

ESTIMATION OF COINTEGRATING VECTORS WITH TIME SERIES MEASURED AT DIFFERENT PERIODICITY

GABRIEL PONS
University of Aarhus

ANDREU SANSÓ
University of the Balearic Islands

We discuss the effects of temporal aggregation on the estimation of cointegrating vectors and on testing linear restrictions on this vector. We adopt a discrete time approach and demonstrate, in contrast with the findings of Chambers (2003, *Econometric Theory* 19, 49–77), who adopts a continuous time approach, that in some situations, when the regressand must be aggregated, systematic sampling is preferable to average sampling for estimation purposes. Like Chambers, we show that the best aggregation scheme for regressors, in terms of asymptotic estimation efficiency, is always average sampling. We also show that different types of aggregation have no influence on the relative size of tests of linear restrictions on the cointegration vector.

1. INTRODUCTION

Economic theories are frequently tested not with data measured at their generating time interval, but with temporally aggregated time series because most economic time series are measured at a lower sampling frequency than economic agents make decisions. Temporal aggregation changes many time series properties such as weak exogeneity or Granger causality and in some cases deteriorates statistical inference because valuable information is lost with such data transformation and cannot be recovered with the observable data (see Wei, 1989; Marcellino, 1999). However, the estimation of cointegrating vectors does

We thank Soren Johansen, Niels Haldrup, Raquel Waters, the associate editor, and two anonymous referees for their helpful comments. Of course, any remaining error is the responsibility of the authors. The first author gratefully acknowledges the financial support of a Marie Curie Fellowship of the European Community Programme "Improving the Human Research Potential and the Socio-Economic Knowledge Base" under contract HPMF-CT-2002-01662 and the Danish Research Council. The second author gratefully acknowledges the financial support of the Spanish Ministry of Science and Technology SEC2002-01512. Address correspondence to Gabriel Pons, Institute of Economics, University of Copenhagen, Denmark; e-mail: gabriel.pons@econ.ku.dk.

not fit into this pessimistic picture because an estimator of cointegrating vectors constructed with temporally aggregated data is consistent (see Granger, 1990; Phillips, 1991a) and, when all the variables of the cointegrated system are flows, asymptotically as efficient as the estimator based on disaggregated time series (see Chambers, 2003).¹

There are different methodological approaches to discuss the effects of temporal aggregation on cointegrated systems. The continuous time aggregation analysis (see Phillips, 1991a; Comte, 1999; Chambers, 2003) assumes that the long-run dynamics evolves in continuous time. This assumption is not realistic for some cointegrating relations because the continuous approach imposes that the markets react instantaneously to a long-run disequilibrium, but in practice many cointegrating relations, such as the purchasing power parity (PPP) (see Johansen and Juselius, 1992), correct a situation of long-run disequilibrium very slowly as a result of information costs or barriers to trade. Another approximation to the issue is the discrete time aggregation that considers that the long-run dynamics are generated at equally spaced time intervals (see Granger, 1990; Granger and Siklos, 1995; Marcellino, 1999). This approach also can be considered an approximation to discuss temporal aggregation of cointegrating relations because it assumes an equally spaced time interval between shocks and long-run reactions, and in practice the time interval between the shock and the long-run reaction may vary along the sample period.²

The aim of this paper is an in-depth examination of the theory of the effects that discrete time aggregation has on inference on cointegrating vectors. We consider the effects of different types of aggregation on the estimation of cointegrating relations with $I(1)$ and $I(2)$ variables and on hypothesis testing on cointegrating vectors. As stressed by Johansen (1994, 1995) the marginal distribution of the estimator is not relevant for inference on cointegrating vectors. What is of interest is the joint distribution of the estimator and the observed information. We provide evidence from a simulation study to confirm this statement for our specific aggregation scenario. More specifically, we show that the aggregation scheme most able to guarantee the estimation's precision is not necessarily the best suited for testing a hypothesis on the cointegrating vector and, in addition, although different aggregation schemes have different effects on the estimator, they have similar effects on the test.

This paper is organized as follows. In Section 2 we present a representation of the cointegrated system with a varying, but equally spaced, sampling interval. The dependency of the optimal regression-type estimators and the observed information on temporal aggregation are discussed in Section 3. Next, we evaluate the finite-sample effects of temporal aggregation on the Phillips and Hansen (1990) fully modified ordinary least squares (FM-OLS) estimator. Finally, Section 5 concludes the paper. The Appendix contains the proofs of a lemma and theorems.

We shall use the following notation: \mathbf{B} denotes a vector of Brownian motions, $\mathbf{\Omega}$ the long-run variance matrix, and \mathbf{W} a vector of standardized indepen-

dent Brownian motions; $CI(d, b)$ stands for a cointegrated process of orders (d, b) ; $[\cdot]$ denotes the integer part and \Rightarrow weak convergence; \mathbf{z}_t represents a k -dimensional disaggregated time series; $T \equiv [t/m]$ denotes the aggregated time unit; $N \equiv [n/m]$ is the aggregated sample size, t is the disaggregated time unit; m is a finite positive integer that determines the length of the sampling interval; and n denotes the span of the sample measured in t -periods. The k -dimensional temporally aggregated time series is defined as

$$\mathbf{Z}_T^{(j)} \equiv [s_m(L)]^j \mathbf{z}_{mT},$$

where $s_m(L) = 1 + L + \dots + L^{m-1}$ is the summation filter. More specifically, $\mathbf{Z}_T^{(0)}$ denotes a *systematically sampled* time series and $\mathbf{Z}_T^{(1)}$ an *average sampled* time series.³ Finally, we decompose the disaggregated time series into $\mathbf{z}_t = (\mathbf{y}'_t, \mathbf{x}'_t)'$, where \mathbf{y}_t is $k_0 \times 1$ and \mathbf{x}_t is $k_1 \times 1$ ($k_0 + k_1 = k$) and we express the mixed sampled time series as

$$\mathbf{Z}_T^{(j_0, j_1)} \equiv (\mathbf{Y}_T^{(j_0)'}, \mathbf{X}_T^{(j_1)'})' = ([s_m(L)]^{j_0} \mathbf{y}'_{mT}, [s_m(L)]^{j_1} \mathbf{x}'_{mT})',$$

where $\mathbf{Y}_T^{(j_0)}$ is temporally aggregated by scheme j_0 and $\mathbf{X}_T^{(j_1)}$ is temporally aggregated by a different scheme j_1 ($\neq j_0$).

2. A REPRESENTATION OF THE COINTEGRATED SYSTEM FOR A VARYING SAMPLING INTERVAL

In this section we present a representation of the cointegrated system where the length of the sampling interval m is not fixed. This representation is used to discuss the problem of temporal aggregation that occurs when a practitioner makes inference on cointegration with data measured at a longer sampling interval than the long-run dynamics are generated. Hence, we use the term *disaggregated model* to describe the cointegrated system measured at the frequency where the long-run dynamics are generated, i.e., the time interval between a shock that alters the equilibrium and the moment the markets react to the disequilibrium. We use the triangular system representation of Phillips (1991b) to do so, because this model avoids having to specify the short-run dynamics and the aggregation effect on the long-run variance can be determined exactly.

2.1. Disaggregated Model

Suppose that the long-run dynamics of a k -dimensional $I(1)$ time series $\mathbf{z}_t = (\mathbf{y}'_t, \mathbf{x}'_{1,t})'$ were generated at time interval Δt by the process

$$\Delta \mathbf{x}_{1,t} = \mathbf{u}_{2,t}, \tag{1}$$

$$\mathbf{y}_t - \mathbf{x}'_{1,t} \boldsymbol{\beta}_1 = \mathbf{u}_{1,t}, \tag{2}$$

where \mathbf{y}_t is k_0 -dimensional, $\mathbf{x}_{1,t}$ is k_1 -dimensional, and $\mathbf{u}_t \equiv (\mathbf{u}'_{1,t}, \mathbf{u}'_{2,t})'$ is a stationary process that satisfies the invariance principle

$$n^{-1/2} \sum_{t=1}^{\lfloor nr \rfloor} \mathbf{u}_t \Rightarrow \mathbf{B} \equiv \boldsymbol{\Omega}^{1/2} \mathbf{W} \quad (3)$$

for $r \in [0, 1]$ as $n \rightarrow \infty$, where \mathbf{W} denotes a k -dimensional standardized Brownian motion with variance matrix $\boldsymbol{\Omega} \equiv \lim_{n \rightarrow \infty} n^{-1} E[(\sum_1^n \mathbf{u}_j)(\sum_1^n \mathbf{u}_j)']$, the long-run variance of \mathbf{u}_t . Let us partition $\boldsymbol{\Omega}$ according to \mathbf{u}_t :

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix}$$

and define the conditional long-run variance $\boldsymbol{\Omega}_{11.2} \equiv \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12} \boldsymbol{\Omega}_{22}^{-1} \boldsymbol{\Omega}_{21}$. We assume that the matrix $\boldsymbol{\Omega}_{22}$ is positive definite in such a way that $\mathbf{x}_{1,t}$ is individually $I(1)$ but not $CI(1, 1)$. Under these assumptions, the process \mathbf{z}_t is a $CI(1, 1)$ process with k_0 cointegrating relations and k_1 common stochastic trends.

Representation (1)–(3) has been used to analyze optimal inference on cointegrating vectors (see Phillips, 1991b) and to design a number of optimal regression-type estimators. Phillips and Hansen (1990) and Park (1992) propose semiparametric methods, Saikkonen (1991) and Stock and Watson (1993) dynamic regression methods, and Phillips (1991c) a frequency domain method. A very important property of optimal estimators of cointegrating vectors is their asymptotic mixed Gaussian distribution, meaning that a test for g restrictions on $\boldsymbol{\beta}_1$ is χ_g^2 distributed (Phillips, 1991b). However, regression-type methods are not very well suited to model building because they impose the location and number of unit roots on the model rather than testing them. This has two important consequences. First, it is not possible to test sequentially for the number of cointegration relations, because a triangular model with k'_0 cointegrating relations is not nested into a triangular model with k''_0 ($>k'_0$) cointegrating relations (see Johansen, 1994). Second, regression-type estimators impose normalization restrictions on cointegrating vectors, and this may lead to inconsistent inference if such restrictions do not hold. In addition, these methods were originally designed to test for a single cointegrating relation ($k_0 = 1$), and this restricts their applicability to very small systems. However, even small systems can be governed by more than one cointegrating relation, and although it is possible to estimate more than one cointegrating relation with regression methods, the properties of the estimator and tests for the extra cointegrating relations are not clear (see Hargreaves, 1994). For all these reasons, system methods that test for the dimension of the cointegrating space and do not impose normalization restrictions, such as the full maximum likelihood analysis of cointegrated vector autoregression (VAR) (see Johansen, 1991), are commonly used for model building.

The aim of this paper is to study the effects of different aggregation options on inference rather than on model building, and the triangular model is more suitable for this purpose than other representations of the cointegrated system,

such as the VAR model or the moving average (MA) model, because of its nonparameterization of the short-run dynamics.

The main focus of the paper is CI(1,1) relations, but we consider two particular cases of cointegration with I(2) variables. To do so, let us suppose that the long-run dynamics of a k -dimensional time series $\mathbf{z}_t \equiv (\mathbf{y}'_t, \mathbf{x}'_{1,t}, \mathbf{x}'_{2,t})'$ are generated by the following model:

$$\Delta^2 \mathbf{x}_{2,t} = \mathbf{u}_{3,t}, \tag{4}$$

$$\Delta \mathbf{x}_{1,t} - \Delta \mathbf{x}'_{2,t} \boldsymbol{\beta}_3 = \mathbf{u}_{2,t}, \tag{5}$$

$$\mathbf{y}_t - \mathbf{x}'_{1,t} \boldsymbol{\beta}_1 - \mathbf{x}'_{2,t} \boldsymbol{\beta}_2 = \mathbf{u}_{1,t}, \tag{6}$$

where $\mathbf{x}_{2,t}$ is k_2 -dimensional ($k_0 + k_1 + k_2 = k$), $\mathbf{u}_t = (\mathbf{u}'_{1,t}, \mathbf{u}'_{2,t}, \mathbf{u}'_{3,t})'$ satisfies (3), and $\boldsymbol{\Omega}_{22}$ and $\boldsymbol{\Omega}_{33}$ are positive definite. This is a very general model that allows for many different cointegrating relationships. Matrix $[I_{k_1}, -\boldsymbol{\beta}_3]$ denotes the CI(2,1) cointegrating vectors of $(\mathbf{x}'_{1,t}, \mathbf{x}'_{2,t})'$, matrix $[I_{k_0}, -\boldsymbol{\beta}_2]$ contains CI(2,2) cointegrating vectors of $(\mathbf{y}'_t, \mathbf{x}'_{2,t})'$ when some or all the elements of $\boldsymbol{\beta}_1$ are zero, and matrix $[I_{k_0}, -\boldsymbol{\beta}_1]$ contains the CI(1,1) cointegrating vectors of $(\mathbf{y}'_t, \mathbf{x}'_{1,t})'$ when some or all the vectors of matrix $\boldsymbol{\beta}_2$ are zero. For the nonrestricted case, $[I_{k_0}, -\boldsymbol{\beta}_2]$ denotes the CI(2,1) cointegrating vectors of $(\mathbf{y}'_t, \mathbf{x}'_{2,t})'$, and $[I_{k_0}, -\boldsymbol{\beta}_1]$ denotes the CI(1,1) vectors of $(\mathbf{y}'_t - \mathbf{x}'_{2,t} \boldsymbol{\beta}_2, \mathbf{x}'_{1,t})'$. In this general setting, Stock and Watson (1993) proposed dynamic optimal inference methods and Kitamura (1995) full and partial information maximum likelihood methods. Under the very restrictive assumption that $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{13} = \boldsymbol{\Omega}_{23} = \mathbf{0}$, the ordinary least squares (OLS) estimator is optimal in the sense that it has a mixed Gaussian distribution allowing for a standard χ^2 inference (see Haldrup, 1994).

2.2. Aggregated Model

Let us use $\mathbf{Z}_T^{(j_0, j_1, j_2)}$ (with $j_0, j_1, j_2 = \{0, 1\}$) to denote the observable vector time series that is measured at a longer sampling interval $\Delta_m T = m \Delta t$ than \mathbf{z}_t . The representation of this temporally aggregated process is given in the following lemma.

LEMMA 2.1. *If \mathbf{z}_t is generated by (3)–(6), the observable temporally aggregated time series $\mathbf{Z}_T^{(j_0, j_1, j_2)}$ can be represented as*

$$\Delta_m^2 \mathbf{X}_{2,T}^{(j_2)} = \mathbf{U}_{3,T}^{(j_2+2)}, \tag{7}$$

$$\Delta_m \mathbf{X}_{1,T}^{(j_1)} - m^{j_1-j_2} \Delta_m \mathbf{X}_{2,T}^{(j_2)'} \boldsymbol{\beta}_3 = \mathbf{U}_{2,T}^{(j_1+1)}, \tag{8}$$

$$\mathbf{Y}_T^{(j_0)} - m^{j_0-j_1} \mathbf{X}_{1,T}^{(j_1)'} \boldsymbol{\beta}_1 - m^{j_0-j_2} \mathbf{X}_{2,T}^{(j_2)'} \boldsymbol{\beta}_2 = \mathbf{U}_{1,T}^{(j_0)}, \tag{9}$$

where $L^m \mathbf{Z}_T^{(j_0, j_1, j_2)} = \mathbf{Z}_{T-1}^{(j_0, j_1, j_2)}$, $\Delta_m \equiv 1 - L^m$, and $\mathbf{U}_T^{(j_0, j_1+1, j_2+2)} \equiv (\mathbf{U}_{1,T}^{(j_0)'}, \mathbf{U}_{2,T}^{(j_1+1)'}, \mathbf{U}_{3,T}^{(j_2+2)'})'$ satisfies the multivariate invariance principle

$$N^{-1/2} \sum_1^{[Nr]} \mathbf{U}_T^{(j_0, j_1+1, j_2+2)} \Rightarrow \begin{bmatrix} \boldsymbol{\Omega}_{11}^{(j_0)} & m^{j_0+j_1} \boldsymbol{\Omega}_{12} & m^{j_0+j_2+1} \boldsymbol{\Omega}_{13} \\ m^{j_0+j_1} \boldsymbol{\Omega}_{21} & m^{2j_1+1} \boldsymbol{\Omega}_{22} & m^{j_1+j_2+2} \boldsymbol{\Omega}_{23} \\ m^{j_0+j_2+1} \boldsymbol{\Omega}_{31} & m^{j_2+j_1+2} \boldsymbol{\Omega}_{32} & m^{2j_2+3} \boldsymbol{\Omega}_{33} \end{bmatrix}^{1/2} \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \mathbf{W}_3 \end{bmatrix},$$

where $\boldsymbol{\Omega}_{11}^{(0)} = m^{-1}(\boldsymbol{\Omega}_{11} + \Lambda(m))$, $\boldsymbol{\Omega}_{11}^{(1)} = m\boldsymbol{\Omega}_{11}$, $\Lambda(m) = 4\pi \sum_{j=1}^{m-1} F_{11}(2\pi j/m)$, and $F_{11}(\omega)$ is the spectral density matrix of $\mathbf{u}_{1,t}$.

Proof. See the Appendix.

From Lemma 2.1, the number of cointegrating relations is invariant with any of the different combinations of aggregation schemes. However, the cointegrating spaces are not generally invariant with temporal aggregation. More specifically, when the aggregation schemes applied to the regressand and the respective regressors are the same ($j_t = j_h$), the cointegrating relations are invariant. If not, when some of the variables are aggregated with systematic sampling and others with average sampling, the cointegrating space changes with m . The weights $m^{j_t-j_h}$ capture this change in cointegrating relations, which is exclusively attributable to a change in the unit of measurement. Ericsson, Hendry, and Tran (1994) analyze the effects of linear filters, $z_t^a = g(L)z_t$, on cointegrating vectors. They show that when all the filtered time series are measured by the same units, $g(1) = I$, then the filtered series z_t^a and the original series z_t have the same cointegrating space. This condition does not hold for mixed sampling, because the flow variables are measured in relation to the sampling interval, i.e., that month's income but not the stocks, i.e., money supply.

The second main effect of temporal aggregation on the cointegrated system is a change in the short-run dynamics. The different effects of temporal aggregation on the error terms of the cointegrated system change the short-run dynamics and also their contribution to the long run, represented by the aggregated long-run variance $\boldsymbol{\Omega}^{(j_0, j_1+1, j_3+2)}$. For any possible combination of aggregation schemes, the short-run dynamics will always vary. For example, when the aggregation schemes are the same ($j_0 = j_1 = j_3 = j$), the long-run variance of $\mathbf{U}_{2,T}^{(j+1)}$ increases in relation to the long-run variance of $\mathbf{U}_{1,T}^{(j)}$, and the long-run variance of $\mathbf{U}_{3,T}^{(j+2)}$ increases in relation to the long-run variances of the other errors. Note that when the regressand is systematically sampled, $\mathbf{Y}_T^{(0)}$, the long-run variance of $\mathbf{U}_{1,T}^{(0)}$ is an average of the disaggregated long-run variance $\boldsymbol{\Omega}_{11}$ and the variance of those seasonal cycles that emerge as a result of the aliasing effect, $\Lambda(m)$. For instance, a quarterly stock variable with an important semiannual seasonal cycle ($F_{11}(\pi)$) systematically sampled as a semiannual variable ($m = 2$) will experience a substantial increase in the long-run

variance of the semiannual systematically sampled cointegrating error $\mathbf{U}_{1,T}^{(0)}$. As discussed in the next section, this type of aggregation will make the mixed normal distribution of the estimator of the cointegrating vectors more disperse.

Let us focus on the CI(1,1) model. The following corollary particularizes the previous results to this situation.

COROLLARY 2.2. *If \mathbf{z}_t is generated by (1)–(3), the temporally aggregated time series $\mathbf{Z}_T^{(j_0, j_1)}$ can be represented as*

$$\Delta_m \mathbf{X}_{1,T}^{(j_1)} = \mathbf{U}_{2,T}^{(j_1+1)},$$

$$\mathbf{Y}_T^{(j_0)} - m^{j_0-j_1} \mathbf{X}_{1,T}^{(j_1)'} \boldsymbol{\beta}_1 = \mathbf{U}_{1,T}^{(j_0)},$$

where $\mathbf{U}_T^{(j_0, j_1+1)} \equiv (\mathbf{U}_{1,T}^{(j_0)}, \mathbf{U}_{2,T}^{(j_1+1)})$ satisfies the multivariate invariance principle

$$N^{-1/2} \Sigma_1^{[Nr]} \mathbf{U}_T^{(j_0, j_1+1)} \Rightarrow \begin{bmatrix} \boldsymbol{\Omega}_{11}^{(j_0)} & m^{j_0+j_1} \boldsymbol{\Omega}_{12} \\ m^{j_0+j_1} \boldsymbol{\Omega}_{21} & m^{2j_1+1} \boldsymbol{\Omega}_{22} \end{bmatrix}^{1/2} \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix},$$

where $\boldsymbol{\Omega}_{11}^{(j_0)}$ is defined as before.

Proof. It follows from Lemma 2.1.

In this case, we can differentiate four different temporal aggregation situations. We use the same classification as Chambers (2003) and identify the pure systematic sampling situation as aggregation type I ($\mathbf{Z}_T^{(0)} = (Y_T^{(0)'}, X_{1,T}^{(0)'})'$), the mixed sampled process $\mathbf{Z}_T^{(0,1)} = (Y_T^{(0)'}, X_{1,T}^{(1)'})'$ as aggregation type II, the other mixed sampled situation $\mathbf{Z}_T^{(1,0)} = (Y_T^{(1)'}, X_{1,T}^{(0)'})'$ as aggregation type III, and finally the pure average sampled process ($\mathbf{Z}_T^{(1,1)} = (Y_T^{(1)'}, X_{1,T}^{(1)'})'$) as aggregation type IV. Table 1 shows the aggregation effect of these different aggregation schemes for $m = 3$ on the long-run variances $\boldsymbol{\Omega}_{11,2}^{(j_0)}$, $\boldsymbol{\Omega}_{22}^{(j_1+1)}$, and on $\Xi^i(m) \equiv m^{j_0-j_1} (\boldsymbol{\Omega}_{11,2} + (1 - j_0)\Lambda(m))^{1/2}$ ($i = \text{I, II, III, IV}$), which, as shown in the next section, is the relative dispersion of the mixed normal distribution across m for the different aggregation schemes.

TABLE 1. Temporally aggregated long-run variances of the CI(1,1) model for $m = 3$

Aggregation scheme	$\boldsymbol{\Omega}_{11,2}^{(j_0)}$	$\boldsymbol{\Omega}_{22}^{(j_1+1)}$	$\Xi^i(3)$
I ($j_0 = 0, j_1 = 0$)	$\frac{1}{3}(\boldsymbol{\Omega}_{11,2} + \Lambda(3))$	$3\boldsymbol{\Omega}_{22}$	$[\boldsymbol{\Omega}_{11,2} + \Lambda(3)]^{1/2}$
II ($j_0 = 0, j_1 = 1$)	$\frac{1}{3}(\boldsymbol{\Omega}_{11,2} + \Lambda(3))$	$27\boldsymbol{\Omega}_{22}$	$\frac{1}{3}[\boldsymbol{\Omega}_{11,2} + \Lambda(3)]^{1/2}$
III ($j_0 = 1, j_1 = 0$)	$3\boldsymbol{\Omega}_{11,2}$	$3\boldsymbol{\Omega}_{22}$	$3\boldsymbol{\Omega}_{11,2}^{1/2}$
IV ($j_0 = 1, j_1 = 1$)	$3\boldsymbol{\Omega}_{11,2}$	$27\boldsymbol{\Omega}_{22}$	$\boldsymbol{\Omega}_{11,2}^{1/2}$

Note: The aliasing component is $\Lambda(3) = 4\pi F_{11}(2\pi/3)$.

Let us consider a particular case of the I(1) cointegrated model. Assume that $k_0 = k_1 = 1$, $u_{1,t} = \rho u_{1,t-1} + \varepsilon_{1,t}$ with $|\rho| < 1$, $\varepsilon_{1,t}$ is i.i.d. $N(0,1)$, $u_{2,t} = \varepsilon_{2,t}$ is i.i.d. $N(0,\sigma^2)$, and $E(\varepsilon_{1,t} \varepsilon_{2,t-j}) = 0$ for all j . The systematically sampled model is given by

$$\Delta_m X_{1,T}^{(0)} = U_{2,T}^{(1)},$$

$$Y_T^{(0)} - \beta_1 X_{1,T}^{(0)} = U_{1,T}^{(0)},$$

where $U_{2,T}^{(1)} = E_{2,T}^{(1)} = s_m(L)\varepsilon_{2,mT}$, and $U_{1,T}^{(0)} = \rho^m U_{1,T-1}^{(0)} + E_{1,T}^{(0)}$, $E_{1,T}^{(0)} = [(1 - \rho^m L^m)/(1 - \rho L)]\varepsilon_{1,mT}$ is serially uncorrelated and not correlated with $E_{2,T}^{(1)}$ for all lags and leads. From Pesaran and Shin (1996) the impact of a unit shock to the variable $Y_T^{(0)}$ on the cointegrating relation after one T -period is given by the autoregressive coefficient ρ^m . Then, given $|\rho| < 1$, the longer the sampling interval with respect to the long-run dynamics generating time interval, the faster the adjustment of the system to a disequilibrium. Let us consider the two limiting aggregation cases. In the maximum aggregation case, i.e., when $m \rightarrow \infty$, all the adjustment takes place in one period ($\lim_{m \rightarrow \infty} \rho^m = 0$), whereas for the maximum disaggregation case, when $m \rightarrow 0$, no adjustment takes place ($\lim_{m \rightarrow 0} \rho^m = 1$). Therefore, the finer the sampling interval the closer the cointegrated model is to a noncointegrated system.

This example illustrates that when the long-run dynamics are measured at a much longer sampling interval than their generating time interval, then the speed of adjustment of the cointegrating relation should be very fast. In practice, for many cointegrated relations the estimated speed of adjustment with monthly or quarterly time series is low, and this suggests that the long-run dynamics for many cointegrated systems are not generated at a much shorter time interval.

3. ASYMPTOTIC DISTRIBUTIONS OF AGGREGATED OPTIMAL COINTEGRATED REGRESSIONS

We will now discuss the asymptotic effects of different types of aggregation on the estimation of β_1 , β_2 , and β_3 and on the observed information. Even though we focus on optimal CI(1,1) regressions we also provide theoretical results for two examples of CI(2,2) and CI(2,1) processes.

3.1. Temporally Aggregated Mixed Normal Distribution

Suppose that a k -dimensional time series $\mathbf{z}_t = \{(y_t, \mathbf{x}'_{1,t})'\}_{t=1}^n$, where y_t is a scalar time series⁴ and $\mathbf{x}_{1,t}$ is a k_1 -dimensional time series, is generated by

$$\Delta \mathbf{x}_{1,t} = \mathbf{u}_{2,t}, \tag{10}$$

$$y_t - \mathbf{x}'_{1,t} \beta_1 = u_{1,t}, \tag{11}$$

where $\mathbf{u}_t = (u_{1,t}, \mathbf{u}'_{2,t})'$ is a stationary process that satisfies (3). Let $\hat{\boldsymbol{\beta}}_1^i$ denote an optimal regression-type estimator of $\boldsymbol{\beta}_1$ using the time series $\mathbf{Z}_T^{(j_0, j_1)}$ where $i = \text{I}$ for $\mathbf{Z}_T^{(0)}$, $i = \text{II}$ for $\mathbf{Z}_T^{(0,1)}$, $i = \text{III}$ for $\mathbf{Z}_T^{(1,0)}$, and $i = \text{IV}$ for $\mathbf{Z}_T^{(1)}$. The following theorem presents its limiting distribution.

THEOREM 3.1. *The asymptotic distribution of the optimal regression-type estimator of $\boldsymbol{\beta}_1$ in model (10) and (11) with temporally aggregated time series is*

$$n(\hat{\boldsymbol{\beta}}_1^i - \boldsymbol{\beta}_1) \Rightarrow \Xi^i(m) \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1},$$

where the relative dispersion across the sampling interval $\Xi^i(m)$ is $\Xi^{\text{I}}(m) = (\boldsymbol{\Omega}_{11.2} + \Lambda(m))^{1/2}$, $\Xi^{\text{II}}(m) = m^{-1}(\boldsymbol{\Omega}_{11.2} + \Lambda(m))^{1/2}$, $\Xi^{\text{III}}(m) = m\boldsymbol{\Omega}_{11.2}^{1/2}$, and $\Xi^{\text{IV}}(m) = \boldsymbol{\Omega}_{11.2}^{1/2}$.

Proof. See the Appendix.

From Theorem 3.1, comparing these asymptotic distributions with the one for an optimal estimator of the disaggregated model, $\boldsymbol{\Omega}_{11.2}^{1/2} \int dW_1 \mathbf{B}'_2 (\int \mathbf{B}_2 \mathbf{B}'_2)^{-1}$, there is no loss of efficiency when the regressand and the regressor are average sampled (aggregation IV), whereas there is a loss of efficiency when the regressors are systematically sampled, i.e., aggregation schemes I and III. These results are in keeping with those of Chambers (2003), who considers a similar scenario where the disaggregated model is a continuous time cointegrated model.

However, in contrast with Chambers (2003), in case II three situations are possible depending on the relative size of the aliasing component $\Lambda(m)$ in relation to the conditional long-run variance $\boldsymbol{\Omega}_{11.2}$. To be more specific, when $\Lambda(m) = (m^2 - 1)\boldsymbol{\Omega}_{11.2}$, which implies that $2\sum_{j=1}^{m-1} F_{11}(2\pi j/m) = (m^2 - 1)F_{11.2}(0)$, there is no loss of efficiency, whereas for $2\sum_{j=1}^{m-1} F_{11}(2\pi j/m) > (m^2 - 1)F_{11.2}(0)$, there is a loss of asymptotic efficiency, and when $2\sum_{j=1}^{m-1} F_{11}(2\pi j/m) < (m^2 - 1)F_{11.2}(0)$ there is a gain in asymptotic efficiency. In any case $\Xi^{\text{II}}(m) < \Xi^{\text{I}}(m)$, and when $\Lambda(m) < (m^4 - 1)\boldsymbol{\Omega}_{11.2}$, then $\Xi^{\text{II}}(m) < \Xi^{\text{III}}(m)$.

To summarize, type II aggregation is always more efficient than type I aggregation, and type IV aggregation is always more efficient than types I and III aggregation. When $\Lambda(m) < (m^2 - 1)\boldsymbol{\Omega}_{11.2}$ type II is more efficient than IV aggregation and therefore is the most efficient option.

Continuing with the example of Section 2.2 and for $m = 3$, a situation of no loss of asymptotic efficiency with aggregation II occurs when $\rho = -0.27$, for $-1 < \rho < -0.27$ there is a loss of efficiency, whereas for $-0.27 < \rho < -1$ there is a gain of efficiency.

Let us compare the relative dispersions $\Xi^i(m)$ with those derived with the continuous time aggregation approach (see Chambers, 2003), which we denote as Ξ^i_C . To do so, let us consider a bivariate triangular continuous time cointegrated model:

$$dx_1(t) = u_2(t) dt, \quad t > 0,$$

$$y(t) - \beta_1 x_1(t) = u_1(t), \quad t > 0,$$

where $(u_1(t), u_2(t))'$ is a stationary continuous time process with spectral density function $F_C(\omega)$ with $-\infty < \omega < \infty$. From Chambers (2003, p. 59) and given that for $j \neq 0$ $h(2\pi ij) = (2\pi ij)^{-1}$, $h(2\pi ij) + h(-2\pi ij) = 0$ and $|h(2\pi ij)|^2 = (4\pi^2 j^2)^{-1}$, we obtain the expressions

$$(\Xi_C^I)^2 = \Omega_{11,2} + 2\pi \sum_{j \neq 0} F_{C,11}(2\pi j),$$

$$(\Xi_C^{II})^2 = \Omega_{11,2} + 2\pi \sum_{j \neq 0} F_{C,11}(2\pi j) + 2\pi \beta_1^2 \sum_{j \neq 0} (4\pi^2 j^2)^{-1} F_{C,22}(2\pi j),$$

$$(\Xi_C^{III})^2 = \Omega_{11,2} + 2\pi \beta_1^2 \sum_{j \neq 0} (4\pi^2 j^2)^{-1} F_{C,22}(2\pi j),$$

$$(\Xi_C^{IV})^2 = \Omega_{11,2}.$$

There are important differences between Ξ_C^i and $\Xi^i(m)$. First of all, in the continuous time setting, type II aggregation is the most inefficient option for any $F_C(\omega)$, a result that contrasts with the findings obtained in the discrete time approach where type II aggregation may be the most efficient way to aggregate the time series and in any case is less efficient than schemes I. Second, the cointegration coefficient β_1 appears at the limiting distributions of the discrete time optimal estimator derived from the continuous time cointegrated system, a result that is in contradiction with the optimal theory of estimation of cointegrating vectors (see Saikkonen, 1991; Phillips, 1991b). Finally, in the continuous time approach, the aliasing affects the distributions through the error process of the cointegrated relations $(2\pi \sum_{j \neq 0} F_{C,11}(2\pi j))$, and through the error process of the common stochastic trends $(2\pi \beta_1^2 \sum_{j \neq 0} (4\pi^2 j^2)^{-1} F_{C,22}(2\pi j))$, a surprising result because for both types of aggregation the process $u_2(t)$ is at least cumulated once $\int u_2(t - s) d_s$, a transformation that eliminates the components $F_{C,22}(2\pi j)$ in scheme IV of continuous time aggregation for the errors of the cointegrating relations. Thus, the continuous time aggregation fails to explain the discrete time aggregation results.

Theorem 3.1 has practical implications. If we consider the common situation where the practitioner may decide how to aggregate a stock time series, then we have two different situations depending on whether it is the regressand or the regressor that is temporally aggregated. When a practitioner has to aggregate a stock regressand and the regressor is a stock, the best aggregation option depends on the sign of $\Lambda(m) - (m^2 - 1)\Omega_{11,2}$. If this magnitude is negative, the best option, in terms of estimation efficiency, is to use systematic sampling (aggregation I), whereas if it is positive, the optimal choice is average sampling (aggregation III). When the regressor is a flow, the best option depends on $\Lambda(m) - (m^2 - 1)\Omega_{11,2}$, in the sense that when this parameter is positive the stock regressand should be aggregated by average sampling (aggregation IV), whereas when it is negative the regressand should be aggregated by systematic

sampling (aggregation II). When a practitioner has to aggregate a stock regressor, the best option, in terms of the asymptotic dispersion of the estimator, is always to apply average sampling to the regressor because aggregations II and IV are superior in efficiency to aggregations I and III, respectively. Therefore, the picture that emerges from the discrete time aggregation analysis is more complex than the one obtained from the continuous time approach where the best option seems to be in any situation to apply average sampling. We provide finite-sample evidence favoring our theoretical findings in Section 4.

In practice, when deciding whether to aggregate the stock regressand by systematic or average sampling, we need to compare $\Lambda(m)$ with $(m^2 - 1)\Omega_{11.2}$. However, because the flow variables are observable at a longer sampling interval than the stock variables, it is necessary to get $\Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ from the observable $\Omega_{12}^{(j_0,2)}\Omega_{22}^{(3)^{-1}}\Omega_{21}^{(2,j_1)}$. Then, from the relation

$$\Omega_{12}\Omega_{22}^{-1}\Omega_{21} = \begin{cases} m\Omega_{12}^{(0,2)}\Omega_{22}^{(3)^{-1}}\Omega_{21}^{(2,0)} & \text{if } j_0 = 0 \\ m^{-1}\Omega_{12}^{(0,2)}\Omega_{22}^{(3)^{-1}}\Omega_{21}^{(2,0)} & \text{if } j_1 = 0 \end{cases},$$

we can estimate $\Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ and $\Omega_{11.2}$ and compare the estimations of $\Lambda(m)$ and $(m^2 - 1)\Omega_{11.2}$.

To summarize, when a stock regressand must be temporally aggregated because some of the regressors are not available at the finest sampling interval, then depending on the magnitudes $\Lambda(m)$ and $(m^2 - 1)\Omega_{11.2}$ systematic or average sampling are the best aggregation options in terms of the dispersion of the mixed normal distribution. When a practitioner has to choose an aggregation option for the stock regressor, the best choice is always to apply average sampling.

The main results found for case CI(1,1) can be extended to models with I(2) variables. We do not aim to cover all possible situations of cointegration in models with I(2) variables but to show the robustness of the aggregation theory for some relevant examples. First, let us assume the following disaggregated model:

$$\Delta^2 \mathbf{x}_{2,t} = \mathbf{u}_{3,t}, \tag{12}$$

$$\Delta \mathbf{x}_{1,t} = \mathbf{u}_{2,t}, \tag{13}$$

$$y_t - \mathbf{x}'_{1,t}\boldsymbol{\beta}_1 - \mathbf{x}'_{2,t}\boldsymbol{\beta}_2 = u_{1,t}, \tag{14}$$

where $\mathbf{u}_t = (u_{1,t}, \mathbf{u}'_{2,t}, \mathbf{u}'_{3,t})'$ satisfies the assumptions described in Section 2 and the long-run covariances are zero $\Omega_{12} = \Omega_{13} = \Omega_{23} = 0$. For this model the optimal estimator is the OLS estimator. Theorem 3.2 shows the asymptotic distribution of the optimal estimator of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ for the aggregation schemes $\mathbf{Z}_T^{(0)}$ (aggregation I), $\mathbf{Z}_T^{(0,1,1)}$ (aggregation II), $\mathbf{Z}_T^{(1,0,0)}$ (aggregation III), and $\mathbf{Z}_T^{(1)}$ (aggregation IV).

THEOREM 3.2. *The asymptotic distribution of the optimal regression-type estimator of $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2)'$ in model (12)–(14) with temporally aggregated time series is*

$$\text{Diag}(n, n^2)(\hat{\boldsymbol{\beta}}^i - \boldsymbol{\beta}) \Rightarrow (\Omega_{11} + \Lambda(m))^{1/2} \int dW_1 \mathbf{B}'_* \left(\int \mathbf{B}_* \mathbf{B}'_* \right)^{-1},$$

where $\Xi^I(m) = (\Omega_{11} + \Lambda(m))^{1/2}$, $\Xi^{II}(m) = m^{-1}(\Omega_{11} + \Lambda(m))^{1/2}$, $\Xi^{III}(m) = m\Omega_{11}^{1/2}$, $\Xi^{IV}(m) = \Omega_{11}^{1/2}$, $\mathbf{B}_* = (\mathbf{B}'_2, \overline{\mathbf{B}}'_3)'$, and $\overline{\mathbf{B}}_3(r) \equiv \int_0^r \mathbf{B}_3(s) ds$ with $r, s \in [0, 1]$.

Proof. See the Appendix.

From Theorem 3.2, temporal aggregation has the same effect on the separate distribution of the estimator of $\boldsymbol{\beta}_1$ as on the separate distribution of the estimator of $\boldsymbol{\beta}_2$, and this effect is exactly like that found for case CI(1, 1).

Now consider the estimation of CI(2, 1) relations, which we assume are generated by the following model:

$$\Delta^2 \mathbf{x}_{2,t} = \mathbf{u}_{3,t}, \tag{15}$$

$$\Delta x_{1,t} = \Delta \mathbf{x}'_{2,t} \boldsymbol{\beta}_3 + u_{2,t}, \tag{16}$$

where $\mathbf{u}_t = (u_{2,t}, \mathbf{u}'_{3,t})'$ satisfies the general assumptions for the cointegrated model with I(2) variables described in Section 2. In this situation, the optimal regression-type estimators of $\boldsymbol{\beta}_3$ are those estimators for the CI(1, 1) model with differenced variables. The limiting distribution for the different aggregation schemes is given in the next corollary.

COROLLARY 3.3. *The asymptotic distribution of the optimal regression-type estimator of $\boldsymbol{\beta}_3$ in model (15) and (16) with temporally aggregated time series is*

$$n(\hat{\boldsymbol{\beta}}^i_3 - \boldsymbol{\beta}_3) \Rightarrow \Omega_{22,3}^{1/2} \int dW_2 \mathbf{B}'_3 \left(\int \mathbf{B}_3 \mathbf{B}'_3 \right)^{-1},$$

where $\Xi^I(m) = \Omega_{22,3}^{1/2}$, $\Xi^{II}(m) = m^{-1}\Omega_{22,3}^{1/2}$, $\Xi^{III}(m) = m\Omega_{22,3}^{1/2}$, and $\Xi^{IV}(m) = \Omega_{22,3}^{1/2}$.

In this case, as a result of the absence of the aliasing effect because the errors are average and double-average sampled, we find slightly different results. Now, for pure aggregation schemes I and IV the limiting distribution remains invariant despite the sampling interval, whereas in the case of mixed sampling the dispersion is reduced with m for aggregation II, and it increases with m for aggregation III.

Now, when a practitioner has to aggregate a stock regressand, the best option is always to apply systematic sampling. This is slightly different from case

CI(1,1) where in some situations the best way to aggregate the regressand was average sampling when the regressor was a flow. When it is the stock regressor that needs to be aggregated, the practitioner has to apply average sampling in all cases, in keeping with case CI(1,1).

The limiting properties of the optimal estimator of β_1 are derived under the assumption that the long-run variance Ω must be consistently estimated, and there may be a loss of valuable information when this matrix is estimated with temporally aggregated time series. Consequently, the aggregation effects on the estimation of Ω may alter the preceding conclusions reached on the best aggregation option for the estimation of long-run relations in finite samples. However, the Monte Carlo evidence provided in Section 4 shows that the estimation of Ω does not make any significant difference to the order in which the types of aggregation are ranked.

3.2. Temporally Aggregated Observed Information

When the purpose of the analysis is to test a hypothesis on the cointegrating vectors, such as the PPP theory or permanent income hypothesis, the aggregation effect on the marginal distribution of the estimator provides us with half the story, because for nonstationary cointegrating regressions, the observed information $J_n(\beta_1) \equiv -\partial^2 \log L(\beta_1) / \partial \beta_1 \partial \beta_1'$ where $\log L(\beta_1)$ is the log likelihood function, and not the variance of $\hat{\beta}_1$ ($V(\hat{\beta}_1)$), is used as the normalizing matrix of the χ^2 -distributed Wald test (see Johansen, 1995):

$$W = (\hat{\beta}_1 - \beta_1)' J_n(\beta_1) (\hat{\beta}_1 - \beta_1)'$$

This situation is different from a stationary regression where $V(\hat{\beta}_1)^{-1} = J_n(\beta_1)$, and the marginal distribution of the estimator is sufficient to discuss the aggregation effect. However, the observed information in a cointegrated model is different from the inverse of the variance of the estimator, and the aggregation effect must be considered on both the estimator and the information. This task is straightforward when we only consider the asymptotic properties, because the χ^2 distribution of the Wald test implies that the aggregation effect on the observed information should neutralize the aggregation effect on the estimator. Thus, an increase in the dispersion of the asymptotic distribution of the estimator due to aggregation should be counterbalanced by a reduction in the dispersion of the distribution of the information. For example, for $\Lambda(m) = (m^2 - 1)\Omega_{11,2}$, the asymptotic dispersion of the observed information increases for cases I and III, whereas it is invariant for cases II and IV. Therefore, in asymptotic terms it is not relevant which type of aggregation we select to test restrictions on β_1 .

However, this asymptotic result may be different when we consider the more realistic finite-sample framework. In such a setting, as shown by Johansen (2002), the distribution of the test is slightly different from the χ^2 distribu-

tion, depending not only on the number of observations but also on the parameters of the model. As shown in Section 2, different combinations of aggregation schemes have different effects on the short-run dynamics, and therefore in finite samples one aggregation scheme may be clearly superior to another for hypothesis testing. Moreover, the faster convergence of the observed information $n^{-2}J_n^i(\beta_1)$ than the estimator $n(\hat{\beta}_1^i - \beta_1)$ may contribute to the fact that the aggregation effect on $J_n(\beta_1)$ has a bigger influence on the finite-sample properties of the test than the aggregation effect on $n(\hat{\beta}_1^i - \beta_1)$. However, the Monte Carlo experiment in Section 4 shows that, for the selected model, all the aggregation options have similar effects on the size of the test.

4. MONTE CARLO EXPERIMENT

4.1. Design of the Experiment

In this section we study the finite-sample aggregation effect on the FM-OLS.⁵ We do so using a very simple data generation process (DGP) used in many simulation studies (see, among others, Gonzalo, 1994):

$$\Delta x_t = u_{2,t},$$

$$y_t - x_t = u_{1,t}, \quad t = 1, \dots, n,$$

where $u_{1,t} = \rho u_{1,t-1} + \varepsilon_{1,t}$, $u_{2,t} = \varepsilon_{2,t}$, and

$$(\varepsilon_{1,t}, \varepsilon_{2,t})' \sim \text{i.i.d.} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right].$$

The length of the sampling interval was fixed as $m = 3$. The relevant magnitudes for the aggregation effect are the aliasing component, $\Lambda(3)$, and the conditional long-run variance of the cointegrating errors, $8\Omega_{11.2}$. This very simple model allows us to control these magnitudes through parameter ρ . More specifically, we consider three sizes for the aliasing effect, so that for $\rho = -0.5$ the aliasing component is very big in comparison with $8\Omega_{11.2}$ ($\Lambda(3) = 4.54$ and $8\Omega_{11.2} = 2.64$); for $\rho = 0$ the aliasing component is slightly smaller than $8\Omega_{11.2}$ ($\Lambda(3) = 4$ and $8\Omega_{11.2} = 6$), and for $\rho = 0.5$ the aliasing component is much bigger than $8\Omega_{11.2}$ ($\Lambda(3) = 2.28$ and $8\Omega_{11.2} = 24$).

The experiment is performed with two spans, $n = 120$ and $n = 360$. We only present the results for the shorter span, $n = 120$, to compare the different aggregation effects on the estimator and test (see Table 2) because a comparison of the aggregation schemes is not (qualitatively) affected by the size of the span. As a measure of the aggregation effect on the precision of the estimator, we provide the ratio of the mean square error (MSE) of the aggregated estimator in relation to the disaggregated estimator, denoted $M(i)$ where $i \in \{I, II, III, IV\}$ is an aggregation scheme. As a measure of the aggregation effect on the test,

we compute the variation of the size of the test for $H_0: \beta = 1$ based on the 5% level, denoted $S(i)$ with $i \in \{I, II, III, IV\}$. That is, $S(i)$ stands for the difference between the size of the test in the aggregated models and the size in the disaggregated model. Three thousand replications of the experiment are used to obtain Monte Carlo estimations of the MSE and 10,000 for the size of the test. We also compare the aggregation effect with the span effect. To do so, with the disaggregated model we compute the variation in the MSE ($M(n)$) and size ($S(n)$) for a small sample $n = 120$ and a big sample $n = 360$. To avoid the influence of initial conditions, a constant is estimated together with the cointegrating coefficient and the first 50 observations are discarded.⁶

4.2. Results

Table 2 shows the variation in the MSE and the size of the test when different types of aggregation are applied to the DGP described previously for $n = 120$. When $8\Omega_{11.2} > \Lambda(3)$, the best aggregation scheme, in terms of the precision of the estimator, is aggregation type II, with a lower MSE than the disaggregated estimator. The second best aggregation option is aggregation type IV. With this data transformation, the MSE hardly varies in comparison with the disaggregated estimator. Much less precise estimations of the cointegrated vector are obtained when the regressor is systematically sampled, with aggregation III being the worst option and there being a tremendous relative increase in the dispersion of the estimator. The performance of the estimator, when both variables are systematically sampled ($i = I$), is largely dependent on the relative size of the aliasing component, with the result that when the aliasing component is smaller than the long-run conditional variance, the precision of the estimator is closer to the precision of the estimator with aggregation IV.

If we look at the aggregation effect on the test, we find a different story. In this case, all the aggregation schemes increase the distortion of the test by approximately 3 to 5 points. The main difference, when compared with the aggregation effect on the precision of the estimator, is that now the different types of aggregation all share a very similar effect. The few differences that can be

TABLE 2. Comparison of the different temporal aggregation effects

ρ	$M(I)$	$S(I)$	$M(II)$	$S(II)$	$M(III)$	$S(III)$	$M(IV)$	$S(IV)$
-0.5	9.8	4.9	1.4	4.9	22.9	4.4	1.1	3.5
0	4.1	4.8	0.6	4.8	14.2	4.9	1.2	4.9
0.5	1.6	4.5	0.2	4.5	11.3	3.1	1.3	4.0

Note: See Table 1 for the definition of the aggregation schemes $i = \{I, II, III, IV\}$. Here $n = 120$, $M(i)$ stands for the relative MSE of the aggregated cointegration estimator in relation to the disaggregated one, and $S(i)$ stands for the variation in the size of the test of the hypothesis $H_0: \beta = 1$ in the aggregated and the disaggregated models.

observed reveal that the aggregation effect on the test cannot be explained by the aggregation effect on the estimator. The most significant example is the case $\rho = 0.5$, where the least distorted test is obtained with aggregation III, the worst aggregation option in terms of the MSE. However, we cannot infer from this that the worst aggregation for estimation purposes is the best one for testing, because if we look at case III for $\rho = -0.5$, where there is the biggest loss in estimator precision (22.9), it is not the one with the least size distortion.

It is interesting to compare the aggregation effect with the effect of the span on the estimator and the test. In Table 3 we present the variation in the MSE and size when the sample is reduced from 360 to 120 observations, either by reducing the span by 3 ($M(n)$ and $S(n)$), or by temporal aggregation ($M(i)$ and $S(i)$, $i \in \{I, II, III, IV\}$). If we examine the effects on the precision of the estimators, the reduction of the span has similar repercussions to the worst aggregation schemes (I, III), a situation similar to that one expects to find with a stationary regression. However, the reduction of the span clearly has worse consequences than aggregation schemes II and IV, which shows that the aggregation effect on the estimation of cointegrating regressions is not as important as the aggregation effect on stationary regressions.

Focusing on the test, reducing the span has a slightly bigger impact on size than aggregation. However, once again the few exceptions that can be observed show how the effects on the test are different than on the estimator. For example, when $\rho = 0$, the size distortion of aggregation II is practically the same (2.2) as the distortion caused by the reduction of the span (2.1). However, for this case, if we consider the effect on the precision of the estimator, the reduction of the span affects the quality of the estimation far more than aggregation with scheme II.

To summarize the Monte Carlo results, different aggregation schemes have very different effects on the precision of the estimator, thus confirming the asymptotic theory derived in the paper. More specifically, when the regressor is average sampled, the precision is not affected or even improved by temporal aggregation, whereas when the regressor is systematically sampled the estimator is clearly less precise, especially when the regressand is average sampled.

TABLE 3. Comparison between the span effect and the temporal aggregation effect

ρ	$M(n)$	$S(n)$	$M(I)$	$S(I)$	$M(II)$	$S(II)$	$M(III)$	$S(III)$	$M(IV)$	$S(IV)$
-0.5	9.8	2.3	9.6	1.4	1.2	1.4	21.6	1.3	1.0	0.8
0	9.8	2.1	3.4	1.4	0.6	2.2	15.3	1.5	1.0	1.7
0.5	9.6	2.4	1.5	1.1	0.2	1.5	10.2	1.7	1.1	0.6

Note: See Table 2 for the definition of $M(i)$ and $S(i)$ $i = \{I, II, III, IV\}$. Here $n = 360$, $M(n)$ stands for the relative MSE of the cointegration estimator with $n = 120$ in relation to the estimator for $n = 360$, and $S(n)$ is the variation in the size of the test of the hypothesis $H_0: \beta = 1$ between $n = 120$ and $n = 360$.

The aggregation effects on the test are very different from the effects on the estimator. In this case, there are no significant differences among the different aggregation options. Finally, we found that reducing the span has a similar effect on precision to those aggregation schemes where the regressor is systematically sampled (I, III) and a much greater effect than those aggregations where the regressor is average sampled (II, IV). Also, a reduction in the span has a slightly greater effect on size than the temporal aggregation of the variables.

5. CONCLUSIONS

We have discussed the effects of different types of discrete time aggregation on the regression-type optimal inference on cointegrating vectors and have shown how different types of aggregation have different effects on the limiting properties and finite-sample properties of optimal estimators but similar effects on the hypothesis test. The theoretical aggregation results have potential empirical implications, because in many situations, a practitioner decides how to temporally aggregate certain variables before making inferences on the cointegrating vectors, because stock variables such as exchange rates, interest rates, or the money supply can be temporally aggregated either by systematic sampling or by average sampling. As for the practical implications of our study, if a stock regressand must be temporally aggregated, because some of the regressors are not available at the finest sampling interval, then when $\Lambda(m) < (m^2 - 1)\Omega_{11.2}$ the best aggregation scheme is systematic sampling whereas otherwise average sampling is the best option in terms of estimation precision. When a practitioner has to choose an aggregation option for the regressor, the best choice is always to apply average sampling. These results only partially corroborate those of Chambers (2003), where the main recommendation was always to use average sampling. From our conclusions it seems that, as long as the aliasing effect is not very important, it is a very bad idea to apply average sampling to a regressand when the regressor is a stock, because this combination of aggregations leads to the highest increase in the noise-to-signal ratio. These differences are discussed theoretically and illustrated with a Monte Carlo study where aggregation type II outperforms aggregation type IV for the situation predicted by the discrete time aggregation theory.

It is a very different story when a practitioner plans to test a long-term theory, rather than just estimating a long-run relation. In this case the different aggregation options have a very similar effect on the test, and none of them, whatever the situation, leads to a least size-distorted test.

We have also compared the effect of the span with the effect of temporal aggregation, and the main conclusion is that the span has a greater effect on inference than temporal aggregation. To be more precise, the span has a much greater impact on the estimator than the best aggregation schemes, II and IV, and a slightly bigger effect on the test than any of the aggregation schemes.

As a further research issue it would be interesting to analyze which aggregation option is best suited for other purposes such as prediction, because in this case the relevant aggregation effects are not those on the cointegrating vector but on the adjusting vector (see Johansen, 1994).

NOTES

1. The invariance of cointegrating vectors with the sampling interval holds whenever the seasonal unit roots do not alias into the zero frequency (see Granger and Siklos, 1995).

Chambers (2003) shows that when both the regressand and the regressor are temporally aggregated by average sampling there is no loss of asymptotic efficiency for the estimation of the cointegrating vectors.

2. A recently developed approach, the random time aggregation (see Jordà, 1999), allows the dynamics to be generated at random time intervals.

3. Systematic sampling is typically applied to stock variables such as prices or population, and average sampling is applied to flow variables such as consumption or income and to stock variables also.

4. Further research needs to be done for the case $k_0 > 1$.

5. We use the FM-OLS routine available at the COINT Gauss library, programmed by Sam Ouliaris and Peter Phillips. More specifically, the long-run variance is estimated with the Parzen spectral window, AR(1) prewhitening, and automatic bandwidth selection. We also considered the dynamic ordinary least squares estimator, which led to qualitatively similar results that are not reported here to save space.

6. Additionally we considered many other values for the parameters, i.e., different correlations between the errors ($\delta = 0$ and $\delta = -0.5$ instead of $\delta = 0.5$) and different values for the variance of $u_{2,t}$ ($\sigma^2 = 2$). However, the aggregation effect does not depend on the value of these parameters, and therefore we do not present all these results here. Tables with the additional results are available upon request.

REFERENCES

- Chambers, M.J. (2003) The asymptotic efficiency of cointegration estimators under temporal aggregation. *Econometric Theory* 19, 49–77.
- Comte, F. (1999) Discrete and continuous time cointegration. *Journal of Econometrics* 88, 207–226.
- Ericsson, N.R., D. Hendry, & H.-A. Tran (1994) Cointegration, seasonality, encompassing, and the demand for money in the UK. In C. Hargreaves (ed.), *Nonstationary Time Series Analysis*, pp. 179–224. Oxford University Press.
- Gonzalo, J. (1994) Five alternative methods of estimating long-run equilibrium relationships. *Journal of Econometrics* 60, 203–233.
- Granger, C.W.J. (1990) Aggregation of time-series variables: A survey. In T. Barker and M.H. Pesaran (eds.), *Disaggregation in Econometric Modelling*, pp. 17–34. Routledge.
- Granger, C.W.J. & P.L. Siklos (1995) Systematic sampling, temporal aggregation, seasonal adjustment, and cointegration: Theory and evidence. *Journal of Econometrics* 66, 357–369.
- Haldrup, N. (1994) The asymptotics of single-equation cointegration regressions with I(1) and I(2) variables. *Journal of Econometrics* 63, 153–181.
- Hargreaves, C. (1994) A review of methods of estimating cointegrating relationships. In C. Hargreaves (ed.), *Nonstationary Time Series Analysis*, pp. 87–131. Oxford University Press.
- Johansen, S. (1991) Estimation and hypothesis testing of cointegrating vectors in gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Johansen, S. (1994) Estimating systems of trending variables. *Econometric Reviews* 13, 351–386.

- Johansen, S. (1995) The role of ancillarity in inference for non-stationary variables. *Economic Journal* 13, 302–320.
- Johansen, S. (2002) A small sample correction for tests of hypothesis on the cointegrating vectors. *Journal of Econometrics* 111, 195–221.
- Johansen, S. & K. Juselius (1992) Testing structural hypothesis in a multivariate cointegration analysis of the PPP and UIP for UK. *Journal of Econometrics* 53, 211–244.
- Jordà, O. (1999) Random-time aggregation in partial adjustment models. *Journal of Business & Economic Statistics* 17, 382–395.
- Kitamura, Y. (1995) Estimation of cointegrated systems with I(2) processes. *Econometric Theory* 11, 1–24.
- Marcellino, M. (1999) Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics* 17, 129–136.
- Niemi, H. (1984) The invertibility of sampled and aggregated ARMA models. *Metrika* 31, 43–50.
- Park, J.Y. (1992) Canonical cointegrating regressions. *Econometrica* 60, 119–143.
- Pesaran, M.H. & Y. Shin (1996) Cointegration and speed of convergence to equilibrium. *Journal of Econometrics* 71, 117–143.
- Phillips, P.C.B. (1991a) Error correction and long-run equilibrium in continuous time. *Econometrica* 59, 967–980.
- Phillips, P.C.B. (1991b) Optimal inference in cointegrated systems. *Econometrica* 59, 282–306.
- Phillips, P.C.B. (1991c) Spectral regression for cointegrated time series. In W.A. Barnett, J. Powell, & G.E. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 413–436. Cambridge University Press.
- Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.
- Saikkonen, P. (1991) Asymptotically efficient estimation of cointegrating regressions. *Econometric Theory* 7, 1–21.
- Stock, J.H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.
- Stram, D.O. & W.W.S. Wei (1986) Temporal aggregation in the ARIMA process. *Journal of Time Series Analysis* 7, 279–292.
- Telser, L.G. (1967) Discrete samples and moving sums in stationary stochastic processes. *Journal of the American Statistical Association* 62, 484–499.
- Wei, W.S.W. (1989) *Time Series Analysis: Univariate and Multivariate Methods*. Addison-Wesley.

APPENDIX: Proofs

Proof of Lemma 2.1.

(a) *Temporally Aggregated Triangular Representation.* Let us consider the representation of the systematically sampled process $\mathbf{Z}_T^{(0)}$. To apply systematic sampling to equations (4)–(6), we need to express the model in such a way that the variables are observable every m t -period. This is the case of all the variables in equation (6) but not in (4) or (5) because of the presence of differenced ($\Delta \mathbf{x}_{1,t}$, $\Delta \mathbf{x}_{2,t}$) and double-differenced variables $\Delta^2 \mathbf{x}_{2,t}$. Because $\Delta_m = \Delta s_m(L)$, we multiply both sides of (5) by the summation filter and both sides of (4) by $[s_m(L)]^2$ to get the following representation:

$$\Delta_m^2 \mathbf{x}_{2,t} = [s_m(L)]^2 \mathbf{u}_{3,t}, \quad (\text{A.1})$$

$$\Delta_m \mathbf{x}_{1,t} - \Delta_m \mathbf{x}'_{2,t} \boldsymbol{\beta}_3 = s_m(L) \mathbf{u}_{2,t}, \quad (\text{A.2})$$

$$\mathbf{y}_t - \mathbf{x}'_{1,t} \boldsymbol{\beta}_1 - \mathbf{x}'_{2,t} \boldsymbol{\beta}_2 = \mathbf{u}_{1,t}. \quad (\text{A.3})$$

Now this representation of the disaggregated cointegrated model is observable at the longer sampling interval $m\Delta t$, and so we just apply systematic sampling to (A.1)–(A.3), thus obtaining the representation of $\mathbf{Z}_T^{(0)}$:

$$\Delta_m^2 \mathbf{X}_{2,T}^{(0)} = \mathbf{U}_{3,T}^{(2)},$$

$$\Delta_m \mathbf{X}_{1,T}^{(0)} - \Delta_m \mathbf{X}_{2,T}^{(0)'} \boldsymbol{\beta}_3 = \mathbf{U}_{2,T}^{(1)},$$

$$\mathbf{Y}_T^{(0)} - \mathbf{X}_{1,T}^{(0)'} \boldsymbol{\beta}_1 - \mathbf{X}_{2,T}^{(0)'} \boldsymbol{\beta}_2 = \mathbf{U}_{1,T}^{(0)},$$

where $\mathbf{U}_{3,T}^{(2)} = [s_m(L)]^2 \mathbf{u}_{3,mT}$, $\mathbf{U}_{2,T}^{(1)} = s_m(L) \mathbf{u}_{2,mT}$, and $\mathbf{U}_{1,T}^{(0)} = \mathbf{u}_{1,mT}$.

To obtain the representation of the pure average sampled process $\mathbf{Z}_T^{(1)}$, we just need to apply the summation polynomial to equations (A.1)–(A.3) and then apply systematic sampling. The aggregated model for mixed sampling is obtained by similar transformations to the two preceding cases. Let us focus on equation (A.3) and consider the cases $\mathbf{Z}_T^{(0,1,0)}$ and $\mathbf{Z}_T^{(1,1,0)}$. For the first case, we multiply $\mathbf{x}_{1,t}$ in (A.3) by $s_m(L)^{-1} s_m(L)$:

$$\mathbf{y}_t - s_m(L) \mathbf{x}'_{1,t} \boldsymbol{\beta}_1 s_m(L)^{-1} - \mathbf{x}'_{2,t} \boldsymbol{\beta}_2 = \mathbf{u}_{1,t}$$

and then apply systematic sampling, thereby obtaining

$$\mathbf{Y}_T^{(0)} - m^{-1} \mathbf{X}_{1,T}^{(1)'} \boldsymbol{\beta}_1 - \mathbf{X}_{2,T}^{(0)'} \boldsymbol{\beta}_2 = \mathbf{U}_{1,T}^{(0)}.$$

For the case $\mathbf{Z}_T^{(1,1,0)}$, we apply the summation polynomial to both sides of (A.3):

$$s_m(L) \mathbf{y}_t - s_m(L) \mathbf{x}'_{1,t} \boldsymbol{\beta}_1 - s_m(L) \mathbf{x}'_{2,t} \boldsymbol{\beta}_2 = s_m(L) \mathbf{u}_{1,t}$$

and then, through sampling, obtain

$$\mathbf{Y}_T^{(1)} - \mathbf{X}_{1,T}^{(1)'} \boldsymbol{\beta}_1 - m \mathbf{X}_{2,T}^{(0)'} \boldsymbol{\beta}_2 = \mathbf{U}_{1,T}^{(1)}.$$

The remaining cases are obtained in a similar way.

(b) Temporally Aggregated Multivariate Invariance Principle. Given the fact that m is finite, the application of any of the temporal aggregation schemes as a linear filter does not alter the multivariate invariance principle, but rather the variance of the Brownian motion, the long-run variance of \mathbf{u}_t . Consequently, we derive the aggregation effects on the long-run variance, and for this purpose we determine the effects of the different aggregation filters on the covariance function.

Let $\Gamma(k) \equiv E(\mathbf{u}_t \mathbf{u}'_{t-k})$ denote the covariance function of the stationary process \mathbf{u}_t . The effect of systematic sampling on the covariance function is

$$\Gamma^{(0)}(k) \equiv E(\mathbf{U}_T^{(0)} \mathbf{U}_{T-k}^{(0)'}) = E(\mathbf{u}_{mT} \mathbf{u}'_{mT-mk}) = E(\mathbf{u}_t \mathbf{u}'_{t-mk}) = \Gamma(mk). \tag{A.4}$$

Any combination of aggregation schemes implies the application of a specific lag polynomial to the covariance function before applying systematic sampling. Thus, the effect of the lag operator on the displacement k of the covariance function (see other examples from literature on temporal aggregation such as Telser, 1967; Stram and Wei, 1986) is

$$E((L^j \mathbf{u}_t) \mathbf{u}'_{t-k}) = E(\mathbf{u}_{t-j} \mathbf{u}'_{t-k}) = E(\mathbf{u}_t \mathbf{u}'_{t-k+j}) = \Gamma(k - j), \tag{A.5}$$

$$E(\mathbf{u}_t (L^j \mathbf{u}'_{t-k})) = E(\mathbf{u}_t \mathbf{u}'_{t-k-j}) = \Gamma(k + j). \tag{A.6}$$

From (A.4)–(A.6) it is straightforward to determine the effect of the different combinations of aggregations on the covariance matrix function. All possible combinations are given by

$$\begin{aligned} \Gamma^{(0)}(k) &= \Gamma(mk); & \Gamma^{(0,1)}(k) &= \sum_{g=0}^{m-1} \Gamma(mk + g); \\ \Gamma^{(0,2)}(k) &= \sum_{g=0}^{m-1} \sum_{h=0}^{m-1} \Gamma(mk + g + h), \\ \Gamma^{(0,3)}(k) &= \sum_{g=0}^{m-1} \sum_{h=0}^{m-1} \sum_{i=0}^{m-1} \Gamma(mk + g + h + i); \\ \Gamma^{(1)}(k) &= \sum_{g=0}^{m-1} \sum_{h=0}^{m-1} \Gamma(mk - g + h), \\ \Gamma^{(1,2)}(k) &= \sum_{g=0}^{m-1} \sum_{h=0}^{m-1} \sum_{i=0}^{m-1} \Gamma(mk - g + h + i); \\ \Gamma^{(1,3)}(k) &= \sum_{g=0}^{m-1} \sum_{h=0}^{m-1} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \Gamma(mk - g + h + i + j), \\ \Gamma^{(2)}(k) &= \sum_{g=0}^{m-1} \sum_{h=0}^{m-1} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \Gamma(mk - g - h + i + j); \\ \Gamma^{(2,3)}(k) &= \sum_{g=0}^{m-1} \sum_{h=0}^{m-1} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} \Gamma(mk - g - h + i + j + k), \\ \Gamma^{(3)}(k) &= \sum_{g=0}^{m-1} \sum_{h=0}^{m-1} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} \Gamma(mk - g - h - i + j + k + l). \end{aligned}$$

Thus, the temporally aggregated long-run variances are obtained from the definition of long-run variance:

$$\Omega^{(l,j)} = \sum_{i=-\infty}^{\infty} \Gamma^{(l,j)}(k)$$

in such a way that, with the exception of $\Omega^{(0)}$, they are

$$\Omega^{(l,j)} = m^{l+j-1} \Omega.$$

For the systematically sampled long-run variance $\Omega^{(0)}$, we use the result reached by Niemi (1984), who obtains $F^{(0)}(\omega)$, the spectrum of a systematically sampled process, and the fact that $\Omega^{(0)} = 2\pi F^{(0)}(0)$, getting

$$\Omega^{(0)} = m^{-1} (\Omega + 2\pi \sum_{j=1}^{m-1} F(2\pi j/m)). \quad \blacksquare$$

Proof of Theorem 3.1. To obtain the asymptotic distribution of the different combinations of aggregations, we replace the expressions for the temporally aggregated long-run variances presented in Lemma 2.1 in the formulation of the mixed normal distribution of Theorem 1 of Phillips (1991b, p. 299), so that, after some manipulation of the sampling interval, we get

$$\begin{aligned} \frac{n}{m}(\hat{\beta}_1^I - \beta_1) &\Rightarrow m^{-1}(\Omega_{11.2} + \Lambda(m))^{1/2} \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1}, \\ \frac{n}{m}(\hat{\beta}_1^{II} - \beta_1) &\Rightarrow m^{-2}(\Omega_{11.2} + \Lambda(m))^{1/2} \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1}, \\ \frac{n}{m}(\hat{\beta}_1^{III} - \beta_1) &\Rightarrow \Omega_{11.2}^{1/2} \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1}, \\ \frac{n}{m}(\hat{\beta}_1^{IV} - \beta_1) &\Rightarrow m^{-1} \Omega_{11.2}^{1/2} \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1}. \end{aligned}$$

These expressions cannot be compared with the disaggregated mixed normal distribution because they are normalized by a different sample size. So, to get comparable distributions, we must obtain the expression for the distribution where the span normalizes the estimator bias:

$$\begin{aligned} n(\hat{\beta}_1^I - \beta_1) &\Rightarrow (\Omega_{11.2} + \Lambda(m))^{1/2} \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1}, \\ n(\hat{\beta}_1^{II} - \beta_1) &\Rightarrow m^{-1}(\Omega_{11.2} + \Lambda(m))^{1/2} \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1}, \\ n(\hat{\beta}_1^{III} - \beta_1) &\Rightarrow m \Omega_{11.2}^{1/2} \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1}, \\ n(\hat{\beta}_1^{IV} - \beta_1) &\Rightarrow \Omega_{11.2}^{1/2} \int dW_1 \mathbf{B}'_2 \left(\int \mathbf{B}_2 \mathbf{B}'_2 \right)^{-1}. \quad \blacksquare \end{aligned}$$

Proof of Theorem 3.2. The proof of this theorem is very similar to the preceding one, and therefore we have not presented the details. It is only necessary to replace the expressions for the different long-run variances cited in Lemma 2.2 in Haldrup’s Theorem 2 (Haldrup, 1994, p. 163). ■