Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.3, No.1, 2013 www.iiste.org

Disaggregation Comparison Analysis of Economic Time Series Data

T.O. OLATAYO and K.K. ADESANYA

Department of Mathematical Sciences,Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria. Otimtoy@yahaoo.com and adesanyakehindekazeem@yahoo.com.

Abstract - Econometrics modeling often implies the use of a number of time series data, some of which could be available at lower frequency and therefore, it could be convenient to disaggregate these data to high frequency form instead of estimating with a significant loss of information. The main aim of temporal disaggregation is to derive an estimate of the underlying high frequency (HF) observation of an observed low frequency (LF) time series.

The method adopted by Chow-Lin, Fernadez, Litterman, (static model), and Santo Silvacardoso (dynamic model) were used to make comparison in disaggregation economic analysis of time series data. The parameters employed in this study are Autoregressive test, Correlation and standard Deviation.

Result of analysis in low frequency form (Annual) confirmed that Chow-Lin has the correlation value of 0.9914, Fernandez has the correlation value 0.9914, Litterman has the correlation value 0.9701 and Santo Silvacardoso has the correlation value of 0.9914. Result of analysis in high frequency (monthly) confirmed that Chow-Lin has the correlation value of 0.9899 and Standard Deviation of 212850.48, Fernadez has the correlation value of 0.9899 and Standard Deviation of 78553.54, Litterman has the correlation value of 0.9997 and Standard Deviation of 789109.18 while Santo Silvacardoso has the correlation value of 0.9898 and Standard Deviation of 2337.24.

The performance indicators of disaggregated values for Chow-Lin, Fernandez, Litterman being a static model and Santo Silvacardso being a dynamic model, annual and monthly data confirms that the results of analysis are very good with high correlation figures while the ability of the estimated monthly data captured the true dynamic of the series. Santo Silvacardoso being a dynamic model preformed better with minimum standard deviation while Litterman technique being a classic and static model preformed poorly in disaggregating to high frequency form.

Keywords: *Disaggregation, Low frequency data, High Frequency Data, Static Model, Dynamic model.*

INTRODUCTION

Temporal disaggregation methods play an important role for the estimation of short term economic indicators. The need for temporal disaggregation can stem from a number of reasons, due to the high costs involved in collecting the statistical information needed for estimating national accounts, could decide to conduct large sample surveys only annually.

Consequently, quarterly (or even monthly) national accounts could be obtained through an indirect approach, which is by using related quarterly (or monthly) time series as indicators of the short term dynamics of the annual aggregates. Econometric modeling often implies the use of a number of time series, some of which could be available only at lower frequencies, and therefore, it would be convenient to disaggregate these data instead of estimating, with a significant loss of information, the complete model at lower frequency level. Chow and Lin (1971).

Different strategies have been developed to get an estimate of the autoregressive parameter from the lower frequency data is the most applied procedures are those proposed by Chow and Lin (1971), Bourney and Laroque (1979), other authors have proposed alternatives restriction on the DGP (Data Generation Process) of the disturbance series in the High Frequency regression model. Fernadez (1981) proposes a random walk model for the disturbances that avoids the estimation of parameter at the High Frequency level. Litterman (1983) refines the Fernadez solution by introducing Markov process to take account of serial correlation in the residuals. Moauro and Savio (2002) encompasses the three solution, generalizing the restrictions in the class of ARIMA (Auto Regressive Integrated Moving Average) processes. Recently, some authors have proposed techniques based on dynamic regression models in the identification of the relationship linking the series to be estimated and the related indicators. Aadland (2000).

In this study, efforts will be geared towards disaggregation of low frequency time series data into high frequency time series data, through the comparison of static and dynamic models as propounded by Chow Lin, Litterman, Fernandez and Santo Silvarcardoso.

MATERIALS AND METHODS

The disaggregation of low frequency data (annual) to high frequency data (monthly), thereafter using both the low and high frequency data. Different strategies that have been developed to get an estimate of the autoregressive parameter from the lower frequency data, the method adopted by Chow-Lin, Fenandez, Litterman and Santo Silvacardoso were used to examine the performance indicators of estimate of private consumption expenditure.

(a) The Chow-Lin Model

The Chow-Lin (1971) disaggregation method is based on the assumption that y_t can be represented by a linear regression model with first order autoregressive errors;

$y_t = \propto_t + X_t^1\beta, t = \emptyset \propto_{t-1} + \varepsilon_{t,} \ \varepsilon_t \sim N\emptyset(0,\sigma^2)$

with $/\emptyset/ < 1$ and $\propto_t \ \sim N(0,\sigma^2/(1-\emptyset^2))$

The model is thus a particular case with scalar system matrices Z=1, T=$\emptyset$, H=1. As far the initial coordinates are concerned as α_t is a stationary zero mean AR(1) process, it is assumed that the process applies since time immemorial, giving $\alpha_1 \sim N(0,\sigma^2/(1-\emptyset^2)$ which amounts to setting α =0, W_1=0, and $H_1=(1-\emptyset^2)^{-1/2}$

If some elements of x_t are non stationary, the CL model postulates full cointegration between them and the series y_t.

Deterministic components (such as a linear trend) are handled by including appropriate regressors in the set X_t e.g. by setting X_t, and writing ;

$$y_t = \mu_t + \sum_j \beta;X_{jt} + \propto_t$$

with the first two elements of B being denoted μ and alternatively, they can be accommodated in the transistor equation, which becomes

$$\alpha_t = \emptyset\alpha_{t-1} + M + g_t + \varepsilon_t.$$

The state space form corresponding to this case features $W_t = (1.t.0^1)$ for $t > 1$, whereas $W_1 = [\ (1-\Phi)^{-1}(1-2\Phi)/(1-\Phi)^2,\ 0^1]$. The first two elements of the vector of exogenous regressive X_t are zero, since M and g do not enter the measurement equation.

(b) The Litterman and Fernadez models

According to the litterman (1983) model, the temporally disaggregated process is a regression model with ARIMA disturbances.

$y_t = x^1{}_t\,\beta + \mu_t\,,\ \Delta\mu_t = \Phi\Delta\mu_{t-1} + \varepsilon_t.$

Litterman explicitly assumes that the v_t process has started off at time t=0 with $\mu_o = \Delta\mu_o$, (litterman 1983). This is usually inadequate, unless the set of indicator include constant (which would capture the effect of initial value); the inclusion of a linear trends amounts to allowing for non-zero drift in the ARIMA process.

The Fernandez (1981) model arises in the particular case when Φ=0 and thus μ_t is a random walk.

The state space representation is obtained by defining the state vector and system matrices as follows;

$$A_t = \begin{bmatrix} \mu_{t-1} \\ \Delta_{ut} \end{bmatrix} - Z^1 = (1,1).\ T = \begin{bmatrix} 1 & 1 \\ 0 & \Phi \end{bmatrix} \quad H = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The Litterman initialization implies $\mu_1 = \mu_0 + \Phi\mu_o + \varepsilon_1 = \varepsilon_t$ which is implemented casting;

$$\alpha_1 = 0,\ W_1 = 0,\ H_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Alternatively including μ_1 in the vector β as its first element, in which case x_t features a zero element in first positive, and assuring that the stationary process has started in the indefinite past, the initial conditions are:

$$W_1 = \begin{bmatrix} 1 & 0^1 \\ 0 & 0^1 \end{bmatrix} \quad H_1 = \frac{1}{\sqrt{1-Q^2}} \qquad \begin{bmatrix} 1 \\ \Phi \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

This follows from writing:

and taking $\Delta\mu_o \sim N(0, \delta^2/(1-Ø^2)$, $\varepsilon_1 \sim N(0,\delta^2)$. The diffuse nature arises from the non-stationary of the model.

It should be noticed that in this second setup we cannot include a constant in x_t, since this effect is captured by μ_1.

Finally, the ARIMA process can be extended to include a constant and a trend in $\Delta\mu_t = \Delta\mu_{t-1}\, m + g_t + \varepsilon_t$; the parameters m and g are incorporated in the vector β and the matrices W_1 and W_t are easily extended; for instance, if $\beta = (\mu_{-1}, m, g, \beta^1{}_2)^1$ where β_2 corresponds to the regression effects affecting only the measurement equation.

(c) Santo Silvacardoso Model

Santo Silvacardoso considered a slightly different representative of the underlying high frequency data which emplicitly takes into account the presence of the tagged dependent variable;

$$(1-ФL)\, y_{t\cdot\mu} = x_{t.\mu}\beta_{t\cdot} + \varepsilon_{t\cdot\mu},$$

The solution for y_h is given by $\begin{cases} t=1 & \mu = 2, \ldots s \\ t=2,\ldots T & \mu = 1, 2 \ldots s \end{cases}$

where $\begin{bmatrix} Y_n{}^1 \\ \bar{\bar{Y}}_n{}^2 \end{bmatrix} \begin{bmatrix} A_1A_1 & T_{12} \\ T_{21} & A_2A_2 \end{bmatrix}^{-1} \begin{bmatrix} A_1Z_n + C_i & (C_1 (A_iA_i)^{-1} Ci)^{-1} & yi - C1n \\ A_2Z_n + C_2{}^1 & (C_2 (A^1{}_2A_2)^{-1}Ci)^{-1} & (y_1{}^2 - C_2 (A^1{}_2 A_2)^{-1} A_2Z^*{}_n \end{bmatrix}$

$$T_{12} = A_1A_2 \; -C_1 \quad (C_1(A_1A_1)^{-1}C_1) \quad C_1(A_iA_1)^{-1} \quad A_iA_2$$
$$T_{21} = A_2A_1 \; -C_2{}^1 \quad (C_2 (A_2A_2)^{-1} \quad C_2(A_2A_2)^{-1} \; A^1{}_2 A_1$$

The method of samtos silva and Cardoso. For notation convenience, let t= s(t-1) + μ be the interm running on the periods and re-write as follows;

$$Y_t = \phi y_{r-1} + X_r\,\beta + \varepsilon_T = 1,\ldots.n$$

$$Y_t = \left(\sum_{i=0}^{t-1} O^i x_{i-1}\right) \beta + Ø^T y_o + \left(\sum_{i=0}^{t-1} O^i \varepsilon_{T-1}\right)$$

RESULTS

The evaluation of the results were made with respect to both Low and High Frequencies data, with the results that examined 'economic reasonableness' of the annual and estimated monthly regression model and, eventually, by comparing different estimates. A number of estimates have been calculated according to different specifications, both in the original data and in log-transformed form.

Summary of the results obtained in disaggregation comparison of economic time series data of annual Nigeria GDP from CBN (1981-2009):

Table 1: Unit roots tests for annual Nigeria GDP (1981-2009) :low frequency

statistic	Intercept and trend		
	None	Only intercept	both
Levels			
ADF(1)	0.144685	0.123473	0.041476
	(0.023166)	(0.027735)	(0.045681)
PP	0.144685	0.123473	0.041476
	(0.023166)	(0.027735)	(0.045681)
First difference			
ADF(1)	0.034904	-0.071382	-0.448537
	(0.119209)	(0.145472)	(0.255566)
PP	-0.282981	-0.427759	-0.995286
	(0.136712)	(0.163720)	(0.238537)
Log-levels			
ADF(1)	0.015485	-0.004487	-0.229935
	(0.002449)	(0.016359)	(0.130912)
PP	0.015485	-0.004487	-0.229935
	(0.002449)	(0.016359)	(0.130912)
First difference in logs			
ADF(1)	-0.303820	-0.835713	-0.828735
	(0.140703)	(0.199238)	(0.202855)
PP	-0.303820	-0.835713	-0.828735
	(0.140703)	(0.199238)	(0.202855)

Mackinnon 5% critical values for rejection of hypothesis of a unit root in parentheses. PP test statistics have been calculated using 3 lags truncation for Bartlett Kernel (Newey and West, 1994). From the table of low frequency of GDP, the study reported that ADF and PP at level with their respective values greater than (-3.34) Mackinnon 5% , thus unit root is present, therefore annual GDP is cointegrated, hence no error correction model. At first difference, ADF and PP values are each greater than 5% Mackinnon hence there is present of unit root, therefore there exist cointegration. The study reported at log level both ADF and PP results shows that the GDP cointegrated at both lag 1 and lag 3 for ADF and PP respectively since their respective values each is greater than Mackinnon 5% . Likewise, we found out that ADF and PP indicate cointegration since their values each is greater than Mackinnon 5%. The distributed lagged model specified for their relationships were stable for control of action and prediction.

Table 2 Unit roots tests for annual Nigeria PCE (1981-2009) :low frequency

Statistic	Intercept and trend		
	None	Only intercept	both
Levels			
ADF(1)	0.132804	0.106901	-0.007913
	(0.037072)	(0.045731)	(0.079247)
PP	0.132804	0.106901	-0.007913
	(0.037072)	(0.045731)	(0.079247)
First difference			
ADF(1)	-0.443656	-0.872442	-1.202197
	(0.228404)	(0.197424)	(0.202624)
PP	-0.688948	-0.872442	-1.202197
	(0.187493)	(0.197424)	(0.202624)
Log-levels			
ADF(1)	0.016026	-0.024149	-0.477070
	(0.005752)	(0.036345)	(0.176411)
PP	0.016026	-0.024149	-0.477070
	(0.005752)	(0.036345)	(0.176411)
First difference in logs			
ADF(1)	-0.988757	-1.320202	-1.321400
	(0.196094)	(0.189513)	(0.192816)
PP	-0.988757	-1.320202	-1.321400
	(0.196094)	(0.189513)	(0.192816)

Mackinnon 5% critical values for rejection of hypothesis of a unit root in parentheses. PP test statistics have been calculated using 3 lags truncation for Bartlett Kernel (Newey and West, 1994). From the table of low frequency of PCI, the study reported that ADF and PP at level with their respective values greater than (-3.34) Mackinnon 5%, thus unit root is present, therefore annual PCI is cointegrated, hence no error correction model. At first difference, ADF and PP values are each greater than 5% Mackinnon hence there is present of unit root, therefore there exist cointegration. The study reported at log level both ADF and PP results shows that the PCI cointegrated at both lag 1 and lag 3 for ADF and PP respectively since their respective values each is greater than Mackinnon 5%. Likewise, we found out that ADF and PP indicate cointegration since their values each is greater than Mackinnon 5%. The distributed lagged model specified for their relationships were stable for control of action and prediction.

Annual Nigerian National Account

The annual series aggregated and the chosen indicator is estimated above. As confirmed by the unit roots tests (table 1 and 2). Moreover, the residual based ADF test τe (table 3) is coherent with the hypothesis of cointegration.

Table 3 : Residual-based cointegration tests: ADF(1) on Nigeria national accounts

	τe	$\tau\alpha$
Levels	-2.632536	-2.553634
Log levels	-2.700094	-2.644761
5% asymptotic	-3.34	-3.78

Table 4: estimates of the auxiliary annual regression on Nigeria national account (PCE)

Variants	A	β	η	Φ
1	48.686	0.670884	0.601003	-0.405550
	(19.625)	(0.164094)	(0.129655)	(0.202708)
2		0.684426	0.593367	-0.412778
		(0.151929)	(0.123654)	(0.202077)
3	0.414404	0.392369	0.597102	0.032398
	(0.141003)	(0.058722)	(0.060532)	(0.190810)
4		0.367033	0.651491	0.151401
		(0.066069)	(0.065554)	(0.187903)

Table 4 contains parameters' estimates for dynamic models in both levels and logarithms, and precisely according to variants 1, 2, 3 and 4 (that is, model in levels with or without intercept, and model in logs with intercept or without intercept, which in this last case turns out to be significant). Concentrating on the estimates obtained through variants (1 and 2) with 3 and 4, we find that the Low Frequency estimated values are very similar.

Table 5 Unit roots tests for monthly Nigeria GDP (1981-2009) :high frequency

Statistic	Intercept and trend		
	None	Only intercept	both
Levels			
ADF(4)	0.000239	0.000194	-3.12E-05
	(7.57E-05)	(8.16E-05)	(0.000125)
PP	0.009858	0.008258	0.002204
	(0.000620)	(0.000744)	(0.001295)
First difference			
ADF(4)	-0.016152	-0.021480	-0.031404
	(0.003631)	(0.004167)	(0.005053)
PP	-0.008279	-0.010343	-0.006392
	(0.007090)	(0.008190)	(0.010017)
Log-levels			
ADF(4)	0.000353	-0.000939	-0.047182
	(0.000218)	(0.001156)	(0.008042)
PP	0.001249	0.000787	-0.070416
	(0.000458)	(0.002457)	(0.014402)
First difference in logs			
ADF(4)	-0.242826	-0.252969	-0.252892
	(0.030019)	(0.030444)	(0.030541)
PP	-0.182021	-0.186895	-0.186789
	(0.030246)	(0.030558)	(0.030660)

Mackinnon 5% critical values for rejection of hypothesis of a unit root in parentheses. PP test statistics have been calculated using 4 lags truncation for Bartlett Kernel (Newey and West, 1994).

From the table of high frequency of GDP, the study reported that ADF and PP at level with their respective values greater than (-3.34) Mackinnon 5% , thus unit root is present, therefore annual GDP is cointegrated, hence no error correction model. At first difference, ADF and PP values are each greater than 5% Mackinnon hence there is present of unit root, therefore there exist cointegration. The study reported at log level both ADF and PP results shows that the GDP cointegrated at both lag 1 and lag 3 for ADF and PP respectively since their respective values each is greater than Mackinnon 5% . Likewise, we found out that ADF and PP indicate cointegration since their values each is greater than Mackinnon 5%. The distributed lagged model specified for their relationships were stable for control of action and prediction.

Table 6: Unit roots tests for monthly Nigeria PCE (1981-2009) :high frequency

Statistic	Intercept and trend		
	None	Only intercept	both
Levels			
ADF(4)	0.000383	0.000300	-9.59E-05
	(0.000122)	(0.000141)	(0.000245)
PP	0.010299	0.008759	0.002948
	(0.000990)	(0.001219)	(0.002222)
First difference			
ADF(4)	-0.032360	-0.037469	-0.043962
	(0.004980)	(0.005318)	(0.005725)
PP	-0.017307	-0.020109	-0.022043
	(0.009975)	(0.010740)	(0.011679)
Log-levels			
ADF(4)	8.58E-05	-5.18E-05	-0.000876
	(3.10E-05)	(0.000136)	(0.000525)
PP	0.001616	3.52E-05	-0.003513
	(0.000229)	(0.001064)	(0.003967)
First difference in logs			
ADF(4)	-0.047802	-0.055155	-0.055161
	(0.006439)	(0.006844)	(0.006855)
PP	-0.025820	-0.029747	-0.029693
	(0.012155)	(0.013053)	(0.013073)

Mackinnon 5% critical values for rejection of hypothesis of a unit root in parentheses PP test statistics have been calculated using 4 lags truncation for Bartlett Kernel (Newey and West, 1994).

From the table of high frequency of PCE, the study reported that ADF and PP at level with their respective values greater than (-3.34) Mackinnon 5% , thus unit root is present, therefore annual PCE is cointegrated, hence no error correction model. At first difference, ADF and PP values are each greater than 5% Mackinnon hence there is present of unit root, therefore there exist cointegration. The study reported at log level both ADF and PP results shows that the PCE cointegrated at both lag 1 and lag 3 for ADF and PP respectively since their respective values each is greater than Mackinnon 5% . Likewise, we found out that ADF and PP indicate cointegration since their values each is greater than Mackinnon 5%. The distributed lagged model specified for their relationships were stable for control of action and prediction.

Monthly Disaggregation of Annual Nigerian National Account

The series disaggregated and the chosen indicators are estimated above. As confirmed by the unit roots tests (table 4.4, 4.5 and 4.6a), both series are I(1). Moreover, the residual based ADF test τe (table 4.6b) is coherent with the hypothesis of cointegration.

Table 7 : Residual-based cointegration tests :ADF(4) on Nigeria national accounts

	Te	Tα
Levels	-3.363091	-3.364926
Log levels	-3.450114	-3.483024
5% asymptotic	-3.34	-3.78

*Davidson and Mackinnon(1993), Table 20.2 p.722

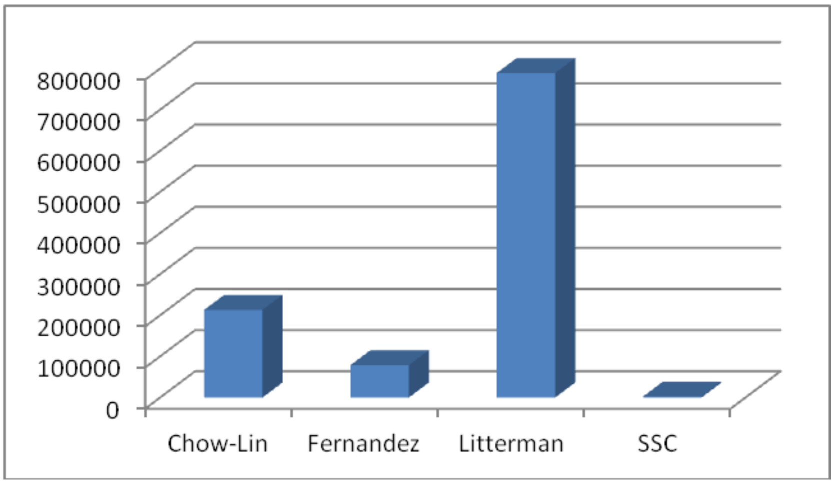
Table 8: estimates of the auxiliary monthly regression on Nigeria national account (PCE)

variants	A	B	η	Φ
1	17.074	0.029475	0.987704	-4.68E-05
	(5.875)	(0.007142)	(0.005033)	(0.003571)
2		0.033687	0.985958	1.19E-05
		(0.007069)	(0.005051)	(0.000103)
3	0.013367	0.022680	0.980698	0.477684
	(0.013570)	(0.011506)	(0.010481)	(0.025889)
4		0.018026	0.986979	0.357109
		(0.010490)	(0.008317)	(0.023430)

Table 8 contains parameters' estimates for dynamic models in both levels and logarithms, and precisely according to variants 1, 2, 3 and 4 (that is, model in levels with or without intercept, and model in logs with intercept or without intercept, which in this last case turns out to be significant). Concentrating on the estimates obtained through variants 2, 3 and 4, we find that the HF estimated values are very similar.

Table 9: Disaggregation comparison indicators of private consumption Expenditure

PCE	Chow-Lin	Fernandez	Litterman	SSC
Annual % changes				
Correlation	0.9914	0.9914	0.9701	0.9914
Monthly % changes				
Correlation	0.9899	0.9899	0.9997	0.9898
Standard dev.	212850.48	78553.54	789109.18	2337.24



Estimated (for Chow-Lin, Fernandez, Litterman (Static models) and Santo Silva-Cardoso(Dynamic model), annual and monthly confirms that the results are surely very good with high correlation figures while the ability of the estimated monthly data capture the 'true' dynamics of the series. Santo Silva Cardoso being a dynamic model performed better with minimum standard deviation while Litterman technique a classical and static model performed poorly from the disaggregation of Monthly national account data.

DISCUSSION OF RESULTS

Qualitative analysis very often has to rely on data whose observation frequency is systematically lower than desired. However, as it would require enormous resources to actually observe this process, most countries use annual estimates of economic activity as the basis of their statistics. In contrast, many other variables such as money stock and interest rates are available at a far higher frequency (and can often also be observed more accurately). Nevertheless, researchers , policy makers and public, all have genuine interest in high frequency information on low frequency data for efficient and timely decision making. Therefore, statistical offices all around the world work on providing temporarily disaggregated data to serve this aim.

CONCLUSION

The performance indicators of disaggregated estimates of private consumption expenditure estimates (for chow-lin, Fernandez, litterman) being a static model and santo silvacardoso being a dynamic model, annual and monthly data confirms that the results of analysis are very good with high correlation figures while the ability of the estimated monthly data capture the true dynamic of the series. Santo silvacardoso being a dynamic model performed better with minimum standard deviation while litterman technique being a classical and static model performed poorly from the disaggregating of monthly national account data.

REFERENCES

Aadland D.M. (2000) Distribution and Interpolation Using Transformed Data, Journal of Applied Statistics.

Bourney and Laroque (1979): An analysis of transformations (with discussion), Journal of the Royal Statistical Society B 26, 211-246.

Chow G., Lin A.L. (1971): Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series. The Review of Economics and Statistics.

Di Fonzo T. (2003), Temporal disaggregation of economic time series: towards a dynamic extension, European Commission (Eurostat) Working Papers and Studies, Theme 1, General Statistics.

Fernandez, P.E.B. (1981), A methodological note on the estimation of time series, the Review of Economics and Statistics, 63, 3, 471-478.

Litterman, R.B. (1983), A random walk, Markov model for the distribution of time series, Journal of Business and Economic Statistics, 1, 2, pp. 169-173.

Moauro F. and Savio G. (2002): Temporal Disaggregation Using multivariate Structural Time Series Models. Fortcoming in the Econometric Journal.

Salazar, E.L., Smith, R.J. and Weale, M. (1997). Interpolation Using a Dynamic Regression Model: Specification and Monte Carlo Properties, NIESR Discussion Paper n. 126.