Conditionally Efficient Estimation of Long-run Relationships Using Mixed-frequency Time Series

J. Isaac Miller[1,2]

Department of Economics, University of Missouri
118 Professional Building, Columbia, Missouri 65211, USA

Abstract

I analyze efficient estimation of a cointegrating vector when the regressand is observed at a lower frequency than the regressors. Previous authors have examined the effects of specific temporal aggregation or sampling schemes, finding conventionally efficient techniques to be efficient only when both the regressand and the regressors are average sampled. Using an alternative method for analyzing aggregation under more general weighting schemes, I derive an efficiency bound that is conditional on the type of aggregation used on the regressand and differs from the unconditional bound defined by the infeasible full-information high-frequency data-generating process. I modify a conventional estimator, canonical cointegrating regression (CCR), to accommodate cases in which the aggregation weights are either unknown or known. In the unknown case, the correlation structure of the error term generally confounds identification of the conditionally efficient weights. In the known case, the correlation structure may be utilized to offset the potential information loss from aggregation, resulting in a conditionally efficient estimator. Efficiency is illustrated using a simulation study and an application to estimating a gasoline demand equation.

First draft: January 27, 2011
Current draft: August 20, 2012

---

[1]Correspondence: millerjisaac@missouri.edu.

1 Introduction

Many models of practical importance to econometricians rely on time series that are observed at different frequencies. Some examples of economic models where mixed-frequency series are unavoidable in estimation include purchasing power parity, where nominal exchange rates fluctuate daily while aggregate prices are sampled only monthly at best; demand models, where the price of a good may be observed daily but the quantity purchased is usually observed only monthly; or just about any macro-level model, which commonly include series that are measured at a mix of quarterly and monthly frequencies.

When the genesis of the problem lies upstream, the analyst must necessarily make a judgment about the best way to incorporate mixed-frequency data into the model at hand. The most commonly used methods for handling mixed frequencies include aggregating or sampling from the higher frequency series to create a low-frequency series.[3] Aggregation generally results in a loss of information, and thus a loss of efficiency.

Cointegrated series, having common long-run relationships and sharing common stochastic trends, offer unique opportunities for overcoming the challenges of inefficiency.[4] Specifically, the superconsistency of the least squares estimator extends to mixed-frequency models, so that an initial consistent estimate may be leveraged to create second-step estimators with asymptotically mixed normal distributions. Chambers (2003) and Pons and Sansó (2005) studied the effects of temporal aggregation of cointegrating regressions on efficiency. Although they explicitly studied models with temporally aggregated regressand *and* regressors, their results shed light on models with *mixed* frequencies considered here, where the lower frequency may be the result of aggregation.

Defining the data-generating process (DGP) in continuous time, Chambers (2003) found that the most efficient estimator is obtained when both regressand and regressors are measured as flows rather than stocks. Such estimators are just as efficient as if the data were continuously recorded. In the practical case in which the type of aggregated data – stock or flow – is dictated by the model, Chambers (2003) suggested mimicking flow data by average sampling, even if the data are not flows.

Pons and Sansó (2005) built on Chambers's (2003) research by taking a discrete-time rather than continuous-time approach. These authors argued that a discrete-time approach may be preferable when agents' reactions to deviations from long-run equilibria are not instantaneous due to transaction costs. Measurement error provides another motivation for

---

[3] *Average* or *flat* sampling, assigning equal weights to each high-frequency observation, is a common but specific type of temporal aggregation. However, the weights used in aggregation may be more complicated. *Selective* or *skip* sampling, which sets most weights to zero, is another common weighting scheme. Specifically, *end-of-period* sampling sets all but the last high-frequency weight in each low-frequency period to be zero. I use temporal *aggregation* and *sampling* almost interchangeably throughout the paper, in spite of a shade of difference in connotation. Although the term *aggregation* ceases to be accurate when only one weight is non-zero, it generally and accurately describes all other weighting schemes.

[4] Granger (1990), Granger and Siklos (1995), Marcellino (1999), Haug (2002), Pons and Sansó (2005), and Rajaguru and Abeysinghe (2008) studied various effects of temporal aggregation on estimation and testing of discrete-time cointegration models. Moreover, Chambers (2003, 2010) and Chambers and McCrorie (2007) studied temporally aggregated cointegrated series in continuous time. Miller (2010) studied the effects of imputation on cointegrated series as a special case of general nonlinear perturbations to the error of a cointegrating regression.

a courser partition of time to detect a long-run relationship. Macroeconomic aggregates, such as consumer price indices, may be measured over a wide cross-section and unevenly over time. Defining the DGP at a monthly frequency circumvents this type measurement error in this case. On the other hand, a continuous-time framework dominates a discrete-time approach when the dynamics cannot be defined at the specified frequency but can be assumed to evolve continuously.

I take a discrete-time approach, which may be viewed as a complement of rather than as a substitute for a continuous-time approach in the following sense. Continuous time may be viewed as the "high frequency" (using discrete-time terminology), while the low frequency is in discrete time. The invariance principle based on a high-frequency DGP considered in this paper may be viewed as the low-frequency invariance principle derived by Chambers (2003, Lemma 2). The key difference lies in the long-run variances. Theorem 1 of Chambers (2003) presents long-run variances when the regressors and regressand are combinations of skip-sampled stock data and average-sampled flow data. If the high-frequency DGP is defined at this discrete frequency, then the benchmark high-frequency long-run variance, by which (unconditional) efficiency is evaluated in this paper and denoted by Ω below, may depend on the type of series, similarly to those in Chambers's (2003) theorem.

Several econometric methods in the literature incorporate different frequencies directly into estimation, thus avoiding potential inefficiency from aggregation. The principal advantage of such models lies in exploiting the high-frequency information without directly imputing the lower frequency series. One such approach is a distributed lag (DL) model with a low-frequency regressand and lagged high-frequency regressors. The DL approach has several advantages over extant mixed-frequency approaches. No numerical optimization or overly restrictive distributional assumptions are required.[5,6]

Like previous authors, I focus on efficient estimation of the cointegrating vector of a single-equation regression model. More specifically, I focus on the case of a low-frequency regressand with high-frequency regressors. Many applications, such as estimating price elasticities or inventory smoothing models, have this feature. As discussed below, the opposite case turns out to be analytically similar. I define the data-generating process at the higher frequency, so that the lower frequency of the regressand is the result of some sort of temporal aggregation or sampling scheme. Specifically, I refer to the scheme as the *regressand aggregation scheme*.

I present a low-frequency invariance principle more general than that of Pons and Sansó (2005) in the sense that it nests both of the two aggregation schemes considered by those authors, as well as more general aggregation schemes, such as the nonlinear DL schemes considered by Ghysels *et al.* (2004, 2006), *inter alia.*

---

[5]Alternatives include state-space models and nonlinear DL models. State-space models specifically for cointegrated mixed-frequency series have been studied by Gomez and Maravall (1994) and Seong *et al.* (2007). Nonlinear DL models have been employed by Ghysels *et al.* (2004, 2006), and a number of subsequent authors. My approach is much less parsimonious than nonlinear DL models, and may not be feasible when the number of high-frequency regressors approaches or exceeds the low-frequency sample size. Therefore, the approach in this paper is not recommended when parameter proliferation is a concern.

[6]However, Miller (2011) recently showed the asymptotic validity of forecasts from stationary or cointegrating nonlinear DL regressions with general error structures.

I take an approach that may be described as a semiparametric DL approach, employing the semiparametric approach of Park (1992) to construct a canonical cointegrating regression (CCR).[7] Although I use a DL approach for the high-frequency regressors, I do not extend the DL approach to account for error correlations, as Saikkonen (1991) does. Rather, CCR allows for a very general error correlation structure. This allowance is particularly convenient in temporal aggregation or mixed-frequency settings, in which the error may contain neglected information from the aggregation.

I derive two efficiency bounds: *aggregation-unconditional* and *aggregation-conditional*. All of the CCR estimators considered here are consistent and asymptotically mixed normal. Thus, the aggregation-unconditional bound is defined by the conditional asymptotic variance of the CCR estimator of the infeasible full-information high-frequency model. I show that this efficiency bound is obtained when both regressand and regressors are average-sampled, which supports the results of Chambers (2003) and Pons and Sansó (2005). Since the econometrician may be presented with series that have already been aggregated, I define the concept of an aggregation-conditional efficiency bound. In general, matching the regressor aggregation scheme with that of the regressand does *not* lead to conditionally efficient estimators.

Within the DL framework, I consider both unknown and known regressand aggregation weights. The cointegrating vector can always be estimated consistently using least squares, but unknown aggregation weights add a stationary term comprised of differenced regressors with coefficients as a function of these weights. Unless the regressors are strictly exogenous, the coefficients cannot be estimated consistently, efficient weights cannot be identified, and so the idiosyncratic error term is compounded with estimation error. The conventional CCR is still asymptotically mixed normal and thus may be used for valid inference. Although not generally efficient, this approach is preferable to arbitrarily aggregating all of the high-frequency series, as is often done in the empirical literature.

The third contribution of this research lies in improving efficiency in the case of *known* weights. Known weights provide additional structure, which may be used to create conditionally efficient estimators. A key insight is that although correlation of the error with the regressors is traditionally viewed as a nuisance in econometric modeling, this correlation may be leveraged in the mixed-frequency case to regain information lost from aggregating the regressand. I show how this correlation may be utilized for conditionally efficient estimation of the cointegrating vector.

The remainder of the paper is organized as follows. In Section 2, I describe the high-frequency DGP and mixed-frequency model, and I derive some preliminary results that generalize some of the results in the literature to more general aggregation schemes. The main results are contained in Section 3, in which I derive efficiency bounds based on the limiting variances, I present asymptotic results for the cases of unknown and known aggregation weights, and I introduce modifications of CCR to address these cases. I demonstrate the efficiency gain for the latter case using finite-sample simulations in Section 4. In Section

---

[7]There are of course many other well-known approaches to estimating cointegrating relationships, including those by Engle and Granger (1987), Johansen (1988), Phillips and Hansen (1990), Saikkonen (1991), Stock and Watson (1993), and Pesaran and Shin (1998). The extension of my results to the FM-OLS estimator of Phillips and Hansen (1990) should be straightforward.

5, I examine an application to estimating short-run price and income elasticities of gasoline demand, providing evidence in favor of estimating the weights when they are unknown. Section 6 concludes, and an appendix contains mathematical proofs.

2   Model and Preliminary Results

I define the DGP and the unconditional efficiency bound of an estimator of the cointegrating vector at the higher frequency. Since the mixed-frequency model is estimated at the lower frequency, I analyze asymptotic approximations at that frequency. The preliminary results of this section generalize some of the results in the existing literature on aggregation by allowing for more flexible aggregation schemes or mixed frequencies.

2.1   High-frequency DGP and Invariance Principle

Consider a high-frequency DGP given by

$$y_{t-i/m} = \beta' x_{t-i/m} + u_{t-i/m}. \tag{1}$$

I normalize the increments of the *lower* frequency to unity so that increments of the higher frequency are $1/m$.[8] Let the number of high-frequency observations be denoted by M and the number of low-frequency observations be denoted by T. It is convenient to let $T = [M/m]$, where $[a]$ denotes the greatest integer not exceeding a. If the high-frequency observations occur in multiples of m, then clearly $T = M/m$. I assume that $m < \infty$, or that the ratio of frequencies is finite. As a result, approximations using standard asymptotics may be more appropriate than those using in-fill asymptotics.

   To analyze statistical properties of the estimators, define $b_{t-i/m} \equiv (u_{t-i/m}, \triangle^{(1/m)} x'_{t-i/m})'$, where $\triangle^{(1/m)}$ is the high-frequency difference operator. Except for the increment normalization above, I maintain standard assumptions about stationarity of $(b_{t-i/m})$, similar to those of Phillips and Hansen (1990), Saikkonen (1991), Park (1992), *inter alia*.

   For the sake of concreteness, I assume that

[A1] $(b_{t-i/m})$ is stationary and ergodic with zero mean, and

[A2] $M^{-1/2} \sum_{i=1}^{[Mr]} b_{i/m} \to_d B^{(1/m)}(r)$, a vector Brownian motion with covariance matrix Ω,

along the lines of Park (1992). The asymptotic approximations below may hold under much more general assumptions, such as those for the weakly dependent heterogeneous processes of Davidson (1994).

---

[8]Such a normalization is similar to Ghysels *et al.* (2004), e.g., but contrasts with that of Pons and Sansó (2005), e.g., who set increments of the higher frequency to unity but those of the lower frequency to m.

Lemma 1. *Under Assumptions [A1]-[A2],*

$$(mT)^{-1/2} \sum_{t=1}^{[Tr]} \sum_{i=0}^{m-1} b_{t-i/m} \to_d B^{(1/m)}(r), \tag{2}$$

as $T \to \infty$,

This invariance principle stems from the invariance principle in [A2]. However, the stochastic process in (2) is convenient for comparing variances of time series observed at different frequencies. Making use of the conventional partition,

$$\lim_{T \to \infty} var \left((mT)^{-1/2} \sum_{t=1}^{T} \sum_{i=0}^{m-1} b_{t-i/m} \right) \equiv \Omega = \begin{bmatrix} \omega_{uu} & \omega_{ux} \\ \omega_{xu} & \Omega_{xx} \end{bmatrix}, \tag{3}$$

and $B^{(1/m)}(r)$ may also be partitioned conformably as $(B_u^{(1/m)}, B_x^{(1/m)\prime})'$. I assume that

[A3] $\Omega, \Omega_{xx} > 0$.

It is worth reiterating that the *high*-frequency invariance principle in Lemma 1 may be viewed analogously to the *low*-frequency invariance principle of Chambers (2003, Lemma 2), since he defines a continuous-time DGP. In this light, the variance Ω depends on whether the vector series $(b_{t-i/m})$ contains stocks, flows, or combinations. By defining a discrete-time DGP, I abstract from this distinction. Instead, I use Ω as the benchmark for efficiency of estimators using the mixed-frequency model below. However, Chambers's (2003) results suggest that efficient estimators defined in this way may not be efficient compared to infeasible estimators using continuously sampled series.

2.2 Aggregation and Mixed-frequency Model

Suppose that some aggregation scheme is imposed on the series (y_t) such that $y_t^{\varpi} \equiv \sum_{i=0}^{m-1} \varpi_{i+1} y_{t-i/m}$, where (ϖ_i) is a sequence of m deterministic weights. The superscript ϖ on y_t is simply to denote that temporal aggregation is present, but not to denote any particular aggregation scheme. Rather, the weights are quite general. I refer to the weights and the scheme interchangeably as the *regressand aggregation scheme*.

In order for the low-frequency model to be consistent with the high-frequency DGP, the regressors and error must inherit the regressand aggregation scheme. The aggregated low-frequency model becomes

$$y_t^{\varpi} = \beta' x_t^{\varpi} + u_t^{\varpi}, \tag{4}$$

where $x_t^{\varpi} \equiv \sum_{i=0}^{m-1} \varpi_{i+1} x_{t-i/m}$ and $u_t^{\varpi} \equiv \sum_{i=0}^{m-1} \varpi_{i+1} u_{t-i/m}$.

The model in (4) is equivalent to

$$y_t^{\varpi} = \beta' \sum_{i=0}^{m-1} \varpi_{i+1} x_t - \beta' \sum_{k=0}^{m-2} \sum_{i=k+1}^{m-1} \varpi_{i+1} \triangle^{(1/m)} x_{t-k/m} + u_t^{\varpi}, \tag{5}$$

using a finite-order Beveridge-Nelson decomposition.

In order to identify β and the cointegrating space, I assume that

$$\sum_{i=0}^{m-1} \varpi_{i+1} = 1 \tag{6}$$

in line with the nonlinear ADL literature, but in contrast to Pons and Sansó (2005). One of the two aggregation schemes analyzed by those authors features weights summing to m rather than to unity. Consequently, some of my results differ from theirs by a factor of m or $1/m$. This identification avoids the difficulty noted by those authors of changes in measurement from one scheme to another.

Introducing additional notation to simplify exposition, let $z_t \equiv (\triangle^{(1/m)} x_t', ..., \triangle^{(1/m)} x_{t-(m-2)/m}')'$, $\Pi \equiv (\varpi_2, ..., \varpi_m) \otimes I$, and $D \equiv D_\# \otimes I$, where $D_\#$ is an $m-1$ square matrix with units along the main diagonal and below and with zeros above, and where the identity matrices have dimensions given by the number of regressors in the DGP in (1). The aggregated models in (4) and (5) may be rewritten simply as

$$y_t^\varpi = \beta'(x_t - \Pi D z_t) + u_t^\varpi \tag{7}$$

using this notation.

In case the regressors are aggregated differently than the regressand, I denote the matrix of *regressor* aggregation weights as Υ. The model in (7) may be rewritten as

$$y_t^\varpi = \beta'(x_t - \Upsilon D z_t) + \psi' z_t + u_t^\varpi \tag{8}$$

by defining the vector $\psi \equiv D'(\Upsilon - \Pi)'\beta$. Note that all of the aggregated models in (4), (5), (7), and (8) have a common high-frequency DGP given by (1).

A key difference between my DGP and that of Pons and Sansó (2005) lies in the term $\psi' z_t$. Since I assume that the data are generated at the high frequency, a low-frequency model necessarily includes $\psi' z_t$, unless aggregation of the regressand and regressors is identical. The vector ψ may also be viewed as estimation error in a mixed-frequency model in which $\Upsilon = \hat{\Pi}$, an estimator of the weight matrix Π of the regressand.

The models in (4), (5), (7), and (8) may be either low-frequency or mixed-frequency models. If the regressor vector is $(x_t - \Pi D z_t)$ or $(x_t - \Upsilon D z_t)$, they are strictly (aggregated) low-frequency models. If the regressor vector is $(x_t', z_t')'$, they are mixed-frequency models.

It is worth noting at this point that the choice of analyzing an aggregated regressand rather than aggregated regressors is arbitrary. In applications such as the gasoline price elasticity model discussed below, the regressand is aggregated. On the contrary, in applications such as using macroeconomic covariates to explain a price (PPP, e.g.), the regressor is aggregated. Although I focus on the former case in this paper, the latter case could be easily accommodated by instead examining $y_t = \beta' x_t^\varpi + \Upsilon D w_t - \psi' w_t + u_t^\varpi$ with $w_t \equiv (\triangle^{(1/m)} y_t, ..., \triangle^{(1/m)} y_{t-(m-2)/m})'$ in place of (8). The error term suffers from aggregation and an omitted stationary regressor, just as that in (8), but it would also be explicitly correlated with the stationary regressors. Since such correlation is allowed, much of the analysis below should hold qualitatively if in fact the regressors are aggregated rather than regressand. Of course, both of these cases are nested by a cointegrated mixed-frequency system, such as that examined by Ghysels and Miller (2012). However, since cointegrated systems are typically examined in differences rather than in levels, that analysis is quite a bit different.

2.3   Low-frequency Invariance Principle

Because estimation of β in mixed-frequency models occurs at the lower frequency, asymptotic approximations at the lower frequency (as $T \to \infty$) are more convenient. However, since the DGP, which is also the benchmark model against which the mixed-frequency model should be compared, is defined at the higher frequency, all low-frequency asymptotics should be defined in terms of high-frequency limits.

All of the stationary series in (8) – error, low-frequency increment of the I(1) regressors, and I(0) regressors – may be expressed as $b_t^\varpi \equiv (u_t^\varpi, \triangle x_t', z_t')'$ at the lower frequency. Defining $v_t \equiv (u_t, ..., u_{t-(m-1)/m}, \triangle^{(1/m)} x_t', ..., \triangle^{(1/m)} x_{t-(m-1)/m}')'$,

$$A \equiv \begin{bmatrix} \varpi_1 & \cdots & \varpi_m & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & I & \cdots & I & I \\ \vdots & & \vdots & I & 0 & 0 & 0 \\ \vdots & & \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 0 & 0 & I & 0 \end{bmatrix} \quad \text{and} \quad F \equiv \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & I & \cdots & I \end{bmatrix}$$

allows $\sum_{i=0}^{m-1} b_{t-i/m} = Fv_t$ and $b_t^\varpi = Av_t$. An invariance principle for (b_t^ϖ), used for the subsequent asymptotic analysis, may be deduced from that for (v_t), which may be deduced from that for $(b_{t-i/m})$ in (2).

It follows from Lemma 1 that $m^{-1/2} F(T^{-1/2} \sum_{t=1}^{[Tr]} v_t) \to_d B^{(1/m)}(r)$. Denote the limiting distribution of $T^{-1/2} \sum_{t=1}^{[Tr]} v_t$ by $V(r)$, and define

$$\sum_{k=-\infty}^{\infty} \Gamma(k + i/m) \equiv \Omega^i = \begin{bmatrix} \omega_{uu}^i & \omega_{ux}^i \\ \omega_{xu}^i & \Omega_{xx}^i \end{bmatrix} \quad \text{and} \quad \Xi \equiv \begin{bmatrix} \Xi_{uu} & \Xi_{ux} \\ \Xi_{xu} & \Xi_{xx} \end{bmatrix},$$

where $\Gamma(z)$ is the autocovariance function,

$$\Xi_{uu} \equiv \begin{bmatrix} \omega_{uu}^0 & \cdots & \omega_{uu}^{m-1} \\ \vdots & \ddots & \vdots \\ \omega_{uu}^{1-m} & \cdots & \omega_{uu}^0 \end{bmatrix}, \Xi_{xx} \equiv \begin{bmatrix} \Omega_{xx}^0 & \cdots & \Omega_{xx}^{m-1} \\ \vdots & \ddots & \vdots \\ \Omega_{xx}^{1-m} & \cdots & \Omega_{xx}^0 \end{bmatrix}, \Xi_{xu} \equiv \begin{bmatrix} \omega_{xu}^0 & \cdots & \omega_{xu}^{m-1} \\ \vdots & \ddots & \vdots \\ \omega_{xu}^{1-m} & \cdots & \omega_{xu}^0 \end{bmatrix},$$

and $\Xi_{ux} \equiv \Xi_{xu}'$. The distribution of $B^{(1/m)}(r)$ coincides with that of $m^{-1/2} F V(r)$ where $V(r)$ is a vector Brownian motion with variance given by Ξ, so that $\Omega = m^{-1} F \Xi F'$.

Because of the normalization of the high-frequency increments to $1/m$, $\Omega^0 = \sum_{k=-\infty}^{\infty} \Gamma(k)$ but

$$\Omega = \sum_{k=-\infty}^{\infty} \sum_{i=0}^{m-1} \Gamma(k - (i+j)/m) = \sum_{i=0}^{m-1} \Omega^{i+j} \tag{9}$$

for any integer j. Note the slight difference from the standard definition of the long-run variance.

I assume that Ξ_{xx} is full rank. Otherwise, some of the short-differenced increments of the regressor vector contain superfluous information, causing perfect collinearity. Such superfluous increments may be dropped or else the length of the difference may be increased.[9]

Finally, using $b_t^\varpi = Av_t$, an invariance principle for (b_t^ϖ) is given by the following lemma.

---

[9]A linear interpolated series provides a counter-example. In that case, all of the short differences of the

Lemma 2. *Under the above assumptions, $T^{-1/2} \sum_{t=1}^{[Tr]} b_t^{\varpi} \to_d B(r)$, where $B(r) = AV(r)$ is a vector Brownian motion that may be partitioned as $B = (B_u^{\varpi}, B_x', B_z')'$ with variance given by*

$$A \Xi A' \equiv \Theta = \begin{bmatrix} \theta_{uu}^{\varpi} & \theta_{ux} & \theta_{uz}^{\varpi} \\ \theta_{xu} & \Theta_{xx} & \Theta_{xz} \\ \theta_{zu}^{\varpi} & \Theta_{zx} & \Theta_{zz} \end{bmatrix}, \tag{10}$$

where

$$\theta_{uu}^{\varpi} \equiv \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \varpi_{i+1} \varpi_{j+1} \omega_{uu}^{j-i}$$

$$\theta_{zu}^{\varpi} \equiv (\sum_{i=0}^{m-1} \varpi_{i+1} (\omega_{xu}^{i})', ..., \sum_{i=0}^{m-1} \varpi_{i+1} (\omega_{xu}^{i-(m-2)})')')'$$

$$\theta_{uz}^{\varpi} \equiv (\sum_{i=0}^{m-1} \varpi_{i+1} \omega_{ux}^{-i}, ..., \sum_{i=0}^{m-1} \varpi_{i+1} \omega_{ux}^{-i+(m-2)})$$

$\theta_{xu} \equiv \omega_{xu}$, $\theta_{ux} \equiv \omega_{ux}$, $\Theta_{xx} \equiv m\Omega_{xx}$, $\Theta_{xz} \equiv \iota_{m-1}' \otimes \Omega_{xx}$ *with defined ι_{m-1} as an $(m-1)$-vector of ones, $\Theta_{zx} \equiv \Theta_{xz}'$, and where Θ_{zz} is a matrix of the first $(m-1)$ blocks of rows and columns of Ξ_{xx}.*

The proof of the lemma follows directly from the preceding discussion with some straight-forward matrix manipulations. The subvectors θ_{xu} and θ_{ux} are invariant with respect to the aggregation weights, due to the properties in (6) and (9).

The invariance principle in Lemma 2 may be compared with that of Pons and Sansó (2005, Lemma 2.1), since both of these results present low-frequency invariance principles with limiting variances defined at the higher frequency. I allow for a general weighting scheme, nesting not only the two types of aggregation considered by Chambers (2003) and Pons and Sansó (2005), but also nesting the flexible nonlinear weighting schemes of Ghysels *et al.* (2004, 2006). (See Miller, 2011, for an application of this invariance principle to models with these schemes.)

Define $\Sigma \equiv var(b_t^{\varpi})$ and $\Lambda \equiv \sum_{k=0}^{\infty} cov(b_t^{\varpi}, b_{t-k}^{\varpi})$ to be the "contemporaneous" variance and a "one-sided" long-run variance.[10] Both Σ and Λ may be partitioned as

$$\Sigma = \begin{bmatrix} \sigma_{uu}^{\varpi} & \sigma_{ux}^{\varpi} & \sigma_{uz}^{\varpi} \\ \sigma_{xu}^{\varpi} & \Sigma_{xx} & \Sigma_{xz} \\ \sigma_{zu}^{\varpi} & \Sigma_{zx} & \Sigma_{zz} \end{bmatrix} \quad \text{and} \quad \Lambda = \begin{bmatrix} \lambda_{uu}^{\varpi} & \lambda_{ux}^{\varpi} & \lambda_{uz}^{\varpi} \\ \lambda_{xu}^{\varpi} & \Lambda_{xx} & \Lambda_{xz} \\ \lambda_{zu}^{\varpi} & \Lambda_{zx} & \Lambda_{zz} \end{bmatrix},$$

similarly to Θ.

2.4 Least Squares Estimation

As in the conventional case with a single frequency, least squares is consistent but neither asymptotically unbiased nor efficient. In the mixed-frequency case, the bias and inefficiency

---

regressors, $\triangle^{(1/m)} x_{t-i/m}$, are functions of the same long difference, $\triangle x_t$. Therefore, each short-difference has no marginal informational content, and the terms causing the singularity may be dropped with no efficiency loss.

[10]The variance of (b_t^{ϖ}) is not really contemporaneous, because the vector b_t^{ϖ} contains random variables observed at different times. However, it is nearly contemporaneous in the sense that b_t^{ϖ} contains random variables observed during the same low-frequency interval.

originate not only from the usual non-Gaussian distribution and nuisance parameters, but also from the stationary covariates (z_t) in (8). The coefficients of these covariates may or may not be estimated consistently, and they are not zero if the regressand and regressors are aggregated differently, creating additional nuisance parameters.

The result for least squares estimation of β is summarized in the following lemma.

Lemma 3. *Let the weight matrix Υ be either (i) chosen to be a nonstochastic matrix $\Upsilon_T = \Upsilon^*$ or (ii) estimated by Υ_T such that $\mathrm{plim}_{T\to\infty}\Upsilon_T \equiv \Upsilon^*$, where $\|\Upsilon^*\| < \infty$. The least squares estimator $\hat{\beta}_{LS}$ of β in (8) has a limiting distribution given by*

$$T(\hat{\beta}_{LS} - \beta) \to_d \left(\int B_x B_x' dr\right)^{-1} \left(\int B_x dB_u^{\varpi*} + \lambda_{ux}^{\varpi*\prime} - \Upsilon^* D\sigma_{zu}^{\varpi*}\right)$$

where $B_u^{\varpi*} \equiv B_u^{\varpi} + B_z'\psi^*$, $\lambda_{xu}^{\varpi*} \equiv \lambda_{xu}^{\varpi} + \Lambda_{xz}\psi^*$, $\sigma_{zu}^{\varpi*} \equiv \sigma_{zu}^{\varpi} + \Sigma_{zz}\psi^*$, and

$$\psi^* \equiv D'(\Upsilon^* - \Pi)'\beta \tag{11}$$

under the above assumptions and as $T \to \infty$.

At this point, Υ_T (and thus Υ^*) may be quite general. Υ_T may be selected arbitrarily or else be an estimator. Its limit is not generally Π.

The keys to efficiency are (i) to find a class of estimators to orthogonalize $B_u^{\varpi*}$ with respect to B_x and eliminate the nuisance parameter $\lambda_{xu}^{\varpi*\prime} - \Upsilon^* D\sigma_{zu}^{\varpi*}$, and (ii) to minimize the variance of the estimators within this class. All of the estimators I consider beyond least squares accomplish the former and are asymptotically unbiased and mixed normal by construction, providing valid asymptotic inference. However, not all provide *efficient* inference by accomplishing the latter.

For arbitrary regressand and regressor weighting schemes Π and Υ, a CCR or FM-OLS regression may be constructed based on the results of the lemma. Efficiency comparisons like those of Chambers (2003) and Pons and Sansó (2005) for particular weighting schemes may be conducted on these regressions. With only a low-frequency model in mind, the preliminary results of this section generalize the results of Pons and Sansó (2005) by allowing more general weighting schemes.

3 Efficiency Bounds and Efficient Estimation

One of the main innovations of this paper lies in using the preliminary results of the previous section to obtain efficiency bounds for either mixed-frequency or (aggregated) low-frequency cointegrating regressions, providing precise metrics for efficiency of such models. The remaining theoretical innovations lie in creating estimators that are more efficient using these metrics.

3.1 Efficiency Bounds

Saikkonen (1991) and Phillips (1991) extensively discussed efficient estimation of cointegrating relationships, synthesizing earlier literature that focused primarily on inference. As mentioned above, all of the estimators in this section are consistent and asymptotically normal. To focus on the issues of aggregation and mixed frequencies, I consider the efficiency bound to be the asymptotic variance of the CCR estimator of the infeasible full-information high-frequency model in (1).

Using the invariance principle in Lemma 1, an efficient high-frequency estimator $\hat{\beta}_{EFF}$ of β must have a limiting distribution given by

$$M(\hat{\beta}_{EFF} - \beta) \to_d \left(\int B_x^{(1/m)} B_x^{(1/m)\prime} dr \right)^{-1} \int B_x^{(1/m)} dB_{u \cdot xx}^{(1/m)\prime},$$

where $B_{u \cdot xx}^{(1/m)}(r) \equiv B_u^{(1/m)}(r) - \omega_{ux} \Omega_{xx}^{-1} B_x^{(1/m)}(r)$. By Lemmas 1 and 2, the vectors of Brownian motion $B_x^{(1/m)}(r)$ and $B_x(r)$ are related by $B_x^{(1/m)} = m^{-1/2} B_x$. Moreover, since $M/(Tm) \to 1$ as $T \to \infty$, the limiting distribution above may be rewritten as

$$T(\hat{\beta}_{EFF} - \beta) \to_d m^{-1/2} \left(\int B_x B_x' dr \right)^{-1} \int B_x dB_{u \cdot xx}^{(1/m)},$$

which has conditional variance given by

$$\left(\int B_x B_x' dr \right)^{-1} \theta_{uu \cdot x} \tag{12}$$

where $\theta_{uu \cdot x} \equiv m^{-1} (\omega_{uu} - \omega_{ux} \Omega_{xx}^{-1} \omega_{xu})$.

This conditional variance is a translation of the high-frequency lower bound into the lower frequency. The factor $(\int B_x B_x' dr)^{-1}$ is common to the distributions of both efficient and inefficient estimators across different types of aggregation. Ignoring the common factor, the comparisons in the ensuing discussion focus on the remaining factor, which is bounded below by $\theta_{uu \cdot x}$. Since this bound is clearly not conditional on the type of aggregation used, I refer to this variance as the *aggregation-unconditional efficiency bound*.

The fact that $\theta_{uu \cdot x}$ is smaller than $\omega_{uu} - \omega_{ux} \Omega_{xx}^{-1} \omega_{xu}$ by a factor of m^{-1} reflects the necessity that distributions of low-frequency estimators must collapse faster to obtain the high-frequency bound. Chambers (2003) and Pons and Sansó (2005) showed that aggregating high-frequency data *may* accomplish this task. A lower bound that is conditional on aggregation would be useful, since aggregation is already present in mixed-frequency series.

A canonical cointegrating regression performed on the high-frequency DGP in (1) has an error term given by $\kappa' b_{t-i/m}$, where $\kappa \equiv (1, -\omega_{ux} \Omega_{xx}^{-1})'$. At the lower frequency with known weights $\Upsilon = \Pi$, so that $\psi = 0$, the error term becomes $\kappa' w_t^\varpi$ with $w_t^\varpi \equiv (u_t^\varpi, \triangle x_t')'$ and with $\kappa \equiv (1, -\theta_{ux} \Theta_{xx}^{-1})'$ redefined. With possibly unknown weights, the error term is augmented by $\psi' z_t \neq 0$, so that it becomes $\kappa_\psi' b_t^\varpi$, where I define

$$\begin{aligned} \kappa_\psi &\equiv (1, -(\theta_{ux} + \psi' \Theta_{zx}) \Theta_{xx}^{-1}, \psi')' \\ &= (1, -m^{-1}(\omega_{ux} \Omega_{xx}^{-1} + \psi'(\iota_{m-1} \otimes I)), \psi')'. \end{aligned}$$

The long-run covariance of any of these error terms with the first difference of the regressors is zero, providing Gaussian alternatives to the limiting distribution in Lemma 3.

It is straightforward to minimize the long-run variance of $\kappa_\psi' b_t^\varpi$ over ψ. Some straightforward matrix manipulations show that this variance is

$$\kappa_\psi' \Theta \kappa_\psi = \theta_{uu \cdot x}^\varpi + 2\theta_{uz \cdot x}^\varpi \psi + \psi' \Theta_{zz \cdot x} \psi$$

where $\theta_{uu \cdot x}^\varpi \equiv \theta_{uu}^\varpi - m^{-1}\omega_{ux}\Omega_{xx}^{-1}\omega_{xu}$, $\Theta_{zz \cdot x} \equiv \Theta_{zz} - \Theta_{zx}\Theta_{xx}^{-1}\Theta_{xz}$, $\theta_{uz \cdot x}^\varpi \equiv \theta_{uz}^\varpi - \theta_{ux}\Theta_{xx}^{-1}\Theta_{xz}$, and $\theta_{zu \cdot x}^\varpi \equiv \theta_{uz \cdot x}^{\varpi\prime}$. The minimum variance

$$\min_\psi \kappa_\psi' \Theta \kappa_\psi = \theta_{uu \cdot x}^\varpi - \theta_{uz \cdot x}^\varpi \Theta_{zz \cdot x}^{-1} \theta_{zu \cdot x}^\varpi$$

is attained at

$$\psi_{\min} = -\Theta_{zz \cdot x}^{-1} \theta_{zu \cdot x}^\varpi, \tag{13}$$

and I refer to the conditional variance

$$\left(\int B_x B_x' dr \right)^{-1} \left(\theta_{uu \cdot x}^\varpi - \theta_{uz \cdot x}^\varpi \Theta_{zz \cdot x}^{-1} \theta_{zu \cdot x}^\varpi \right) \tag{14}$$

as the *aggregation-conditional efficiency bound*.

In general, ψ^* in (11) does not equal $\psi_{\min}$. In other words, using the least squares residual in a conventional procedure, such as FM-OLS or CCR, does not yield the aggregation-conditional efficiency bound.

Given a known regressand aggregation scheme, an intuitive way to try to achieve the aggregation-conditional efficiency bound is to set $\psi^* = 0$. This means using the same aggregation scheme for the regressors, thus avoiding the mixed-frequency problem entirely. From Lemma 3, the limiting distribution of the least squares estimator resembles that of the least squares estimator in the conventional case. It is straightforward to deduce that a conventional method such as FM-OLS or CCR yields an estimator with a conditional variance having a second factor of $\theta_{uu \cdot x}^\varpi$ in this case, rather than that in (12) or in (14).

The following proposition formalizes the relationship between $\theta_{uu \cdot x}$, $\theta_{uu \cdot x}^\varpi$, and $\theta_{uu \cdot x}^\varpi - \theta_{uz \cdot x}^\varpi \Theta_{zz \cdot x}^{-1} \theta_{zu \cdot x}^\varpi$.

Proposition 4. *In general,*

[a] $\theta_{uu \cdot x} \leq \theta_{uu \cdot x}^\varpi - \theta_{uz \cdot x}^\varpi \Theta_{zz \cdot x}^{-1} \theta_{zu \cdot x}^\varpi \leq \theta_{uu \cdot x}^\varpi$.

If the regressors are strictly exogenous, such that $\Xi_{xu} = 0$, then

[b] $\theta_{uu \cdot x} \leq \theta_{uu \cdot x}^\varpi - \theta_{uz \cdot x}^\varpi \Theta_{zz \cdot x}^{-1} \theta_{zu \cdot x}^\varpi = \theta_{uu \cdot x}^\varpi$.

If either (i) the aggregation weights are constant or (ii) there is no variation across the rows of Ξ_{ux} and Ξ_{uu}, then

[c] $\theta_{uu \cdot x} = \theta_{uu \cdot x}^\varpi - \theta_{uz \cdot x}^\varpi \Theta_{zz \cdot x}^{-1} \theta_{zu \cdot x}^\varpi = \theta_{uu \cdot x}^\varpi$.

Part [a] may seem to be somewhat counterintuitive. If the regressand aggregation scheme is *known*, it is *not* efficient to use these weights for aggregating the regressors, except in the special cases in parts [b] and [c]. Instead, the optimal weights are functions of the covariance of the error term with differences of the regressors.

For aggregated low-frequency models, the results of Chambers (2003) and Pons and Sansó (2005) may be re-interpreted in light of the proposition. Those authors considered two types of aggregation – average sampling, in which $\varpi_{i+1} = m^{-1}$ for all i,[11] and selective sampling, in which $\varpi_1 = 1$ and $\varpi_{i+1} = 0$ for $i \geq 1$.

Clearly, it might be possible to attain the aggregation-unconditional bound in aggregated low-frequency models. Part [c] of the above proposition supports the findings of Chambers (2003) and Pons and Sansó (2005) that average sampling both sides of the regression is just as efficient as the infeasible high-frequency DGP.[12] Pons and Sansó (2005) further suggest average sampling the regressors even if the regressand is selectively sampled. In this case, Π is a matrix of zeros, but $\Upsilon^* = m^{-1}(\iota'_{m-1} \otimes I)$, so that $\psi^* = m^{-1}(D'_\# \iota_{m-1} \otimes \beta)$. Since $\psi_{\min} = -\Theta_{zz \cdot x}^{-1} \theta_{zu \cdot x}^{\varpi}$, their suggested scheme is not efficient.

Proposition 4 is also useful in evaluating estimators in mixed-frequency models. Given a regressand aggregation scheme, I henceforth consider a conditionally efficient estimator to be one that attains the aggregation-conditional efficiency bound. As Proposition 4 suggests, such estimators are *not* generally efficient in the unconditional high-frequency sense.

3.2   Estimation of β with Unknown Weights

Some of the discussion so far has focused on aggregated low-frequency models, in order to compare results with the existing literature. I now focus exclusively on mixed-frequency models, in which (z_t) is observable. I first consider an *unknown* regressand aggregation scheme. The unknown case may be considered to be a proxy for systematic measurement error. Even if the weighting scheme is known, there may be hidden systematic adjustments, such as seasonal adjustment or regular but mistimed sampling. The latter might occur, for example, if a weekly regressand is sampled on a day unknown to the analyst. On the other hand, allowing unknown weights does not accommodate cases of irregular or random weights, such as that considered by Jordá (1999), e.g., where series are observed at the frequency at which agents make decisions. The weights are constant here.

With Π unknown, it may either be (i) ignored and left to the error term, so that $\Upsilon^* = 0$, (ii) set arbitrarily to some other Υ^*, or (iii) estimated by $\Upsilon = \hat{\Pi}$ for some estimator $\hat{\Pi}$ of Π. Choice (i) amounts to selective sampling of the last high-frequency observation in each low-frequency period. Choices (i) and (ii) amount to much the same thing: choosing Υ^* arbitrarily and suboptimally.

Choice (iii) may be more reasonable in the sense that its goal is to estimate Π, so that $\hat{\psi} \equiv D'(\hat{\Pi} - \Pi)'\beta$ may be close to zero to diminish the effect of $\hat{\psi}'z_t$ on the error term. With a consistent estimator of β from Lemma 3, the following theorem holds.

---

[11]Constant weights given by $\varpi_{i+1} = 1$ considered by Pons and Sansó (2005), do not satisfy the identifying restriction that I impose in (6).

[12]Further, Chambers (2003) showed that average sampling is just as efficient as having continuously recorded series when all series are flows.

Theorem 5. *Given a consistent estimator $\hat{\beta}$ of β, the least squares estimator $\hat{\Pi}_{LS}$ of Π in (7) is such that*

$$\psi^*_{LS} = \text{plim}_{T \to \infty} \hat{\psi}_{LS} = \text{plim}_{T \to \infty} D'(\hat{\Pi}_{LS} - \Pi)'\hat{\beta} = -\Sigma_{zz}^{-1}\sigma_{zu}^{\varpi}$$

under the above assumptions and as $T \to \infty$.

Using this strategy, $-\Sigma_{zz}^{-1}\sigma_{zu}^{\varpi} = \psi^*_{LS} \neq \psi_{\min} = -\Theta_{zz \cdot x}^{-1}\theta_{zu \cdot x}^{\varpi}$ in general. It is possible that $\psi^*_{LS} = 0$, but only if the stationary regressors are contemporaneously uncorrelated with the error term, so that $\sigma_{zu}^{\varpi} = 0$. In general, $\hat{\Pi}_{LS}$ is *not* a consistent estimator of Π.

The advantage of this approach is a value of ψ^* such that $\sigma_{zu}^{\varpi*} \equiv \sigma_{zu}^{\varpi} + \Sigma_{zz}\psi^* = 0$, eliminating the final nuisance parameter $\Upsilon^* D\sigma_{zu}^{\varpi*}$ in the limiting distribution of the least squares estimator of β in Lemma 3. However, ψ^*_{LS} is not identified, since its definition involves unknown Π. The penultimate nuisance parameter thus remains.

Using least squares, the residual is $\hat{u}_t^{\varpi} + \hat{\psi}'_{LS}z_t$. In this case, $\hat{\psi}_{LS}$ is estimation error, but when $\psi^*_{LS} \neq 0$, it is part of the residual. As such, a feasible version of b_t^{ϖ} cannot be identified. Alternatively, let

$$C_1' \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad \text{and} \qquad C_2' \equiv \begin{bmatrix} 1 & 0 & \psi^{*\prime}_{LS} \\ 0 & 1 & 0 \end{bmatrix},$$

and consider $\kappa'_\psi C_1 C_2' b_t^{\varpi}$. The linear transformation $C_2' b_t^{\varpi} = (u_t^{\varpi} + \psi^{*\prime}_{LS}z_t, \triangle x_t')'$ is identified.

The long-run variance $C_2'\Theta C_2$ and associated variances $C_2'\Sigma C_2$ and $C_2'\Lambda C_2$ can be feasibly estimated using the residuals identified in this way. It only remains to estimate $-(\theta_{ux} + \psi'\Theta_{zx})\Theta_{xx}^{-1}$ in $C_1'\kappa_\psi$. Estimating Θ_{xx} is accomplished from the long difference of the regressors, from the limit in (10) of Lemma 2. Also from that limit, the long-run covariance between the residuals $\hat{u}_t^{\varpi} + \hat{\psi}'_{LS}z_t$ and the long difference of the regressors $\triangle x_t$ is clearly $\theta_{ux} + \psi^{*\prime}_{LS}\Theta_{zx}$, so that ψ^*_{LS} does not need to be identified.

As a result, the conventional feasible procedures are still feasible when the unidentified term $\psi^{*\prime}_{LS}z_t$ becomes part of the error term. The estimator is asymptotically mixed normal, and the nuisance parameters in Lemma 3 are eliminated – the first $\lambda_{ux}^{\varpi*\prime}$ by the procedure itself and the second $\Upsilon^* D\sigma_{zu}^{\varpi*}$ by least squares estimation. However, not knowing Π means that the aggregation-conditional efficiency bound is not attained, since $\psi^*_{LS} \neq \psi_{\min}$ in general.

3.3   Estimation of β with Known Weights

With known Π, the least squares estimator in Lemma 3 can be calculated directly with $\psi^* = 0$ – i.e., by using the same aggregation scheme on both sides. Setting $\psi^* = 0$ guarantees $\omega_{uu \cdot x}^{\varpi}$, but this is not the aggregation-conditional bound in the general case of part [a] of Proposition 4. A smaller variance is possible. The mixed-frequency CCR proposed in this section attains the aggregation-conditional bound asymptotically.

An example in which the regressand in selectively sampled provides some intuition. Information is lost from the selection, inflating the variance. If the regressors are selectively sampled in the same way, $\psi^* = 0$, the information loss is neither worsened nor ameliorated. However, if the short differences of the regressors are correlated with the error

term, $\Xi_{xu} \neq 0$, including lagged regressors provides additional information about the omitted high-frequency errors. This information may be exploited to recover the aggregation-conditional bound.

Estimation proceeds in several steps. First, the regression in (8) is estimated using least squares with $\Upsilon = \Pi$, so that $\psi^* = 0$. This regression provides consistent estimates of both β and u_t^ϖ by Lemma 3. The residual series $(\hat{u}_t^\varpi)$ is used in $(\hat{b}_t^\varpi)$ to estimate the variances Σ, Θ, and Λ.

Once these variances are identified using $\psi^* = 0$, the minimum variance may be obtained by estimating $\psi_{\min}$ to be used in κ_ψ. Either FM-OLS or CCR is valid, but using κ_ψ rather than $C_1' \kappa_\psi$ as in the case of unknown weights.

The novelty of this approach lies in estimating $\psi_{\min}$ before the final step. Once Θ is estimated, the formula

$$\psi_{\min} = -(\Theta_{zz} - \Theta_{zx}\Theta_{xx}^{-1}\Theta_{xz})^{-1}(\theta_{zu}^\varpi - \Theta_{zx}\Theta_{xx}^{-1}\theta_{xu}^\varpi)$$

from (13) may be applied.

A CCR may be constructed using either the model in (4) or one of the equivalent models in (5), (7), or (8). However, a single long difference of each integrated regressor $\triangle x_t$ should be used, which is intuitively based on the models in (5), (7), or (8). Basing the CCR on that in (4) would require weighing short differences of each lag of each regressor, adding superfluous terms and an unnecessary computational burden.

The mixed-frequency CCR differs from the conventional CCR in two ways. First, $\kappa_{\psi\min}$ in (13) used in the second step is more complicated than $\psi = 0$ corresponding to the conventional case. Second, the coefficients of Dz_t are known up to a consistent estimate of β, so this term may be subtracted from both sides.

Specifically, let

$$x_t^{\varpi*} \equiv x_t - \Lambda_{\bullet x}'\Sigma^{-1}b_t^\varpi$$
$$y_t^{\varpi*} \equiv y_t^\varpi + \beta'(\Pi Dz_t - \Lambda_{\bullet x}'\Sigma^{-1}b_t^\varpi) - (\theta_{ux} + \psi_{\min}'\Theta_{zx})\Theta_{xx}^{-1}\triangle x_t + \psi_{\min}'z_t$$

where $\Lambda_{\bullet x}$ is a submatrix of Λ given by all rows but only the columns corresponding to $\triangle x_t$. The model in (7) may be rewritten as

$$y_t^{\varpi*} = \beta' x_t^{\varpi*} + u_t^{\varpi*} \tag{15}$$

where $u_t^{\varpi*} \equiv u_t^\varpi - (\theta_{ux} + \psi_{\min}'\Theta_{zx})\Theta_{xx}^{-1}\triangle x_t + \psi_{\min}'z_t = \kappa_{\psi\min}'b_t^\varpi$. This model reduces to the conventional CCR considered by Park (1992) if $\psi_{\min} = 0$ and $\Pi = 0$.

Theorem 6. *Given weights Π, variances Σ, Θ, and Λ, and a consistent estimator $\hat{\beta}$ of β such that $y_t^{\varpi*}$ is defined using $\hat{\beta}$, the least squares estimator $\hat{\beta}_{CCR}$ of β in the model in (15) has a limiting distribution given by*

$$T(\hat{\beta}_{CCR} - \beta) \to_d \left(\int B_x B_x' dr\right)^{-1} \int B_x dB' \kappa_{\psi\min},$$

under the above assumptions and as $T \to \infty$. This estimator attains the aggregation-conditional efficiency bound.

The theorem provides an asymptotically aggregation-conditional efficient estimator of the cointegrating vector.

To the best of this author's knowledge, efficient estimators for mixed-frequency cointegrating regressions have not been analyzed previously, except in the special case in which the regressand has been average sampled. More generally, such estimators do not attain the same level of efficiency in comparison with the infeasible DGP. The proposed estimator attains the efficiency bound defined conditionally on the regressand aggregation scheme, nesting the unconditional bound as a special case.

The proposed mixed-frequency CCR is feasible once the variance matrices are replaced by consistent estimators.[13] Using residuals $(\hat{u}_t^\varpi)$ from the least squares regression in the first step, that standard covariance estimation procedures are consistent may be deduced from the low-frequency invariance principle in Lemma 2.

4   Finite-Sample Comparisons (Known Weights)

Finite-sample experiments provide clear comparisons of the efficiency gains under different data-generating assumptions. All of the procedures discussed above, including least squares, are consistent. Moreover, all of the conventional techniques provide estimators with limiting mixed normal distributions. Since a goal of efficient estimation is to improve the power of hypothesis tests, I base the comparisons on size-adjusted power functions of simple t-tests of β, which is a scalar in this exercise.

I use a data-generating process similar to that of Pons and Sansó (2005) for the simulations. Specifically, I let

$$b_{t-i/m} = \begin{bmatrix} \rho & 0 \\ 0 & 0 \end{bmatrix} b_{t-(i+1)/m} + \varepsilon_{t-i/m}, \quad \varepsilon_{t-i/m} \sim \mathbf{N}\left(0, \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}\right),$$

varying both ρ and α. Clearly, ρ controls the serial correlation of the error, while α controls the correlation of the error with the regressors.

In order to simulate 60 years of quarterly and monthly data, I let $M = 720$ and $m = 3$, so that $T = 240$. I repeat all simulations $10{,}000$ times. I then repeat the whole exercise for 60 years of annual and monthly data, with $m = 12$ and $T = 60$. In order to most clearly illustrate the efficiency of each technique, I use the true variances based on the DGP above in calculating $y_t^{\varpi*}$ and $x_t^{\varpi*}$, rather than estimating the variances. Any loss of efficiency therefore stems from aggregation rather than from a smaller sample size than the full-information benchmark.

---

[13]The feasibility of the mixed-frequency CCR presupposes that the sample size is large enough to accommodate the requisite number of high-frequency regressors. It would *not* be feasible with 50 years of annual and weekly data, for example. In this case, temporal aggregation may be required. As a first approximation, aggregation of the regressors should match that of the regressand, as in the first step of the mixed-frequency CCR above. Tests to distinguish between these schemes, such as the tests proposed by Andreou *et al.* (2010) and Miller (2011), would be useful for this purpose.

I consider the same types of aggregation as Chambers (2003) and Pons and Sansó (2005). However, I show power functions only for a selectively sampled regressand. The power functions for an average-sampled regressand are not informative, since part [c] of Proposition 4 suggests no difference in the efficiency bounds in that case.

Figures 1 and 2 show size-adjusted power functions for $\rho = (0, 0.5, 0.9)$, rows of panels, and $\alpha = (0, 0.5, 0.9)$, columns of panels. Specifically, Figure 1 shows power functions with $m = 3$, while Figure 2 shows those with $m = 12$.

Each panel displays the power functions from t-tests based on four different estimators. The benchmark estimation is least squares estimation of a CCR created from the high-frequency model in (1), similar to Park's (1992) CCR. The remaining three estimators are based on a selectively-sampled regressand. The case in which both the regressand and regressors are selectively sampled is labeled "same schemes" in the figures, while the label "mixed schemes" refers to the case in which the regressors are average sampled, even though the regressand is selectively sampled. The fourth estimator uses the same aggregation schemes, but with $\psi_{\min}$ derived above. I label this as "same schemes, adjusted" in the figures and refer to it as *adjusted estimation* below.

I set $\beta = 10$. Since this is a relatively large value, the extra term with $\psi = m^{-1}(D'_{\#}\iota_{m-1} \otimes \beta)$ for mixed aggregation schemes has a large impact. In each case, the size-adjusted power shows substantially weaker tests based on mixed aggregation schemes. The tests are even biased over a range of alternatives. (Similar results – not shown – hold for an average-sampled regressand but selectively sampled regressor, the opposite mix.) The results suggest that aggregation schemes should *not* be mixed, in contrast to the recommendation of Pons and Sansó (2005). They found no appreciable differences in test sizes, but recommended always average-sampling the regressors, even when the regressand is selectively sampled, based on precision of the estimator when the cointegrating space is allowed to vary. My results instead suggest that matching aggregation schemes, so that $\psi = 0$, is preferable to mixing aggregation schemes based on power. The tests are unbiased and more powerful when the schemes are matched.

The resulting power from matched schemes lies below that from the full-information case, because selective sampling necessarily suppresses information. However, as the serial correlation in the error term increases from 0 to 0.9 in the figures (top to bottom), two things are apparent. First, the overall power of tests using all of the techniques decreases, as the root in the error term approaches unity. Recent theoretical results of Kurozumi and Hayakawa (2009) suggest that CCR remains consistent even when the error contains a root near unity. Second, increasing the error dependence decreases the marginal value of the information loss from selective sampling. At $\rho = 0.9$, the power functions from matching schemes are nearly indistinguishable from the corresponding full-information benchmarks.

Matching regressor and regressand aggregation schemes clearly does not generally attain the aggregation-unconditional efficiency bound. As part [a] of Proposition 4 suggests, it does not generally attain the aggregation-conditional bound, either. No improvement from adjusted estimation is expected when $\alpha = 0$, since $\theta^{\varpi}_{zu}, \theta^{\varpi}_{xu} = 0$ so that $\psi_{\min} = 0$. However, the figures show an improvement in power when $\alpha \neq 0$. The intuition underlying this result is as follows. When the marginal information loss from selectively sampling the regressand is high – i.e., ρ is small, and the correlation between the error and regressors is high – i.e., α
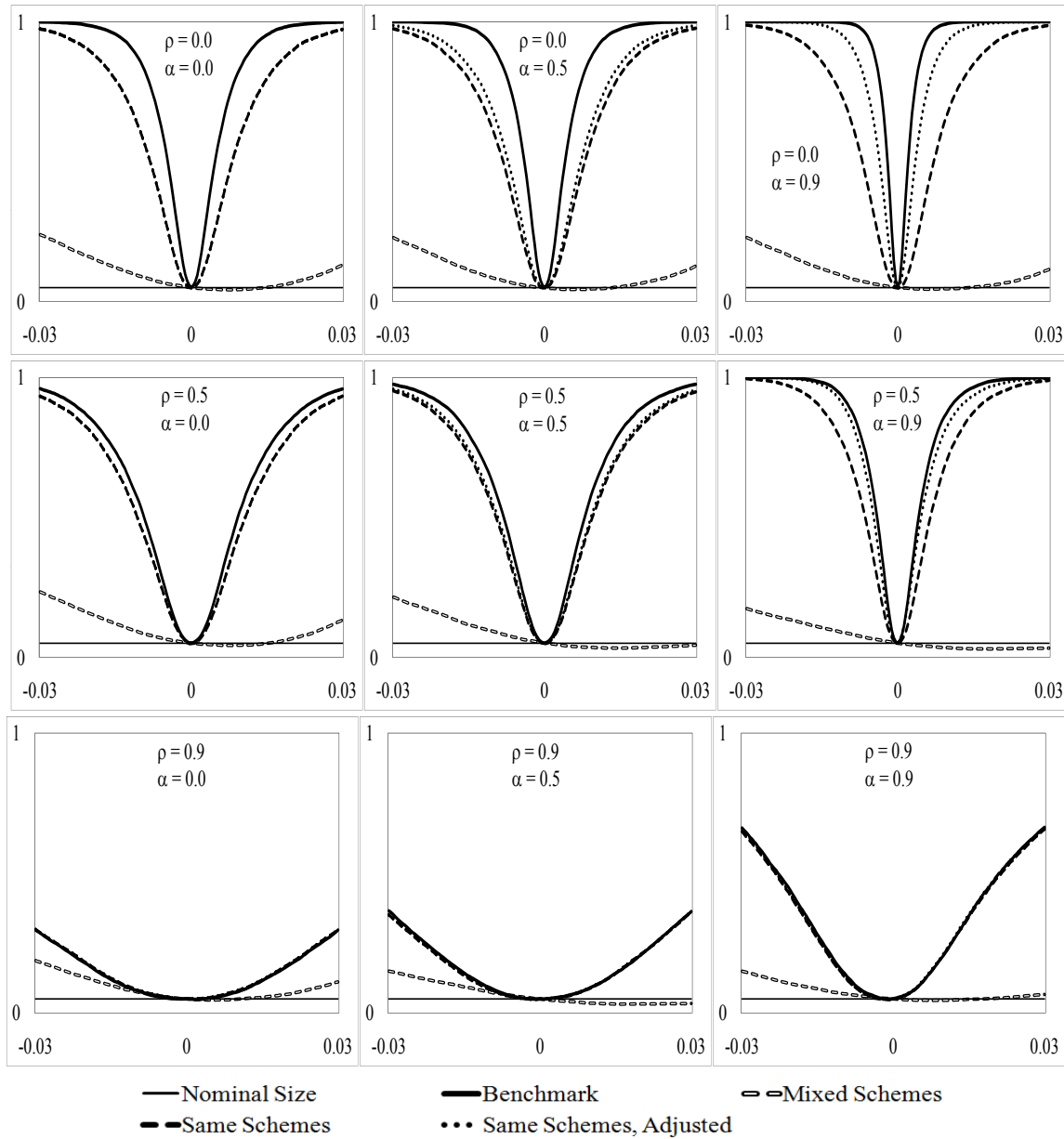
Figure 1: Size-adjusted power for $m = 3$, $\beta = 10$, $\rho, \alpha = (0, 0.5, 0.9)$ with a selectively sampled regressand.
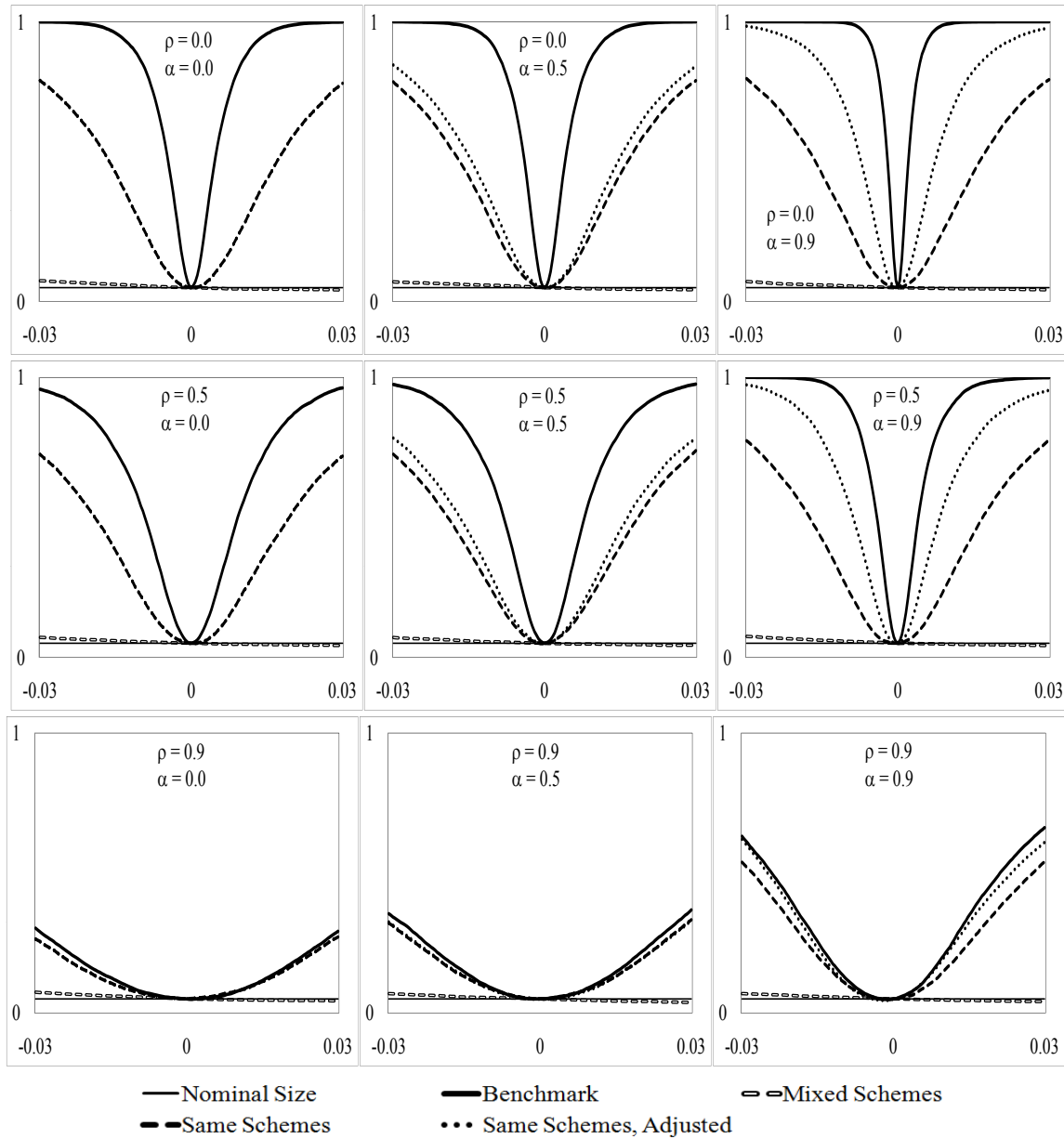
Figure 2: Size-adjusted power for $m = 12$, $\beta = 10$, $\rho, \alpha = (0, 0.5, 0.9)$ with a selectively sampled regressand.

is large, then using the correlated regressors to compensate for the information loss in the error provides more power. The inequalities of the factors $\theta_{uu\cdot x}$, $\theta_{uu\cdot x}^{\varpi} - \theta_{uz\cdot x}^{\varpi}\Theta_{zz\cdot x}^{-1}\theta_{zu\cdot x}^{\varpi}$, and $\theta_{uu\cdot x}^{\varpi}$ are clearly illustrated in these cases, and the aggregation-conditional efficiency bound cannot be attained without the proposed $\psi_{\min}$ adjustment.

5 Application: Gasoline Price Elasticity (Unknown Weights)

Estimating the demand equation for a good, such as gasoline, provides an application in which one of the regressors, nominal price, may be obtained at a high frequency than the regressand, quantity consumed. Gasoline price inelasticity provides, to some extent, a measure of consumers' dependence on gasoline. Consumers who cannot easily substitute away from gasoline consumption (e.g., using alternative means of transportation) will exhibit an inelastic short-run price elasticity, because they cannot easily decrease gasoline consumption in response to price increases. As a result, the elasticity has implications for any policy that could affect oil or gasoline prices, including taxation and environmental policies.

Since the oil price hikes of the 1970s, there has been a vast literature on estimating the short-run price elasticity of gasoline demand. See Dahl and Sterner (1991), e.g., for an extensive review of the early literature. With gasoline prices soaring, price elasticity has once again become an important concern. A number of studies (e.g., Schmalensee and Stoker, 1999; Small and Van Dender, 2007; Hughes *et al.*, 2008) have examined more recent data. While many of these studies use cross-sectional or panel data, a few estimate cointegrating regressions using aggregate data.[14]

Let Q_t, R_t, and I_t denote the quantity of gasoline consumed, the real price of gasoline, and real income, respectively, and let lower case letters q_t, r_t, and i_t represent their logarithmically transformed quantities. A regression to estimate gasoline short-run price elasticities may be written as

$$q_t = \alpha' d_t + \beta r_t + \gamma i_t + u_t,$$

where β is the short-run price elasticity and γ is the short-run income elasticity. d_t is vector of monthly indicators included to control for seasonality, but the coefficient vector α may also account for scale transformations of the other regressors.

Suppose that the real price of gasoline may be represented as $R_t \equiv P_t/C_t$, where P_t is the nominal gasoline price and C_t is a price deflator, such as the consumer price index. Clearly, $r_t = p_t - c_t$ using the same notational convention. If nominal prices are available daily, the monthly nominal price is a weighted geometric mean of daily prices:

$$P_t = \sqrt[m_t]{\prod_{i=0}^{m_t-1} P_{t-i/m}^{\pi_{i+1}}} \quad \text{or} \quad p_t = \frac{1}{m_t}\sum_{i=0}^{m_t-1} \pi_{i+1}p_{t-i/m},$$

so that the log monthly price is a weighted arithmetic mean of the log daily price. (Here, m_t denotes the number of days in month t.) The regression may be rewritten as

$$q_t = \alpha d_t + \beta \sum_{i=0}^{m_t-1} \varpi_{t,i+1}(p_{t-i/m} - c_t) + \gamma i_t + u_t \tag{16}$$

---

[14]Hughes *et al.* (2008) provides an example for gasoline demand. Silk and Joutz (1997) apply a cointegrating regression to the related problem of estimating electricity demand.

by defining $\varpi_{t,i+1} \equiv \pi_{i+1}/m_t$ and imposing $\sum_{i=0}^{m_t-1} \varpi_{t,i+1} = 1$.

Of the four covariates in the model, three may be viewed as subject to aggregation: q_t, c_t, and i_t. From the analysis above, it is clear that differences between the aggregation scheme of the regressand q_t and those of each of the regressors may cause a loss of efficiency due to an omitted variable. However, aggregation of c_t and i_t are upstream of the analyst in practice. In contrast, a measure of p_t, spot price for gasoline, is available daily. The focus of this exercise is to show how the daily price may be used and the implications for efficiency.

I consider explicitly four approaches using the daily price. In the absence of measurement error, monthly gasoline consumption should simply be the sum of daily consumption, so that equal weights should be assigned: $\varpi_{t,i+1} = \varpi_t$ (average sampling). Second, I consider sampling at the end of each month (end-of-period sampling, EOP), so that $\varpi_{t,1} = 1$ and $\varpi_{t,i+1} = 0$ otherwise. Third, I consider sampling at the beginning of each month (beginning-of-period sampling, BOP), so that $\varpi_{t,m_t} = 1$ and $\varpi_{t,i+1} = 0$ otherwise. These approaches correspond to choices (i) and (ii) discussed in Section 3.2 above.

As the fourth approach, I consider estimating $\varpi_{t,i+1}$ by specifying different regressors for different days of the month, corresponding to choice (iii) in Section 3.2. In order to eliminate missing variables, I estimate weights using only the last 18 trading days of each month, since every month in the sample contains at least that many.[15] Conceptually, when the weights are unknown, the model above is estimated by either setting using $m_t = 18$ or by restricting $\varpi_{t,i+1} = \varpi_{s,i+1}$ for $i + 1 \leq 18$ and all $s, t = 1, ..., T$, and $\varpi_{t,i+1} = 0$ for $i + 1 > 18$. The fourth approach is still *less* restrictive than the first three approaches, which restrict *all* of the weights.

I impose $\sum_{i=0}^{17} \varpi_{i+1} = 1$ as an identifying restriction in the fourth approach. That is, I estimate 18 OLS coefficients of the form $\widehat{\beta \varpi_{i+1}}$ and then identify $\hat{\beta}$ as $\sum_{i=0}^{17} \widehat{\beta \varpi_{i+1}}$. A standard error for $\hat{\beta}$ is calculated accordingly.

For q_t and i_t, I use data that match those of Hughes *et al.* (2008) as closely as possible. Specifically, I use "U.S. product supplied of finished motor gasoline" from the Energy Information Administration (EIA)[16] and "real disposable personal income: per capita" (chained 2005 dollars, seasonally adjusted) from the St. Louis Fed (original source: Bureau of Economic Analysis).[17] Hughes *et al.* (2008) use a monthly real gasoline price, which is aggregated both temporally (across days) and cross-sectionally (across U.S. cities). I use daily spot prices for gasoline (New York Harbor) from the EIA[18] and the CPI from the St. Louis Fed (original source: Bureau of Labor Statistics)[19] to deflate these prices. I am making a trade-off between the time and space domains: I am choosing a relatively poor cross-sectional representation of prices by only looking at one location, but I get observa-

---

[15]This choice is arbitrary and necessitates some omitted variables, which will naturally affect power like the omitted variables discussed above. As long as weights at the beginning of each month are not outliers, I do not expect this to be very problematic. As an alternative, unbalanced months can be handled effectively using nonlinear DL specifications.

[16]<http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=MGFUPUS1&f=M>

[17]<http://research.stlouisfed.org/fred2/series/A229RX0>

[18]<http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm>

[19]<http://research.stlouisfed.org/fred2/series/CPIAUCSL>

tions at a higher temporal frequency by doing so. Even though q_t may be viewed as a result of aggregating daily quantities using equal weights, the fact that my daily price series omits weekends and holidays suggests that equal weights might not be optimal for daily prices. Consequently, the regressand aggregation scheme is unknown and flat sampling may not be the most efficient.

The sample runs from June 1986 through February 2012: $T = 309$ monthly observations. June 1986 is the earliest that I can obtain daily spot prices. Otherwise, mid-1986 is a convenient time to start the sample for two reasons. First, it is just after the oil price collapse early that year. Second, Hughes *et al.* (2008) argue that a structural break in the price elasticity occurred, and it appears from their graphical evidence that the break may have occurred just before 1986.

To estimate the four models, I construct a feasible CCR in multiple steps. I first regress out $\alpha' d_t$ from all other variables in the model. I then construct a feasible CCR in the usual way based on these residuals. In order to calculate the long-run variances and covariances for the feasible CCR, I use a lag window of 12, but I obtained similar results using 8 or 16.

Table 1. Coefficient Estimates.

	Average		BOP		EOP		Daily	
	est	s.e.	est	s.e.	est	s.e.	est	s.e.
β	-0.0359	0.0039	-0.0364	0.0038	-0.0345	0.0038	-0.0352	0.0041
γ	0.7381	0.0123	0.7398	0.0124	0.7361	0.0124	0.7386	0.0129

Table 1 shows the coefficient estimates and standard errors using CCR. The estimators are all superconsistent and they are all asymptotically normal. They differ only in terms of asymptotic efficiency. As a result, the robustness of estimates of β should not be surprising. The short-run price elasticity results are also similar to those obtained by Hughes *et al.* (2008) for their later sample: -0.042 for 2001-2006.

To examine efficiency of the different estimators of the cointegrating vector, I test for cointegration using estimates generated by the different estimators. With no persistent supply disruptions in the gasoline market over the period and similarly to the results of Hughes *et al.* (2008), I expect a long-run equilibrium – i.e., cointegration – in this market. If the series are I(1) and cointegrated, then a more precise estimate of the cointegrating vector will better detect the cointegration, because even a slightly different linear combination of the I(1) variables will not be cointegrated.

Table 2. Residual-based Cointegration Tests.

	Average	BOP	EOP	Daily
$\hat{Z}_\alpha$	-165.18	-167.11	-163.72	-206.32
$\hat{Z}_t$	-3.50	-3.50	-3.52	-3.56

Table 2 shows the results of Phillips-Ouliaris (1990) coefficient tests and t-tests. With a 5% critical value of -26.09, all four estimators produce coefficient test statistics that reject the null of no cointegration. The t-tests are more marginal, with all four statistics falling

between the 5% and 10% critical values of -3.77 and -3.45. The least restrictive (mixed-frequency) model provides the strongest evidence for cointegration using both tests. This result suggests that when the regressand aggregation weights are *unknown*, they should be estimated when feasible.

6   Concluding Remarks

Temporal aggregation and observation of series at disparate frequencies are nearly unavoidable problems in empirical analysis of time series. Since cointegrating regressions are known to retain their long-run properties in the wake of aggregation, it might be tempting to aggregate all series to the lowest frequency and settle for conventional estimation methods. When average sampling of both regressand and regressors is *not* possible, these estimators are inefficient compared to the full-information high-frequency benchmark.

The aggregation-conditional bound defined in this paper provides a convenient metric for evaluating efficiency of cointegrating regressions in mixed-frequency cases. Moreover, the proposed adjustment to conventionally efficient estimators attains this bound asymptotically, allowing more powerful inference on the long-run relationship in a cointegrating regression. When the regressand aggregation scheme is *known* but such an adjustment is infeasible, I recommend matching the regressor aggregation scheme to that of the regressand, in order to increase power and efficiency. When the regressand aggregation scheme is *unknown*, I recommend leaving the regressor disaggregated and estimating a distributed lag model when feasible.

Appendix: Proofs of the Asymptotic Results

Proof of Lemma 1.   Since the ratio $(mT/M)^{-1/2}$ has a unit limit as $T \to \infty$, the normalization in the lemma is asymptotically equivalent to that in Assumption [A2]. Now, since

$$\sum\nolimits_{i=1}^{[Mr]} b_{i/m} = \sum\nolimits_{t=1}^{[Tr]} \sum\nolimits_{i=0}^{m-1} b_{t-i/m} + \sum\nolimits_{i=[Tr]m+1}^{[Mr]} b_{i/m}$$

the desired result follows if $(mT)^{-1/2} \sum_{i=[Tr]m+1}^{[Mr]} b_{i/m} = o_p(1)$, which requires only that $[Mr] - [Tr]m < \infty$. Using the definition of T and the properties of the floor operator, $[Mr] - [Tr]m < Mr - ((M/m - 1)r - 1)m$, which simplifies to $(r+1)m$ and is finite. $\square$

Proof of Lemma 3.   Defining

$$M_T \equiv \sum\nolimits_t (x_t - \Upsilon_T D z_t)(x_t - \Upsilon_T D z_t)' \quad \text{and}$$
$$N_T \equiv \sum\nolimits_t (x_t - \Upsilon_T D z_t)(u_t^\varpi + \beta'(\Upsilon_T - \Pi) D z_t),$$

the least squares estimator is such that $\hat{\beta}_{LS} - \beta = M_T^{-1} N_T$. By assumption, Υ_T is either nonstochastic at Υ^* or ergodic with finite limit Υ^*, so that $M_T = \sum_t x_t x_t' + O_p(T)$ and

$$N_t = \sum\nolimits_t x_t (u_t^\varpi + z_t' \psi^*) - \Upsilon^* D \sum\nolimits_t z_t (u_t^\varpi + z_t' \psi^*) + o_p(T)$$

by replacing Υ_T with its probability limit. The limit of $T^{-2}M_T$ comes from the invariance principle in Lemma 2. The limit of the first term in $T^{-1}N_T$ follows along the lines of Lemma 3.1 of Park (1992) and using the invariance principle in Lemma 2. That of the second term follows from a law of large numbers for stationary processes. $\qquad \square$

Proof of Proposition 4. The second inequality in part [a] of the lemma requires

$$
\theta_{uz\cdot x}^{\varpi}\Theta_{zz\cdot x}^{-1}\theta_{zu\cdot x}^{\varpi} = (\theta_{uz}^{\varpi} - m^{-1}(\iota_{m-1}' \otimes \omega_{ux}))(\Theta_{zz} - m^{-1}(\iota_{m-1}\iota_{m-1}' \otimes \Omega_{xx}))^{-1}
$$
$$
\times\,(\theta_{zu}^{\varpi} - m^{-1}(\iota_{m-1} \otimes \omega_{xu}))
$$

to be positive semidefinite. Define $P_\iota \equiv \iota_m(\iota_m'\iota_m)^{-1}\iota_m'$ and E to be an $m \times (m-1)$ matrix with identity in the first $m-1$ rows and columns and zeros in the remaining row. Some algebra reveals that

$$
\Theta_{zz} - m^{-1}(\iota_{m-1}\iota_{m-1}' \otimes \Omega_{xx}) = (E' \otimes I)(\Xi_{xx} - P_\iota \otimes \Omega_{xx})(E \otimes I)
$$
$$
= (E'(I - P_\iota) \otimes I)\Xi_{xx}((I - P_\iota)E \otimes I) \equiv \tilde{\Xi}_{xx}
$$

using the fact that $\Xi_{xx}(P_\iota \otimes I) = (P_\iota \otimes I)\Xi_{xx} = P_\iota \otimes \Omega_{xx}$ from (9). $\tilde{\Xi}_{xx} > 0$ since $\Xi_{xx} > 0$ and $(I - P_\iota)E$ has full column rank of $m - 1 < m$.[20]

Parts [b] and [c] require that the second inequality collapses to an equality. Defining $\varpi = (\varpi_1, ..., \varpi_m)'$, allows

$$
\theta_{zu}^{\varpi} - m^{-1}(\iota_{m-1} \otimes \omega_{xu}) = (E' \otimes I)\Xi_{xu}\varpi - (E' \otimes I)\Xi_{xu}P_\iota\varpi
$$
$$
= (E'(I - P_\iota) \otimes I)\Xi_{xu}(I - P_\iota)\varpi \equiv \tilde{\Xi}_{xu}\varpi
$$

using the fact that $\Xi_{xu}P_\iota = (P_\iota \otimes I)\Xi_{xu} = P_\iota \otimes \omega_{xu}$. $\tilde{\Xi}_{xu}\varpi$ is a vector of zeros if either (i) $(I - P_\iota)\varpi = 0$, (ii) $(I - P_\iota)\Xi_{ux} = 0$, or (iii) $(E'(I - P_\iota) \otimes I)\Xi_{xu} = 0$. The first is satisfied if there is no variation in the weights: all weights equal $1/m$. The second and third are satisfied if $\Xi_{xu} = 0$ – i.e., if the regressors are strictly exogenous. The second is also satisfied if there is no variation across the rows of Ξ_{ux}, even if the elements of these rows are not zeros. (The assumption that $\Xi_{xx} > 0$ rules out that the third may also be satisfied by a similar condition on the columns.)

The first inequality in part [a] holds if $\theta_{uu\cdot x}^{\varpi} - \theta_{uz\cdot x}^{\varpi}\Theta_{zz\cdot x}^{-1}\theta_{zu\cdot x}^{\varpi} - \theta_{uu\cdot x} \geq 0$. This difference may be rewritten as

$$
\varpi'(\tilde{\Xi}_{uu} - \tilde{\Xi}_{ux}\tilde{\Xi}_{xx}^{-1}\tilde{\Xi}_{xu})\varpi
$$

with $\tilde{\Xi}_{uu} \equiv (I - P_\iota)\Xi_{uu}(I - P_\iota)$ and $\tilde{\Xi}_{ux} \equiv \tilde{\Xi}_{xu}'$, using the fact that $\Xi_{uu}P_\iota = P_\iota\Xi_{uu} = P_\iota \otimes \omega_{uu}$. Letting

$$
G' \equiv ((I - P_\iota), -\tilde{\Xi}_{ux}\tilde{\Xi}_{xx}^{-1}(E'(I - P_\iota) \otimes I))
$$

allows $\varpi'(\tilde{\Xi}_{uu} - \tilde{\Xi}_{ux}\tilde{\Xi}_{xx}^{-1}\tilde{\Xi}_{xu})\varpi = \varpi'G'\Xi G\varpi$. Finally, $\varpi'G'\Xi G\varpi \geq 0$ since $\Xi \geq 0$. This completes the proof for parts [a] and [b].

Finally, ruling out the possibility that $\tilde{\Xi}_{xx}$ has reduced rank, $\varpi'G'\Xi G\varpi = 0$ holds if either $(I - P_\iota)\varpi = 0$ or $(I - P_\iota)\Xi_{ux} = (I - P_\iota)\Xi_{uu} = 0$. Unless the weights are constant, it

---

[20] The rank is deduced from the properties of idempotent matrices (e.g., Lütkepohl, 1996, pg. 138).

is insufficient that the regressors are strictly exogenous. There must not be any variation across the rows of Ξ_{ux} *and* Ξ_{uu}. Under these conditions, *all* of the inequalities become equalities, completing the proof of part [c]. $\qquad\square$

Proof of Theorem 5. The least squares estimator of $D'\hat{\Pi}'_{LS}\hat{\beta}$ is

$$D'\hat{\Pi}'_{LS}\hat{\beta} = -\left(\sum_t z_t z'_t\right)^{-1}\sum_t z_t(y_t^{\varpi} - x'_t\hat{\beta}),$$

given $\hat{\beta}$. Subtracting $D'\Pi'\hat{\beta}$ from both sides yields

$$D'(\hat{\Pi}_{LS} - \Pi)'\hat{\beta} = -D'\Pi'(\hat{\beta} - \beta) - \left(\sum_t z_t z'_t\right)^{-1}\sum_t z_t(u_t^{\varpi} - x'_t(\hat{\beta} - \beta))$$

from (7). By assumption, $(\hat{\beta} - \beta) = o_p(1)$, so the stated result follows from a standard law of large numbers for stationary processes applied to the factors $\sum_t z_t z_t$ and $\sum_t z_t u_t^{\varpi}$. $\qquad\square$

Proof of Theorem 6. Similarly to the proof of Lemma 3, define $M_T^* \equiv \sum_t x_t^{\varpi*} x_t^{\varpi*\prime}$ and $N_T^* \equiv \sum_t x_t^{\varpi*} u_t^{\varpi*}$ so that $\hat{\beta}_{CCR} - \beta = (M_T^*)^{-1}N_T^*$. That the limit of $T^{-2}M_T^*$ coincides with that of $T^{-2}M_T$ in Lemma 3 follows from the stationarity of b_t^{ϖ}. $T^{-1}N_T^*$ expands to

$$T^{-1}\sum_t x_t b_t^{\varpi\prime}\kappa_{\psi\min} - \Lambda'_{\bullet x}\Sigma^{-1}T^{-1}\sum_t b_t^{\varpi}b_t^{\varpi\prime}\kappa_{\psi\min},$$

and standard asymptotics for integrated series show that the limit is $\int B_x dB'\kappa_{\psi\min}$, since the $\Lambda'_{\bullet x}\kappa_{\psi\min}$ from the limit in first term cancels exactly with the limit of the second. By construction, the conditional variance of the estimator attains the bound, since the long-run variance of $\kappa'_{\psi\min}b_t^{\varpi}$ is $\min_\psi \kappa'_\psi\Theta\kappa_\psi$. $\qquad\square$

References

Andreou, E., E. Ghysels, and A. Kourtellos (2010). "Regression Models with Mixed Sampling Frequencies," *Journal of Econometrics* 158, 246-261.

Chambers, M.J. (2003). "The Asymptotic Efficiency of Cointegration Estimators under Temporal Aggregation," *Econometric Theory* 19, 49-77.

Chambers, M.J. (2010). "Cointegration and Sampling Frequency," *Econometrics Journal*, forthcoming.

Chambers, M.J. and J.R. McCrorie (2007). "Frequency Domain Estimation of Temporally Aggregated Gaussian Cointegrated Systems," *Journal of Econometrics* 136, 1-29.

Dahl, C. and T. Sterner (1991). "Analyzing Gasoline Demand Elasticities: A Survey," *Energy Economics* 3, 203-210.

Davidson, J.E.H. (1994). *Stochastic Limit Theory: An Introduction For Econometricians.* Oxford: Oxford University Press.

Engle, R.F. and C.W.J. Granger (1987). "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55, 251-76.

Ghysels, E. and J.I. Miller (2012). "Mixed Frequency Data and Testing for Cointegration," forthcoming.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). "The MIDAS Touch: Mixed Data Sampling Regression Models," CIRANO Working Paper 2004s-20.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). "Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies," *Journal of Econometrics* 131, 59-95.

Gomez, V. and A. Maravall (1994). "Estimation, Prediction, and Interpolation for Non-stationary Series with the Kalman Filter," *Journal of the American Statistical Association* 89, 611-24.

Granger, C.W.J. (1990). "Aggregation of time-series variables: A survey." In: T. Barker and M.H. Pesaran (eds.), *Disaggregation in Econometric Modelling*. New York: Routledge, pp. 17-34.

Granger, C.W.J. and P.R. Siklos (1995), "Systematic Sampling, Temporal Aggregation, Seasonal Adjustment, and Cointegration: Theory and Evidence," *Journal of Econometrics*, 66, 357-369.

Haug, A.A. (2002). "Temporal Aggregation and the Power of Cointegration Tests: a Monte Carlo Study," *Oxford Bulletin of Economics and Statistics* 64, 399-412.

Hughes, J.E., C.R. Knittel, and D. Sperling (2008). "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand," *The Energy Journal* 29, 113-134.

Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* 12, 231-254.

Jordá, O. (1999). "Random-Time Aggregation in Partial Adjustment Models," *Journal of Business and Economic Statistics* 17, 382-395.

Kurozumi, E., and K. Hayakawa (2009). "Asymptotic Properties of the Efficient Estimators for Cointegrating Regression Models with Serially Dependent Errors," *Journal of Econometrics* 149, 118-35.

Lütkepohl, H. (1996). *Handbook of Matrices*. New York: John Wiley & Sons.

Marcellino, M. (1999). "Some Consequences of Temporal Aggregation in Empirical Analysis," *Journal of Business and Economic Statistics* 17, 129-136.

Miller, J.I. (2010). "Cointegrating Regressions with Messy Regressors and an Application to Mixed-frequency Series," *Journal of Time Series Analysis* 31, 255-277.

Miller, J.I. (2011). "Cointegrating MiDaS Regressions and a MiDaS Test," Department of Economics, University of Missouri, mimeo.

Park, J.Y. (1992). "Canonical Cointegrating Regressions," *Econometrica* 60, 119-43.

Pesaran, M.H. and Y. Shin (1998). "An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis." In: S. Strøm (ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium.* Cambridge: Cambridge University Press, pp. 371-413.

Phillips, P.C.B. (1991). "Optimal Inference in Cointegrated Systems," *Econometrica* 59, 283-306.

Phillips, P.C.B. and B.E. Hansen (1990). "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *The Review of Economic Studies* 57, 99-125.

Phillips, P.C.B. and S. Ouliaris (1990). "Asymptotic Properties of Residual-Based Tests for Cointegration," *Econometrica* 58, 165-93.

Pons, G. and A. Sansó (2005). "Estimation of Cointegrating Vectors with Time Series Measured at Different Periodicity," *Econometric Theory* 21, 735-56.

Rajaguru, R. and T. Abeysinghe (2008). "Temporal Aggregation, Cointegration and Causality Inference," *Economics Letters* 101, 223-226.

Saikkonen, P. (1991). "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory* 7, 1-21.

Schmalensee, R. and T.M. Stoker (1999). "Household Gasoline Demand in the United States," *Econometrica* 67, 645-662.

Seong, B., S.K. Ahn, and P.A. Zadrozny (2007). "Cointegration Analysis with Mixed-Frequency Data," CESifo Working Paper No. 1939.

Silk, J.I. and F.L. Joutz (1997). "Short and Long-Run Elasticities in U.S. Residential Electricity Demand: A Co-Integration Approach," *Energy Economics* 19, 493-513.

Small, K.A. and K. Van Dender (2007). "Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect," *The Energy Journal* 28, 25-51.

Stock, J.H. and M.W. Watson (1993). "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica* 61, 783-820.