business & Economic Statistics*, 17(1), pp. 129–136.
- MARIANO, R. S. E MURASAWA, Y. (2003). *A new coincident index of business cycles based on monthly and quarterly series*. *Journal of applied Econometrics*, 18(4), pp. 427–443.
- MARIANO, R. S. E MURASAWA, Y. (2010). *A coincident index, common factors, and monthly real gdp*. *Oxford Bulletin of Economics and Statistics*, 72(1), pp. 27–46.
- MARONILLI, A. (2019). *Benchmarking e disaggregazione temporale con indicatori di riferimento: un pacchetto per gretl*.
- MORLEY, J. C. (2002). *A state–space approach to calculating the beveridge–nelson decomposition*. *Economics Letters*, 75(1), pp. 123–127.
- PIERSE, R. G. E SNELL, A. J. (1995). *Temporal aggregation and the power of tests for a unit root*. *Journal of Econometrics*, 65(2), pp. 333–345.
- SIMS, C. A. (1971). *Discrete approximations to continuous time distributed lags in econometrics*. *Econometrica* (pre-1986), 39(3), pp. 545.
- SIMS, C. A. (1980). *Macroeconomics and reality*. *Econometrica: journal of the Econometric Society*, pp. 1–48.

STOCK, J. E. E WATSON, M. W. (1991). *A probability model of the coincident economic indicators*. In *The Leading Economic Indicators: New Approaches and Forecasting Records* (curato da MOORE, G. E LAHIRI, K.), pp. 63–90. Cambridge University Press.

SWANSON, N. R. E GRANGER, C. W. J. (1997). *Impulse response functions based on a causal approach to residual orthogonalization in vector autoregressions*. *Journal of the American Statistical Association*, 92(437), pp. 357–367.

WEI, W. W. (1982). *Comment: The effects of systematic sampling and temporal aggregation on causality—a cautionary note*. *Journal of the American Statistical Association*, 77(378), pp. 316–319.

ZADROZNY, P. (1988). *Gaussian likelihood of continuous-time armax models when data are stocks and flows at different frequencies*. *Econometric Theory*, 4(1), pp. 108–124.