DISCUSSION PAPER SERIES

No. 9654

TESTING FOR COINTEGRATION WITH TEMPORALLY AGGREGATED AND MIXED-FREQUENCY TIME SERIES

Eric Ghysels and J. Isaac Miller

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP9654.php

ISSN 0265-8003

TESTING FOR COINTEGRATION WITH TEMPORALLY AGGREGATED AND MIXED-FREQUENCY TIME SERIES

Eric Ghysels, University of North Carolina and CEPR
J. Isaac Miller, University of Missouri

Discussion Paper No. 9654
September 2013

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9654

September 2013

ABSTRACT

Testing for Cointegration with Temporally Aggregated and Mixed-frequency Time Series*

We examine the effects of mixed sampling frequencies and temporal aggregation on standard tests for cointegration. We find that the effects of aggregation on the size of the tests may be severe. Matching sampling schemes of all series generally reduces size, and the nominal size is obtained when all series are skip sampled in the same way. When matching all schemes is not feasible, but when some high-frequency data are available, we show how to use mixed-frequency models to improve the size distortion of the tests. We test stock prices and dividends for cointegration as an empirical demonstration.

JEL Classification: C12 and C32
Keywords: cointegration, mixed sampling frequencies, residual-based cointegration test, temporal aggregation and trace test

Eric Ghysels
Department of Economics
University of North Carolina
Gardner Hall CB 3305
Chapel Hill
NC 27599-3305
USA
Email: eghysels@unc.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=135789

J. Isaac Miller
Department of Economics
University of Missouri
118 Professional Building
Columbia, Missouri 65211
USA

Email: millerjisaac@missouri.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=177553

*The authors are grateful to participants of a research colloquium at the University of Missouri for useful comments. The first author acknowledges support of a Marie Curie FP7-PEOPLE-2010-IIF grant.

Submitted 07 September 2013

1 Introduction

Economic data is sampled at different frequencies, mostly because the cost of collecting or measuring variables can vary considerably. Price (indices) are relatively easy to collect. The most extreme example is the price of financial assets, such as stocks, commodities, etc. Such series are in principle available on a trade-by-trade basis for exchange traded assets. At the other end of the spectrum are demographic data, collected every 10 years via a Census count. Most key macroeconomic variables are collected on a monthly or quarterly basis. An additional complication is that some series are point sampled, such as prices, whereas others are flows, such as the gross domestic product (GDP) which is measured quarterly.

Faced with such data, a typical strategy is to collect same-frequency series, and for most economic relationships of interest a mixture of stock and flow variables are considered. For example, if we were to study the relationship between prices and output (the latter measured via GDP) across different countries we would end up with quarterly CPI (a stock variable) and quarterly GDP (a flow variable) for each country. Note that in this case, CPI data are available monthly but are aligned with GDP observations.

In this paper we show that, even though aggregation does not change the cointegrating vector, aggregation can cause size distortion in cointegration tests.[1] Several cases need to be considered and depending on the case, size distortions can either be absent, mild or severe. Consider a first example of inflation sampled quarterly in several countries and we are interested in cointegration between prices. To be more specific, in the paper we consider Johansen's (1988) likelihood-based trace test (when in general more than one cointegrating relationship is allowed) and Engle and Granger's (1987) residual-based tests, or the modified tests of Phillips and Ouliaris (1990) (when no more than one cointegrating relationship

[1] In contrast, Hooker (1993), Hu (1996), and Haug (2002) addressed power of cointegration tests with aggregated data using simulations. These authors focused on varying the frequency of the series, while keeping the aggregation scheme fixed: Hooker (1993) examined skip sampling, Haug (2002) examined flat sampling, and Hu (1996) examined each separately. In contrast, we focus on the more fundamental problem of size distortion under varying aggregation schemes, which would be of practical use when faced with either aggregated data or data observed at different frequencies.

allowed). The sampling scheme involved in this first example is the same, i.e. both series are skip-sampled at a (low) quarterly frequency. This is a case where all series are skip-sampled. We show that in such cases when all series skip-sampled in the *same* way, then there will be no size distortions. Now consider a second case, where cointegration between GDP from the different countries is of interest. In this case we expect size distortions, although one can characterize an upper bound on the distortions and they are still acceptable. Third, consider cointegration within a given country between output and prices. Here the series are aggregated/sampled differently – namely flow versus stocks with the latter being available at higher frequency. We show that in such cases size distortion can be quite severe. To complicate matters further, suppose statistical agencies across countries skip-sample prices but do so *differently*. In such cases we obtain the most severe size distortions.

The scope of our paper goes beyond characterizing size distortions. We also propose novel ways to solve the size distortion problems. Recall that price and GDP series are actually available at different frequencies: prices are recorded monthly and GDP quarterly. We will take mixed sampling frequencies to our advantage to address size distortions. Namely, instead of running a low-frequency (henceforth LF) trace test, we propose to keep the high frequency (henceforth HF) and run a mixed-frequency (henceforth MF) trace test. The former will have size distortions, depending on the case as discussed above, while the latter will not have size distortions. How do we run a MF trace test? We rely on MF vector autoregressive (henceforth VAR) models to implement the new class of tests. VAR models for MF data were independently introduced by Anderson *et al.* (2012), Ghysels (2012) and McCracken *et al.* (2013).[2] An extension of these models to cointegrated series has recently been considered by Götz *et al.* (2012). An early example of related ideas appears in Friedman (1962). Foroni *et al.* (2013) provide a survey of mixed frequency VAR models and

[2] State space models provide a common alternative method for handling possibly nonstationary series observed at different frequencies by treating the low-frequency series as containing missing observations (see Jones (1980), Ansley and Kohn (1983), Harvey and Pierse (1984), Zadrozny (1988, 1990), Gomez and Maravall, 1994, Mariano and Murasawa, 2003, 2010, *inter alia*, among others). Seong *et al.* (2013) analyzed a cointegrated VAR in this context.

related literature. In particular, MF VAR models can be viewed as a multivariate extension of MIDAS regressions.[3]

The connection with MIDAS regressions also leads us to residual-based tests for cointegration. Miller (2013) studies CoMIDAS regressions involving I(1) processes. We therefore also propose MF residual-based tests, as opposed to LF residual-based tests involving aggregate HF data. Indeed, the latter also feature size distortions, as above (despite given results on efficiency – see Chambers (2003), *inter alia*). We show that CoMIDAS – while featuring some size distortions – works surprisingly well.

The rest of the paper is organized as follows. In Section 2, we present a short and less technical explanation for the size distortions driving our main results and detailed later in the paper. We review popular cointegration testing procedures in the context of a possibly infeasible HF data-generating process in Section 3. We then introduce the LF and MF environments in Section 4, reassessing testing options and asymptotic null distributions in these contexts. A detailed discussion of size distortion in the case of equal weighting schemes is contained in Section 5. Section 6 contains simulation results, Section 7 contains an empirical application to stock prices and dividends, and Section 8 concludes. The paper contains two appendices: Appendix A contains the proofs of the main theoretical results and two ancillary lemmas, and Appendix B extends the discussion of size distortion of Section 5 to additional cases.

We make use of the following notational conventions throughout the paper. $C \oplus D$ (direct sum) creates a block diagonal matrix with diagonal blocks given by C and D. $C \otimes D$ is the usual Kronecker product. We use ι to denote a unit vector of length given by the context in which it is used.

[3] MIDAS, meaning Mi(xed) Da(ta) S(ampling), regression models have been put forward in recent work by Ghysels *et al.* (2004, 2006) and Andreou *et al.* (2010). See Andreou *et al.* (2011) and Armesto *et al.* (2010) for surveys.

2 On the Genesis of Size Distortion

Consider a p-variate I(1) series (y_t) with $t = 1, ..., T$. Leaving aside the possibility of deterministic trends, it is common to write $A'y_t = e_t$, with an r-variate I(0) series (e_t), when such series are cointegrated by $r \leq p$ linearly independent cointegrating vectors given by the r columns of A. In the context of single-equation Engle-Granger cointegrating regressions, where a single cointegrating relationship α is assumed, $\alpha' y_t = e_t$. The most commonly used cointegration tests are Johansen's (1988) likelihood-based trace test for the first case (more than one cointegrating relationship allowed) and Engle and Granger's (1987) residual-based tests, or the modified tests of Phillips and Ouliaris (1990), for the second case (no more than one cointegrating relationship allowed).

Although these test statistics and their limit distributions under the null are quite different, they fundamentally depend on the sample moment $\sum_t y_{t-1} \triangle y_t'$. Size distortion may be traced to deviations of the limiting distribution of this sample moment from what we expect under the respective nulls. Suppose that the data-generating process (DGP) occurs at a higher frequency: $m < \infty$ times more often.[4] The series may be rewritten as $(y_{t-i/m}^{(m)})$ with $i = 0, ..., m-1$, and it will be convenient to let $M = mT$ denote the HF sample size. Stock (1987) and Granger (1990) noted that temporal aggregation and sampling frequency do not affect cointegrating vectors, so $A' y_{t-i/m}^{(m)} = e_{t-i/m}^{(m)}$ is still I(0). Therefore, the sampling frequency does not alter the nulls and alternatives of these tests.

Suppose that the HF series $(y_{t-i/m}^{(m)})$ is subject to end-of-period sampling, so that only $i = 0$ is observed and we observe only a LF series $(y_t^{(m)})$. Because $\triangle y_t^{(m)} = \sum_{i=0}^{m-1} \triangle^{(1/m)} y_{t-i/m}^{(m)}$, where $\triangle^{(1/m)}$ is defined to be the HF difference operator, some algebra shows that

$$T^{-1} \sum_t y_{t-1}^{(m)} \triangle y_t^{(m)\prime} = m \left[M^{-1} \sum_{i=0}^{m-1} \sum_t y_{t-1/m}^{(m)} \triangle^{(1/m)} y_t^{(m)\prime} \right] + o_p(1) \tag{1}$$

holds as T increases. Under standard assumptions, the square-bracketed factor has a limit-

[4] See Chambers (2011) for analysis of single-equation cointegration models as $m \to \infty$.

ing distribution that coincides with the well-known limiting distribution of $T^{-1}\sum_t y_{t-1}\triangle y_t'$ with (y_t) generated at the low frequency. Thus, the limit of the sample moment using the observed LF series $(y_t^{(m)})$ differs from that using the infeasible high-frequency series $(y_{t-i/m}^{(m)})$ by a factor of m. Because the test statistics are standardized, they are robust to such scalar multiples by construction. Test size is therefore *not* distorted by end-of-period sampling.[5]

Size distortion arises when we do not observe the HF series $(y_{t-i/m}^{(m)})$, or even the LF series $(y_t^{(m)})$, but some LF series (z_t^a) resulting from temporally aggregating $(y_{t-i/m}^{(m)})$ by some other method. Denoting the aggregation weights by $\varpi_{s,j+1}$ for $j = 0, ..., m-1$ and $s = 1, ..., p$ and letting $\Pi_j \equiv (\varpi_{1j} \oplus \cdots \oplus \varpi_{pj})$ represent a diagonal matrix of the weights for all series in the vector $(y_{t-j/m}^{(m)})$, the aggregated series may be written as $z_t^a = \sum_{j=0}^{m-1} \Pi_{j+1} y_{t-j/m}^{(m)}$. Its first difference is

$$\triangle z_t^a = \sum\nolimits_{i,j=0}^{m-1} \Pi_{j+1} \triangle^{(1/m)} y_{t-(i+j)/m}^{(m)} \neq \sum\nolimits_{i=0}^{m-1} \triangle^{(1/m)} y_{t-i/m}^{(m)} = \triangle y_t \qquad (2)$$

in general. The inequality in (2) becomes an equality when $\Pi_j = I$ for $j = 0$ and 0 otherwise, which is exactly the end-of-period sampling scheme just discussed.

The inequality in (2) more generally causes size distortion because, with $T^{-1}\sum_t z_{t-1}^a \triangle z_t^a$ on the left-hand side, the square-bracketed factor on the right-hand side of (1) cannot be isolated without leaving non-negligible terms.

We consider size distortion from two causes of the inequality in (2). First, the inequality holds when an aggregation scheme *other* than end-of-period sampling is used on *all* p series. Second, the inequality holds when not all series are aggregated in the same way. We refer to the size distortion from these two cases as type A and type B size distortion, respectively. As it turns out, type A size distortion may be zero asymptotically and has an upper bound within an acceptable range. Type B size distortion may be quite severe.

[5]Shiller and Perron (1985) and Perron (1991) noted that the powers of univariate unit root tests are not affected by the frequency. Hooker's (1993) and Haug's (2002) simulation results suggest that observing data at a higher frequency can increase the power of residual-based cointegration tests, but Hu's (1996) results suggest that span matters much more than frequency, as in the univariate case.

3 High-frequency DGP and Cointegration Tests

In this section, we describe the HF DGP, assumed to be infeasible given the data, but against which we can compare the feasible models. We then review the standard cointegration tests and their asymptotics. All of the results and discussions in this section are well known, with the only complication being the index $t-i/m$ accounting for the different frequencies. This section thus provides a review of the relevant techniques and introduces essential notation.

Consider a p-variate HF DGP given by

$$\triangle^{(1/m)} y_{t-i/m} = \Gamma A' y_{t-(i+1)/m} + \varepsilon_{t-i/m}, \tag{3}$$

where $\triangle^{(1/m)}$ is the HF difference operator described above and $i = 0, ..., m-1$. (We now and henceforth suppress the superscript (m) on $y^{(m)}_{t-i/m}$ employed in the previous section.) The cointegrating rank is r, so that Γ and A are both $p \times r$ matrices. As usual, $0 \leq r \leq p$ and the series are cointegrated if $0 < r < p$.

We assume an invariance principle of the form $M^{-1} \sum_{j=1}^{[rM]} \varepsilon_{j/m} \to_d B(r)$ as $M \to \infty$, where B is a vector Brownian motion with variance Σ. Thus, $B = \Sigma^{1/2} W$, where W is a vector of independent standard Brownian motions. It follows that $(mT)^{-1} \sum_{t=1}^{[rT]} \sum_{i=0}^{m-1} \varepsilon_{t-i/m}$ $\to_d mB(r)$ as $T \to \infty$ and for finite m, which provides a translation of the parameters of the LF asymptotics utilized below into those of the infeasible HF DGP. (See Miller, 2011, for more details.)

Serial correlation is typically allowed by including lags of $\triangle^{(1/m)} y_{t-i/m}$, in which case $\triangle^{(1/m)} y_{t-i/m}$ and $y_{t-(i+1)/m}$ in (3) may be replaced by residuals from regressing these onto the lagged differences. In that case, the variances of $\triangle^{(1/m)} y_{t-i/m}$, etc., are conditional, but the results do not otherwise change. For expositional simplicity, we assume no serial correlation in $(\varepsilon_{t-i/m})$ and let $var(\varepsilon_{t-i/m}) \equiv \Sigma$.

The analysis of a model with deterministic trends would be more involved. As usual, these trends would affect the limiting distributions. In order to avoid overly complicating the presentation of our results and because deterministic trends may be "sampled" at any frequency, we assume that $y_0 = 0$ and do not explicitly consider deterministic trends. We expect that the main intuitions about size distortion gleaned from the simpler model would hold for a more complicated model.

System of Regression Equations. For notational ease, authors typically define $s_{0,t-i/m} \equiv \triangle^{(1/m)} y_{t-i/m}$ and $s_{1,t-i/m} \equiv y_{t-(i+1)/m}$, with the idea that these may be redefined as regression residuals if serial correlation is allowed. The sample moments $S_{gh} \equiv M^{-1} \sum_t \sum_{i=0}^{m-1} s_{g,t-i/m} s'_{h,t-i/m}$ for $g, h = 0, 1$ are in turn defined by these random variables.

Johansen's (1988) trace test allows for the possibility of more than one cointegrating relationship. For known A, the maximum likelihood estimators $\hat{\Gamma}$ and $\hat{\Sigma}$ of Γ and Σ are $\hat{\Gamma}(A) = S_{01} A \left(A' S_{11} A\right)^{-1}$ and $\hat{\Sigma}\left(A\right) = S_{00} - S_{01} A \left(A' S_{11} A\right)^{-1} A' S_{10}$. The likelihood function may be concentrated so that the maximal value (up to an irrelevant constant) is $L_{\max}$ in

$$
\begin{aligned}
L_{\max}^{-2/M} &= |S_{00} - S_{01} A \left(A' S_{11} A\right)^{-1} A' S_{10}| \\
&= |S_{00}||A'(S_{11} - S_{10} S_{00}^{-1} S_{01}) A| / |A' S_{11} A|, \qquad (4)
\end{aligned}
$$

and A is chosen to minimize $L^{-2/M}$. Specifically, A is estimated by finding the r largest eigenvalues of $|S_{11} - S_{10} S_{00}^{-1} S_{01}|$ subject to $|A' S_{11} A| = I$. The p ordered eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ are the same as those obtained by solving the determinantal equation $|\lambda I - S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}| = 0$ to implement the test.

The null of the well-known trace test is $H_0 : r = r_0$, and the alternative is $H_A : r = p$ (all series I(0), no cointegration). The second determinant in (4) equals the product of $(1 - \lambda_i)$ corresponding to the first r_0 eigenvalues under the null and equals the product corresponding to all p eigenvalues under the alternative. The familiar trace test is a likelihood ratio test therefore given by

$$
-2 \log Q(H_{r_0|p}) = -M \sum_{i=r_0+1}^{p} \log(1 - \hat{\lambda}_i),
$$

since the common factor $|S_{00}|$ cancels. Using properties of the log and of the trace,

$$-2\log Q(H_{r_0|p}) = M\ \mathrm{tr}\{S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}\} + o_p(1)$$

as $M \to \infty$. See Johansen (1995) for a very detailed discussion. Further references to the trace test may be understood to refer either to $-2\log Q(H_{r_0|p})$ or to its asymptotic approximation $M\ \mathrm{tr}\{S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}\}$.

Following the lead of Cheung and Lai (1993), Horvath and Watson (1995), and other authors, we consider the null $H_0 : r = 0$ against the full rank alternative.[6] The asymptotic distribution of the test employs the limits $S_{00} \to_p \Sigma$, $S_{10} \to_d \int BdB'$, and $S_{11} \to_d \int BB'$. Canceling out the variance Σ, the test has a limiting null distribution given by

$$M\ \mathrm{tr}\{S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}\} \to_d \mathrm{tr}\{\int (dW)'W(\int WW')^{-1}\int WdW'\} \tag{5}$$

as $M \to \infty$. The number of stochastic trends under the null is the only nuisance parameter.

Single Regression Equation. Supposing that $r \leq 1$, the DGP becomes $\triangle^{(1/m)}y_{t-i/m} = \gamma\alpha' y_{t-(i+1)/m} + \varepsilon_{t-i/m}$, and a typical residual-based test for cointegration is simply a unit root test of the fitted residuals $\hat{e}_{t-i/m} = \hat{\alpha}' y_{t-i/m}$ with one element of α normalized to unity, so that $\alpha = (1, -\beta')'$. In the absence of serial correlation, the unit root test is just a regression of $\triangle^{(1/m)}\hat{e}_{t-i/m}$ onto $\hat{e}_{t-(i+1)/m}$ with a null that the coefficient is zero (no cointegration), along the lines of Engle and Granger (1987). The coefficient test and t-test may be written as

$$\rho_M = M(\hat{\alpha}' S_{11}\hat{\alpha})^{-1}\hat{\alpha}' S_{10}\hat{\alpha} \qquad \text{and} \qquad \tau_M = (\hat{\alpha}' S_{00}\hat{\alpha}(M^{-1}\hat{\alpha}' S_{11}\hat{\alpha}))^{-1/2}\hat{\alpha}' S_{10}\hat{\alpha},$$

[6] Extending to high-order nulls essentially requires redefining the rank of the limiting matrix inside of the trace, but substantially complicates the proofs. This null allows us to focus exposition more specifically but without any substantial loss of generality. Also, this null is analogous to the null in the single-equation case (no cointegration), even though the tests have different alternatives.

in this case.

Under the null, the rank of $\gamma\alpha'$ is zero and $\triangle^{(1/m)} y_{t-i/m} = \varepsilon_{t-i/m}$, so that $e_{t-i/m} = \alpha' \sum_{j=1}^{mt-i} \varepsilon_{j/m}$. Using the above invariance principle, we have $\hat{\beta} \rightarrow_d (\int B_2 B_2')^{-1} \int B_2 B_1$, where $B = (B_1, B_2')'$ is partitioned like α. Phillips and Ouliaris (1990) show that the coefficient test and t-test have limiting distributions given by

$$\rho_M \rightarrow_d \left(\int Q^2 \right)^{-1} \int Q dQ \equiv \rho \quad \text{and} \quad \tau_M \rightarrow_d \left(\kappa' \kappa \int Q^2 \right)^{-1/2} \int Q dQ \equiv \tau \tag{6}$$

where $Q(r) \equiv W_1(r) - \int W_1 W_2' (\int W_2 W_2')^{-1} W_2(r)$ and $\kappa' \equiv (1, -\int W_1 W_2' (\int W_2 W_2')^{-1})$ with $W = (W_1, W_2')'$ partitioned like B.

4 Temporal Aggregation and Mixed Sampling Frequencies

The main premise of this research is that temporal aggregation of at least one series is unavoidable. From the analyst's point of view, some or all of the data have already been aggregated. With the well-known results from the previous section in hand, we now turn to the task of introducing temporal aggregation of one or more of the component series in $(y_{t-i/m})$.

The ensuing analysis will employ LF asymptotics (as $T \rightarrow \infty$) rather than the HF asymptotics of the previous section. To this end, it will be useful to stack the HF series into a single LF vector. We sort by series and then by HF time period. In contrast, Ghysels (2012) sorts by time period and then by series, keeping the vector y_t intact. The order is not important for conducting the tests discussed in this paper, but only for the theoretical analysis.

Define $z_{st} \equiv (y_{st}, y_{s,t-1/m}, ..., y_{s,t-(m-1)/m})'$ and $u_{st} \equiv (\varepsilon_{st}, \varepsilon_{s,t-1/m}, ..., \varepsilon_{s,t-(m-1)/m})'$ for $s = 1, \ldots, p$, and define $z_t \equiv (z_{1t}', \ldots, z_{pt}')'$ and $u_t \equiv (u_{1t}', \ldots, u_{pt}')'$. Transposing and stacking the original DGP in (3) across m allows

$$(\triangle^{(1/m)}z_{1t},\ldots,\triangle^{(1/m)}z_{pt}) = (z_{1,t-1/m},\ldots,z_{p,t-1/m})A\Gamma' + (u_{1t},\ldots,u_{pt}),$$

and vectorizing both sides allows

$$\triangle^{(1/m)}z_t = (\Gamma A' \otimes I)z_{t-1/m} + u_t,$$

equivalently to (3). This expression represents a system of mp equations with an error variance given by $var(u_t) = \Sigma \otimes I$.

The VECM estimated by the analyst will need a LF difference rather than a HF difference, because HF differences are infeasible for any temporally aggregated series. In anticipation of the aggregation, a LF difference of the HF system may be rewritten as

$$\triangle z_t = (\Gamma A' \otimes I)z_{t-1} + \eta_t, \tag{7}$$

where $\eta_t \equiv (I + (\Gamma A' \otimes I))\sum_{i=1}^{m-1}\triangle^{(1/m)}z_{t-i/m} + u_t$. Even though we have assumed no serial correlation in the HF DGP, differencing the HF series at the low frequency in (7) creates first-order serial correlation of (η_t). This correlation results not from aggregation, but simply from the LF difference. Although Cheung and Lai (1993) and other authors have shown that the trace test may suffer size distortion in the presence of serial correlation, most of the serial correlation in (η_t) exists *within* but not between LF increments.

4.1 Low- and Mixed-frequency Models

The system in (7) is still infeasible, because for some or all of the p series the analyst observes $\varpi_k' z_{kt}$, where $\varpi_k \equiv (\varpi_{k1},...,\varpi_{km})'$ is a vector of m non-negative deterministic aggregation weights such that $\varpi_k'\iota = 1$ (the weights sum to unity) for all k. Some algebra allows $z_{kt} = \iota y_{kt} - H\triangle^{(1/m)} z_{kt}$, with H defined as

$$H \equiv \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ 1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 \end{bmatrix},$$

an $m \times m$ matrix. The decomposition makes clear what we intuitively know to be true: that an I(1) series observed every m time periods must be cointegrated with the same series observed every m periods with an overlapping interval. All m series share a single common stochastic trend (y_{kt}), and therefore have $m-1$ linearly independent cointegrating relationships. Further, the restriction on the weights allows a notationally efficient representation of the aggregated series: $\varpi_k' z_{kt} = y_{kt} - \varpi_k' H\triangle^{(1/m)} z_{kt}$.

System of Regression Equations. In the purely LF case, all of the series have been temporally aggregated. We define $\Pi_a \equiv (\varpi_1' \oplus \cdots \oplus \varpi_p')$ to be a $p \times mp$ *full aggregation matrix* (high frequency to low frequency) in order to operationalize this concept. We further define the temporally aggregated series (z_t^a) such that $z_t^a \equiv \Pi_a z_t$. That is, $z_{kt}^a \equiv \varpi_k' z_{kt}$ for each series $k = 1, ..., p$. Full aggregation transforms the system in (7) as

$$\triangle z_t^a = \Gamma A' z_{t-1}^a + \eta_t^a, \tag{8}$$

which is a feasible LF system. Note that under the null of no cointegrating vectors, the error is simply $\eta_t^a = \Pi_a \eta_t$.

The matrices Γ and A in the fully aggregated system in (8) are the same as those in the HF DGP. Temporal aggregation does not change the cointegrating relationships, as pointed out by Stock (1987) and Granger (1990), but it may substantially affect the short-run properties of the series (Marcellino, 1999, *inter alia*).

Moving to a system of equations with series observed at different sampling frequencies, we assume that p_l series in $(\triangle^{(1/m)} y_{t-i/m})$ are aggregated and observed at the low frequency, while p_h series are observed at the high frequency, with $p_l + p_h = p$. Without loss of generality, let the first p_l series in $(\triangle^{(1/m)} y_{t-i/m})$ be those observed only at the low frequency.

The matrix Π_m defined by

$$\Pi_m \equiv \begin{bmatrix} \varpi_1' \oplus \cdots \oplus \varpi_{p_l}' & 0 \\ 0 & I \end{bmatrix}$$

is a $(p_l + mp_h) \times mp$ *partial aggregation matrix* that operationalizes aggregation from the infeasible HF DGP to a feasible MF model. Premultiplying z_t by Π_m temporally aggregates the HF observations of the first p_l series in $(\triangle^{(1/m)} y_{t-i/m})$, but leaves the remaining p_h series. The resulting series (z_t^m), defined by $z_t^m \equiv \Pi_m z_t$, contains p_l LF series. The remaining mp_h series are created from p_h distinct series observed at m different LF intervals. For example, if the first p_l series are observed annually and only a single $(p_h = 1)$ monthly $(m = 12)$ HF series remains, the last 12 series are annual observations of that series observed at distinct months.

A MF system contains both HF series regressed on lags of LF series and LF series regressed on lags of HF series. Premultiplying both sides of the system in (7) by Π_m does *not* yield a VECM directly. Rather, it yields a system in which both HF and LF series are regressed on lags of *only* HF series. That is, $(\triangle z_t^m)$ is regressed on (z_{t-1}), which is still infeasible. To operationalize the VECM, (z_{t-1}^m) should be created using the same partial aggregation matrix as that for $(\triangle z_t^m)$.

Define the notation

$$\Gamma^m \equiv \begin{bmatrix} I & 0 \\ 0 & I \otimes \iota \end{bmatrix} \Gamma \qquad \text{and} \qquad A^m \equiv \begin{bmatrix} I & 0 \\ 0 & \varpi_{p_l+1}^* \oplus \cdots \oplus \varpi_p^* \end{bmatrix} A,$$

where the vectors $\varpi_{p_l+1}^*, ..., \varpi_p^*$ satisfy the properties of ϖ_k above. Partial aggregation results in

$$\triangle z_t^m = \Gamma^m A^{m\prime} z_{t-1}^m + \eta_t^m, \tag{9}$$

which is a feasible MF VECM. The error is simply $\eta_t^m = \Pi_m \eta_t$ under the null of no cointegrating vectors.

The aggregation weight vectors ϖ_k^* introduced here have no practical role, except to impose known cointegrating relationships. In the systems context, we may set them to be equal ($\varpi_k^* = \varpi^*$ for all k). In the HF system in (7), the coefficient matrix $(\Gamma A' \otimes I)$ is $mp \times mp$ with rank of mr. The fact that its rank must be a multiple of m has practical implications for testing, as it requires $(m-1)p$ restrictions to be imposed. Total aggregation of the HF DGP to a LF system imposes $(m-1)p$ restrictions, so no additional restrictions are necessary. Partial aggregation to obtain the MF system imposes $(m-1)p_l$ restrictions, but $(m-1)p_h$ remain to be imposed. We use ϖ_* to impose the remaining $(m-1)p_h$ restrictions on the MF system. The only practical implication is on the rank of A^m, affecting the number of eigenvalues $\hat{\lambda}_i$ in the test, as will be discussed below. The particular choice of ϖ_* is purely theoretical, as may be seen from the fact that $(I \otimes \varpi_*')(I \otimes \iota) = I$ for any choice of ϖ_*, as long as its elements sum to unity. Consequently, the set of eigenvalues of $\Gamma^m A^{m\prime}$ are exactly those of $\Gamma A'$, but with the addition of $(m-1)p_h$ zeros.

Since the eigenvalues are invariant with respect to the ordering of the series, it is worth repeating that the ordering that Ghysels (2012) uses for impulse responses is also valid for these tests.

Single Regression Equation. The fully aggregated LF Engle-Granger regression may be written as

$$z_{1t}^a = (z_{2t}^a, ..., z_{pt}^a)\beta + e_t^a, \tag{10}$$

where $z_{kt}^a = \varpi_k' z_{kt}$ is a scalar aggregate of the vector z_{kt}, as above. The Engle-Granger testing strategy is a unit root test of the fitted residuals $(\hat{e}_t^a)$. The exact structure of the error term e_t^a implied by aggregation is not needed for the analysis of the null of no cointegration.

As in the case of a system, a MF single-equation regression is more complicated than a LF regression. We again let $\alpha = (1, \beta')'$, so that the coefficient on one of the LF series is normalized to unity.[7] Under this convention, the cointegrating vector for the mixed-frequency model may be written as

$$\begin{aligned}\alpha^m &\equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & \varpi_{p_l+1} \oplus \cdots \oplus \varpi_p \end{bmatrix} \begin{bmatrix} 1 \\ -(\beta_2, ..., \beta_{p_l})' \\ -(\beta_{p_l+1}, ..., \beta_p)' \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -(\beta_2, ..., \beta_{p_l})' \\ -(\varpi^*_{p_l+1} \oplus \cdots \oplus \varpi^*_p)(\beta_{p_l+1}, ..., \beta_p)' \end{bmatrix}.\end{aligned}$$

The cointegrating regression to be tested may be expressed as $z_{1t}^a = (z_{2t}^a, ..., z_{p_l t}^a, \varpi'_{p_l+1} z_{p_l+1,t}, ..., \varpi'_p z_{pt})\beta + e_t^a$, where we write $\varpi'_k z_{kt}$ rather than z_{kt}^a to emphasize that the weights ϖ_k for $k = p_l + 1, ..., p$ may be estimated. For the test, we must impose the *known* cointegrating restrictions, which may be done by estimating

$$z_{1t}^a = (z_{2t}^a, ..., z_{p_l t}^a, y_{p_l+1,t}, ..., y_{pt})\beta - (0, ..., 0, \varpi'_{p_l+1} H \triangle^{(1/m)} z_{p_l+1,t}, ..., \varpi'_p H \triangle^{(1/m)} z_{pt})\beta + e_t^a, \tag{11}$$

In this expression, the error has the same structure as the fully aggregated equation in (10). However, the fitted residuals will be different from those in (10), because they contain additional errors from estimating the weights on the last p_h series. These errors do not collapse to zero under the null, leading to additional size distortion of the tests.

[7] From a theoretical perspective, there is no loss of generality. If the coefficient on a high-frequency series is instead normalized to unity, then the I(0) terms in $\iota y_{1t} - H\triangle^{(1/m)} z_{1t}$ may be moved to the right-hand side, so that their weights are estimated.

4.2 Testing in Low- and Mixed-frequency Environments

Facing the task of conducting a likelihood-based or residual-based cointegration test on series observed at different frequencies, the analyst may employ the MF system in (9) or the regression in (11). Π_m may be known to the analyst, but is assumed to be beyond the analyst's control, so that the MF models in (3) and (7) are infeasible. Alternatively, the analyst may aggregate the available HF series and simply employ the LF system in (8) or the regression in (10). The tests themselves do not need to be modified for aggregated series, but some modification is necessary for series observed at different frequencies.

System of Regression Equations. Although we do not explicitly consider a trace test based on the infeasible system in (7), it will be convenient to define some additional notation along these lines. Specifically, let $r_{0t} \equiv \triangle z_t$, $r_{1t} \equiv z_{t-1}$, and $R_{gh} \equiv T^{-1} \sum r_{gt} r'_{ht}$ for $g, h = 0, 1$. Note that the sample moments R_{gh} are LF averages, rather than the HF averages of S_{gh} defined above.

When all series are aggregated, the system contains the same number of series as in the HF DGP. No modifications to the procedure are necessary, except to accommodate the smaller sample size associated with the aggregated series, but the test will generally have a different limiting distribution. The trace test statistic is

$$-2 \log Q(H_{r_0|p}) = T \ \mathrm{tr}\{(R^a_{11})^{-1} R^a_{10} (R^a_{00})^{-1} R^a_{01}\} + o_p\left(1\right),$$

where $R^a_{gh} \equiv \Pi_a R_{gh} \Pi'_a$ for $g, h = 0, 1$.

When only some series are aggregated, the MF model contains $p_l + mp_h$ series, but $(m-1)p_h$ cointegrating relationships are prespecified. Thus, r is such that $(m-1)p_h \leq r \leq$

$p_l + mp_h$ and the null r_0 should be chosen accordingly. The trace test is therefore given by

$$-2\log Q(H_{r_0|p_l+mp_h}) = -T\sum\nolimits_{i=r_0+1}^{p_l+mp_h}\log(1-\hat{\lambda}_i)$$
$$= T\ \mathrm{tr}\{(R^m_{11})^{-1}R^m_{10}(R^m_{00})^{-1}R^m_{01}\} + o_p(1)$$

where $R^m_{gh} \equiv \Pi_m R_{gh}\Pi'_m$ for $g, h = 0, 1$ and $\hat{\lambda}_1, \ldots, \hat{\lambda}_{p_l+mp_h}$ solve the determinantal equation $|\lambda I - (R^m_{11})^{-1}R^m_{10}(R^m_{00})^{-1}R^m_{01}| = 0$.

For the limiting distributions of the trace test for the fully and partially aggregated systems, it will be useful to define $\Xi^{a,m}_{00} \equiv \Pi_{a,m}(\Sigma \otimes m^{-1}H_{00})\Pi'_{a,m}$, $\Xi^{a,m}_{11} \equiv \Pi_{a,m}(\Sigma^{1/2} \otimes I)(\int WW' \otimes \iota\iota')(\Sigma^{1/2\prime} \otimes I)\Pi'_{a,m}$, and $\Xi^{a,m}_{10} \equiv \Pi_{a,m}(\Sigma^{1/2} \otimes I)((\int WdW' \otimes \iota\iota') + (I \otimes m^{-1}H_{10}))(\Sigma^{1/2\prime} \otimes I)\Pi'_{a,m}$, with $H_{00} \equiv (I + H')(I + H) + HH'$, $H_{01} \equiv H(I + H)$, and $H_{10} = H'_{01}$. The limiting distributions are given in the following theorem.

Theorem 1. Consider the null hypothesis of no cointegrating relationships: $r_0 = 0$ in the LF system, $r_0 = (m-1)p_h$ in the MF system.

(a) The trace test statistic $T\ \mathrm{tr}\{(R^a_{11})^{-1}R^a_{10}(R^a_{00})^{-1}R^a_{01}\}$ based on the fully aggregated (LF) system in (8) has an asymptotic distribution coinciding with $\mathrm{tr}\{\Xi^{a\prime}_{10}(\Xi^a_{11})^{-1}\Xi^a_{10}(\Xi^a_{00})^{-1}\}$ as $T \to \infty$. The limiting distribution further simplifies to

$$\mathrm{tr}\{(\int(dW)'W + (Z^a_{10})')(\int WW')^{-1}(\int WdW' + Z^a_{10})(Z^a_{00})^{-1}\}$$

where $Z^a_{i0} \equiv \Sigma^{-1/2}\Pi_a(\Sigma \otimes m^{-1}H_{i0})\Pi'_a\Sigma^{-1/2\prime}$ for $i = 0, 1$.

(b) The trace test statistic $T\ \mathrm{tr}\{(R^m_{11})^{-1}R^m_{10}(R^m_{00})^{-1}R^m_{01}\}$ based on the partially aggregated (MF) system in (9) has an asymptotic distribution coinciding with $\mathrm{tr}\{\Xi^{m\prime}_{10}(\Xi^m_{11})^{-1}\Xi^m_{10}(\Xi^m_{00})^{-1}\}$ as $T \to \infty$.

It is important to note that the benchmark distribution is not generally obtained, and some size distortion may be expected.

If $m^{-1}\varpi_s' H_{00} \varpi_u = 1$ and $m^{-1}\varpi_s' H_{10} \varpi_u = 0$ for all $s, u = 1, ..., p$, then $Z_{00}^a = I$, $Z_{10}^a = 0$, and the benchmark distribution is obtained for the fully aggregated model in part (a). It is straightforward to verify that the diagonal elements of H_{00} are given by m and that those of H_{10} are given by 0. Hence, if ϖ_s is a binary vector with a unit in any element and zeros elsewhere for all $s = 1, ..., p$, the benchmark distribution is obtained. In practical terms, the benchmark is obtained if all series are skip-sampled in any way – but all in *exactly* the same way.

Single Regression Equation. In the full aggregation case, the residual-based test statistics are simply replaced by their LF analogs, $\rho_T^a \equiv T(\hat{\alpha}' R_{11}^a \hat{\alpha})^{-1} \hat{\alpha}' R_{10}^a \hat{\alpha}$ and $\tau_T^a \equiv (\hat{\alpha}' R_{00}^a \hat{\alpha}$ $(T^{-1}\hat{\alpha}' R_{11}^a \hat{\alpha}))^{-1/2} \hat{\alpha}' R_{10}^a \hat{\alpha}$, using the notation from above. Calculating these statistics poses no additional computation complications – one may simply run unit root tests on the fitted residuals of (10).

When some series are aggregated but others are observed at the high frequency, the weights (ϖ_k) for $k = p_l + 1, ..., p$ on the HF observations may be fixed or estimated. If fixed, then the analyst simply aggregates the remaining HF observations to the low frequency, resulting in the LF model in (10). Otherwise and assuming there are sufficient degrees of freedom to do so, the weights may be estimated for each HF regressor. The test statistics are then calculated from these fitted residuals, so that they are $\rho_T^m \equiv T(\hat{\alpha}' R_{11}^m \hat{\alpha})^{-1} \hat{\alpha}' R_{10}^m \hat{\alpha}$ and $\tau_T^m \equiv (\hat{\alpha}' R_{00}^m \hat{\alpha}(T^{-1}\hat{\alpha}' R_{11}^m \hat{\alpha}))^{-1/2} \hat{\alpha}' R_{10}^m \hat{\alpha}$ in this case.

The following theorem shows the limiting distributions of these test statistics. We employ the decomposition of Phillips and Ouliaris (1990): $\Sigma = L'L$, where

$$L = \begin{bmatrix} \sqrt{\sigma_1^2 - \sigma_{12}\Sigma_{22}^{-1}\sigma_{21}} & 0 \\ \Sigma_{22}^{-1/2\prime}\sigma_{21} & \Sigma_{22}^{1/2\prime} \end{bmatrix}$$

with Σ partitioned in the usual way.

Theorem 2. Consider the null hypothesis of no cointegrating relationships: $r_0 = 0$.

(a) The coefficient and t-tests calculated from the fitted residuals of the fully aggregated (LF) regression in (10) have limiting distributions given by

$$\rho_T^a \to_d \frac{\int QdQ + \kappa' L^{-1\prime}\Pi_a(\Sigma \otimes m^{-1}H_{10})\Pi_a' L^{-1}\kappa}{\int Q^2} \equiv \rho(\Pi_a)$$
$$\tau_T^a \to_d \frac{\int QdQ + \kappa' L^{-1\prime}\Pi_a(\Sigma \otimes m^{-1}H_{10})\Pi_a' L^{-1}\kappa}{\sqrt{\kappa' L^{-1\prime}\Pi_a(\Sigma \otimes m^{-1}H_{00})\Pi_a' L^{-1}\kappa \int Q^2}} \equiv \tau(\Pi_a)$$

as $T \to \infty$.

(b) The coefficient and t-tests calculated from the fitted residuals of the partially aggregated (MF) regression in (11) have limiting distributions given by $\rho_T^m \to_d \rho(\Pi_a) + O_p(1)$ and $\tau_T^m \to_d \tau(\Pi_a) + O_p(1)$ as $T \to \infty$.

The notations κ and Q are the same as in (6) above. As in the case of a system, the benchmark distributions in the single-equation case are obtained when $m^{-1}\varpi_s' H_{00}\varpi_u = 1$ and $m^{-1}\varpi_s' H_{10}\varpi_u = 0$ for all $s, u = 1, ..., p$. In other words, size is not distorted when all series are skip-sampled in the *same* way.

5 Aggregation and Size Distortion

With the limiting distributions of the test statistics from the previous section, we now examine size distortion of the tests in more detail. It is already clear that no size distortion occurs if all p HF series in the DGP are skip-sampled in the same way. Of course, this optimum may be infeasible.

System of Regression Equations. Size distortion in the LF system depends critically on the matrices $Z_{i0}^a = \Sigma^{-1/2}\Pi_a(\Sigma \otimes m^{-1}H_{i0})\Pi_a'\Sigma^{-1/2\prime}$ for $i = 0, 1$ in the limiting distribution of Theorem 1(a). Multiplicative separability of Σ from Π eliminates the impact of the error covariance, isolating the effect of ϖ_s on size distortion and allowing tractable analysis. Otherwise, size may be distorted both directly by ϖ_s and indirectly by the fact that aggregation prevents cancellation of the Σ from the limiting distributions.

Multiplicative separability is obtained when either Σ is diagonal or when $\varpi_s = \varpi_0$ for all $s = 1, ..., p$ (each series has the same aggregation scheme). When Σ is diagonal, Z_{i0}^a becomes a diagonal matrix with diagonals given by $m^{-1}\varpi_s' H_{i0}\varpi_s$ for $s = 1, ..., p$. When $\varpi_s = \varpi_0$, Z_{i0}^a becomes a diagonal matrix given by $(m^{-1}\varpi_0' H_{i0}\varpi_0)I$. We analyze size distortion resulting from the latter case, referred to as type A size distortion, relegating discussion of type B size distortion to Appendix B.

Define random variables U_1, U_2, and U_3 with distributions given by $\text{tr}\{\int (dW)'W(\int WW')^{-1} \int WdW'\}$, $2\ \text{tr}\{(\int WW')^{-1} \int WdW'\}$, and $\text{tr}\{(\int WW')^{-1}\}$, respectively, and define $a_{i\varpi} \equiv m^{-1}\varpi_0' H_{i0}\varpi_0$ for $i = 0, 1$. The limiting distribution of the test statistic based on the HF benchmark system coincides with that of U_1. When $\varpi_s = \varpi_0$ the limiting distribution of the test statistic in Theorem 1(a) based on the fully aggregated system coincides with that of $\frac{1}{a_{0\varpi}}U_1 + \frac{a_{1\varpi}}{a_{0\varpi}}U_2 + \frac{a_{1\varpi}^2}{a_{0\varpi}}U_3$.

At the critical value k^*, the test has a size given by

$$\mathbf{P}\left\{\frac{1}{a_{0\varpi}}U_1 + \frac{a_{1\varpi}}{a_{0\varpi}}U_2 + \frac{a_{1\varpi}^2}{a_{0\varpi}}U_3 \geq k^*\right\} = 1 - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\tau_\varpi(k^*, u_2, u_3)} p(u_1, u_2, u_3)du_1du_2du_3$$

where $\tau_\varpi(k^*, u_2, u_3) \equiv a_{0\varpi}k^* - a_{1\varpi}u_2 - a_{1\varpi}^2 u_3$ and $p(u_1, u_2, u_3)$ is the joint density of the three random variables. The size distortion of the test is given by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{k^*} p(u_1, u_2, u_3)du_1 - \int_{-\infty}^{\tau_\varpi(k^*, u_2, u_3)} p(u_1, u_2, u_3)du_1\right] du_2du_3,$$

and the integral in square brackets simplifies to $\int_{\tau_\varpi(k^*, u_2, u_3)}^{k^*} p(u_1, u_2, u_3)du_1$ for non-negative size distortion.

Size distortion clearly increases as τ_ϖ decreases. Because τ_ϖ is quadratic in $a_{1\varpi}$ and since U_3 has no support on the negative part of the real line ($\Sigma^{1/2\prime}\Xi_{11}^{-1}\Sigma^{1/2}$ cannot be negative definite), size is minimized at $a_{1\varpi} = 0$. In order for the test to be properly sized,

$a_{0\varpi}k^* = k^*$ must hold – i.e., $a_{0\varpi} = 1$. As already discussed above, $a_{0\varpi} = 1$ and $a_{1\varpi} = 0$ are obtained by any skip-sampling scheme, as long as each series is skip sampled in the same way.

Because of the quadratic nature of τ_ϖ, size distortion is maximized as $a_{1\varpi}$ increases and as $a_{0\varpi}$ decreases, but these are checked by the constraint that the weights sum to unity. To find the weighting scheme that causes the *worst* size distortion, we choose ϖ_0 to maximize $\epsilon_\varpi \equiv a_{1\varpi} - a_{0\varpi}$. Imposing the constraints that the weights must be positive and must sum to unity yields a Lagrangian of the form

$$\mathcal{L}_1(\varpi_0, \lambda_1, \lambda_0) = \epsilon_\varpi + \lambda_1 \left(\varpi_0'\iota - 1\right) + \lambda_0'\varpi_0$$

to be maximized. λ_1 is a scalar multiplier on an equality constraint, while λ_0 is a vector of m multipliers on inequality constraints. One of the first-order conditions of this Lagrangian is that $\lambda_1\iota = -\partial\epsilon_\varpi/\partial\varpi_0 - \lambda_0$. Moreover, noting that $\varpi_0'\iota - 1 = \left(\varpi_0' - (\iota'\iota)^{-1}\iota'\right)\iota$ and substituting this first-order condition back into the Lagrangian yields a new Lagrangian given by

$$\mathcal{L}_2(\varpi_0, \lambda_1, \lambda_0) = \epsilon_\varpi - \left(\varpi_0' - (\iota'\iota)^{-1}\iota'\right)\partial\epsilon_\varpi/\partial\varpi_0 + (\iota'\iota)^{-1}\iota'\lambda_0.$$

Some algebra reveals that this new Lagrangian has a first-order condition given by

$$-\frac{\partial^2\epsilon_\varpi}{\partial\varpi_0\partial\varpi_0'}\left(\varpi_0 - \iota(\iota'\iota)^{-1}\right) = 0,$$

which is satisfied for $\varpi_0 = \iota(\iota'\iota)^{-1}$ – that is, when flat sampling is employed.

To check for other maxima, note that

$$\frac{\partial^2\epsilon_\varpi}{\partial\varpi_0\partial\varpi_0'} = \frac{1}{m}\left(H_{11} - 2H_{00}\right)$$

where $H_{11} = H_{10} + H_{01}$. The second derivative matrix cannot be zero due to the structure of the deterministic matrices H_{11} and H_{00}.

Noting that $a_{0,\iota(\iota'\iota)^{-1}} = (1+2m^2)/3m^2$ and $a_{1,\iota(\iota'\iota)^{-1}} = (m^2-1)/6m^2$, the asymptotic test size at $\varpi = \iota(\iota'\iota)^{-1}$ is given by

$$\mathbf{P}\left\{U_1 \geq \frac{1+2m^2}{3m^2}k^* - \left(\frac{m^2-1}{6m^2}\right)U_2 - \left(\frac{m^2-1}{6m^2}\right)^2 U_3\right\}, \tag{12}$$

which is illustrated by simulations in Table 1.[8]

The main intuition to be gleaned from the case of equal aggregation weights is that we obtain *no* type A size distortion from skip sampling, but we obtain *maximum* size distortion from flat sampling.

The limiting distribution in Theorem 1(b) provides guidance for the MF case. No additional size distortion – beyond what might already exist from the LF series – is created by running the test on a MF system, as long as it is feasible to do so and only nulls $r_0 \in [(m-1)p_h, p_l + mp_h]$ are considered.

Comparing the size distortions that might result from using the distributions in Theorem 1(a) and (b), we can solidly recommend conducting the trace test on a MF system if feasible, rather than aggregating the HF series to run the test on a low-frequency system. Further aggregation will most likely aggravate any existing size distortion, unless the analyst knows the way in which the existing LF series were aggregated.[9]

Single Regression Equation. Some of the intuition above carries over to residual-based tests. Looking at the distribution in Theorem 2(a), size distortion is zero when all series have been skip sampled, so that $a_{0\varpi} = 1$ and $a_{1\varpi} = 0$, in this case, too. These results are the opposite of those expected from the literature on efficient estimation of the cointegrating vector in a MF cointegrating regression. For efficient estimation of the cointegrating vector, Chambers (2003) and Miller (2011) suggest flat sampling all series, if possible, and Pons and

[8]The table is constructed by simulating the asymptotic test size in (12), with k^* calculated for a nominal size of 0.05, using a bivariate DGP. Empirical probabilities from $1,000$ simulations are shown. Similar results for $p > 2$ suggest that this size distortion increases with p.

[9]Even in this case, the optimal aggregation of the remaining series might not be straightforward, unless the existing LF series were skip sampled in the same way.

Sansó (2005) suggest flat sampling the regressors if the regressand has been skip sampled. Our results suggest that both of these approaches entail size distortion in testing for the presence of a cointegrating vector.

Much of the intuition about systems does not carry over to a single regression equation, however. Since the residual-based test statistics lack a quadratic form similar to the trace test statistic, size distortion may be negative. Optimality is thus difficult to ascertain.

A second, and perhaps more important, fundamental difference is that the limiting distribution in the MF case of Theorem 2(b) includes $O_p(1)$ terms – not $o_p(1)$ – that cause additional size distortion. The presence of these terms may be explained by the fact that the MF regression aims to estimate weights that are otherwise set by the aggregation scheme. Under the null of no cointegration, the MF regression in (11) is spurious and these weights are estimated inconsistently.

The only restriction imposed on the aggregation weights by estimating the model in (11) with least squares is that they sum to unity. They are otherwise unrestricted. Although it seems counterintuitive, it may be preferable to impose arbitrary aggregation weights, using the LF regression in (10), than to estimate them. Even though both approaches will generally lead to inconsistency and size distortion, using arbitrary weights imposes a limit on size distortion. This counterintuitive recommendation runs against our recommendation to base trace tests on MF rather than LF systems.

As an alternative, each of the coefficients may be restricted. A CoMIDAS (cointegrating mixed data sampling) regression, introduced by Miller (2013) and based on the MIDAS regression of Ghysels *et al.* (2004), may accomplish this. MIDAS models typically employ a parsimonious nonlinear distributed lag structure. Many of the lag structures used in the literature, such as the exponential Almon lag (Ghysels *et al.*, 2005), impose that the weights sum to unity and that they are non-negative. An i^{th}-order exponential Almon lag generates weights of the form

$$\varpi_{sk}\left(\gamma_1, \ldots, \gamma_i\right) = \frac{\exp(\gamma_1 k + \cdots + \gamma_i k^i)}{\sum_{j=1}^{m} \exp(\gamma_1 j + \cdots + \gamma_i j^i)},$$

for $s = 1, ..., p$ and $k = 1, ..., m$. This function is chosen for its flexibility in mimicking empirically relevant weighting schemes, including flat sampling and end- or beginning-of-period skip sampling. If feasible, an m^{th}-order (unparsimonious) exponential Almon lag could be used to estimate the model in (11) in a more restrictive way than least squares, in order to limit size distortion.

6 Finite-sample Comparisons

Simulations provide evidence in support of type A size distortions analyzed above and shed light on the possible magnitudes of less tractable type B size distortions. Similarly to Cheung and Lai (1993), we use a bivariate ($p = 2$) HF DGP given by $\triangle^{(1/m)} y_{t-i/m} = \varepsilon_{t-i/m}$ with

$$\Sigma = var(\varepsilon_{t-i/m}) = \begin{bmatrix} 1 & \varsigma \\ \varsigma & 1 \end{bmatrix}$$

under the null. Consequently,

$$\triangle z_t^m = \eta_t^m = \Pi_m(I \otimes (I + H' + HL))u_t, \qquad \text{and} \tag{13}$$

$$\triangle z_t^a = \eta_t^a = \Pi_a(I \otimes (I + H' + HL))u_t \tag{14}$$

are the MF and LF null models.

Four parameters may be varied: m, ϖ_1, ϖ_2, and ς. We consider $m = 12$ and $T = 200$, suggestive of monthly series aggregated to an annual frequency.[10] We vary ς and the aggregation weight vectors ϖ_1, ϖ_2, and we conduct $1,000$ simulations for each model.

[10] Qualitatively similar results obtained with $m = 3$ are not shown.

Specifically, in the Tables 2 and 3 below, F-F denotes that both series have been flat sampled to the low frequency, while F- denotes that the first series has been flat sampled, while the remaining series is observed at the high frequency. We use similar notations for end-of-period sampling (E), beginning-of-period sampling (B), and a seasonal weighting pattern (S), given by $(1, 3, 9, 1, 3, 9, 1, 3, 9, 1, 3, 9)/52$. MF-OLS denotes a single MF regression estimated using least squares, while MIDAS(i) denotes a MIDAS model with an i^{th}-order exponential Almon lag.

We calculate critical values based on those that gave a size of 0.05 for the HF DGP, so that we can eliminate any distortion from the random seeds used in the simulations.

Tables 2 and 3 show no cross-correlation, $\varsigma = 0$, and strong cross-correlation, $\varsigma = 0.9$, respectively. As discussed above, a lack of cross-correlation simplifies the limiting distributions of the test statistics. The optimization in Appendix B suggests that the worst size distortion occurs when all series are aggregated using flat sampling. In other words, type B size distortion is zero, and the worst size distortion occurs when type A size distortion is maximized. Table 2 clearly supports this proposition. Note that the size 0.078 in Table 2 is almost identical to that identified by the analysis of the previous section (see Table 1). It is also clear that skip-sampling (either beginning- or end-of-period) of all series provides the least size distortion across all models and techniques. Overall, it is hard to go wrong with *any* of the aggregation schemes or combinations thereof, since size distortion does not appear to be very large in any case where $\varsigma = 0$. Type A size distortion does not appear to be serious.

The story is drastically different in the more realistic case of cross-correlation, $\varsigma = 0.9$ reported in Table 3. For fully aggregated LF series, the only sizes less than 0.10 are for those models in which both series have been aggregated using the same scheme: type A size distortion only. Moreover, the only sizes close to the nominal size are for those in which the scheme is some kind of skip sampling, consistent with our analysis above. The worst size distortion is caused by mixing different skip-sampling techniques, which suggests the use of extreme caution when aggregating data for the purposes of such tests. If the aggregation scheme for the first series is *unknown*, then skip sampling could either minimize or maximize size distortion! Flat sampling is more conservative, but mixing flat and skip sampling still leads to an unacceptably large size. Clearly, aggregating a MF model to the lowest frequency is risky.

Results for partially aggregated MF series vary greatly across tests and estimation methods when $\varsigma = 0.9$. The *only* test that has acceptable size in *every* MF case is the trace test, supporting our results above. (Incidentally, the trace test is the worst by a small margin when $\varsigma = 0$, but only because the other tests and methods improve so dramatically with no cross-correlation.) The remaining methods are all inconsistent, so size distortion is expected. Using unrestricted least squares on a MF regression performs the worst of all single-equation MF methods, but not necessarily worse than arbitrarily aggregating every series to the low frequency. MIDAS(2) does reasonably well, which is not surprising since the weight restrictions limit the size distortion from inconsistency and since the MIDAS(2) scheme nests F, E, and B. Even though MIDAS(m) is more flexible than MIDAS(2), the inconsistency appears to be problematic in the two skip-sampling cases.

Overall, we recommend using a MF system over an aggregated LF system when feasible. Similarly, we recommend using a MF regression over an aggregated LF regression, but the latter requires some restrictions on the coefficients to limit the size distortion. A MIDAS(2) regression seems to limit size distortion well.

7 Empirical Application

We study data made available by Shiller and pertaining to his 2000 book entitled *Irrational Exuberance*. This data set consists of monthly stock price, annual dividends, and earnings data (and the monthly consumer price index to allow conversion to real values), all starting

January 1871.[11] The dividend (denoted D) and earnings series represent flows, whereas stock prices (denoted P) are point-sampled. Annual data run through 2012, hence $T = 142$. Monthly data are through 2012, hence $m = 12$, $M = mT = 1704$. For the purpose of testing, we use the log of nominal price P and the log of nominal dividends D.

We compare testing cointegration with annual data versus mixed frequency using annual dividends versus annual and monthly stock price data. Regarding the annual P data we consider various configurations. Shiller sets annual price equal to January daily average – which we will refer *Annual-B*, a begin of period sampling. Alternatively, we also consider annual price series which are average monthly prices, which we will refer to as *Annual-F* and finally to contrast beginning of period we also consider end-of-period sampling, denoted *Annual-E* which is a December daily average price.

The appeal of the series which we consider fit models that coincide with the setting of our theoretical analysis. We do not include any lagged first differences, since serial correlation in financial markets is not expected. We expect that each series follows a random walk with drift, so we add a constant in the cointegrating relationship – i.e., in the Engle-Granger regression or in the error correction term – but not in the VECM itself. This is because although each series has a drift, we expect them to have the same drifts, so that the drifts cancel out.[12] Figure 1 displays the log price dividend ratio – i.e., with the cointegrating vector $(1, -1)'$ imposed on log dividends and log prices. The plot suggests that there does not appear to be a cointegrating relationship. A closer look reveals that the series may (or may not) be cointegrated up until mid-1990's, which corresponds to the tech bubble, where tech companies tend not to pay dividends. This drop in dividend paying companies suggests that (log) dividends and prices are not cointegrated after mid-1990's, and the evident structural break means that cointegration should fail for the whole sample.[13] Since

[11]Monthly dividends are available too, but they are linear interpolations of the annual data. Testing with interpolated data brings about a host of issues not covered in the current paper. Data from Robert Shiller's website is available at: `http://www.econ.yale.edu/\symbol{126}shiller/data.htm`.

[12]This corresponds Case 1* (restricted constant) in Osterwald-Lenum's (1992) critical values.

[13]Note that we unfortunately do not have enough annual data to estimate a model for the subsample since 1995.

we believe that cointegration does not hold, we do not impose $(1, -1)'$ as the cointegrating vector in the subsequent tests, in order to err on the side of favoring cointegration.

In light of this, we consider the following hypotheses and tests:

- Trace test:
  - $H_0(0)$: No cointegrating vector (2 distinct unit roots), versus $H_A(2)$: 2 cointegrating vectors (all series stationary).
  - $H_0(1)$: 1 cointegrating vector (1 distinct unit root), versus $H_A(2)$: 2 cointegrating vectors (all series stationary).
- Coefficient test:
  - $H_0(0)$: No cointegrating vector (2 distinct unit roots), versus $H_A(1)$: 1 cointegrating vector (1 distinct unit root).
- t-test:
  - $H_0(0)$: No cointegrating vector (2 distinct unit roots), versus $H_A(1)$: 1 cointegrating vector (1 distinct unit root).

Unit root tests (not shown) strongly support unit roots in the individual series. With each series following a random walk with drift $H_A(2)$ should never be true. Hence, we should *fail to reject* $H_0(0)$ or $H_0(1)$ against $H_A(2)$. Similarly, $H_0(1)$ should not hold and therefore we should *fail to reject* $H_0(0)$ against $H_0(1)$ as well. The empirical results appear in Table 4. Recall that *Annual-F* uses annual D with average P, *Annual-B* uses annual D with January P, *Annual-E* uses annual D with December P. Finally, "MF" is unrestricted MF model, using OLS in the single-equation case, "MIDAS(2)" and "MIDAS(m)" use NLS with 2^{nd}-order and m^{th}-order exponential Almon lags (single-equation case only). Estimation is a two step procedure: We first demean P and D (but not their first differences), and then run cointegration tests on demeaned data. The critical values are taken from Osterwald-Lenum (1992) and Phillips and Ouliaris (1990).

Recall that we expect that all tests reported in Table 4 should fail to reject their respective nulls. The theory in the previous sections tells us that the "safest" strategy (in terms of possible size distortion) would be to average the monthly P. Indeed, if we do this, we fail to reject any of the nulls, as expected. However, we also know that we can do better (lower size distortion) if we use the "right" skip sampling for P. Looking at *Annual-B* and *Annual-E*, we note from the results in Table 4 that the last three test statistics are slightly smaller using beginning of period rather than using flat aggregation (simple average) and much smaller using *Annual-B* versus *Annual-E*. This finding is indeed consistent with the theory if in fact annual D are sampled using something close to beginning-of-period sampling. Using *Annual-B*, which is what Shiller actually did to create his annual data set, we fail to reject any of the nulls as expected.

Now, continuing along the lines of the theory developed in the previous sections, what happens if we use the MF data directly? The results in Table 4 show that the trace test statistics are very close to the *Annual-B* case, which is also consistent with the theory if in fact annual D is sampled using something close to beginning of period sampling. Likewise, the residual-based test statistics are not close to the *Annual-B* case, which is again consistent with the theory, since size distortion is unavoidable in such a case.

To limit – but not eliminate – the size distortion in the residual-based tests, we can use a MIDAS regression strategy with exponential Almon lag. We see from Table 4 that the test statistics are about the same as using *Annual-B,* which appeared to have the least size distortion above. Again, we fail to reject any of the nulls, as expected.

Of course, we may/do not know how exactly D was aggregated. In such cases we should use the MF trace test and MIDAS for single-equation tests. Since the MF trace test and MIDAS gave us test statistics very similar to the *Annual-B,* using annual dividends and sampling January prices from each year, it would be reasonable to use the *Annual-B* strategy, which is what Shiller (1989) actually did to create the annual data set in his earlier book.

8 Concluding Remarks

Standard tests for cointegration are affected by aggregation schemes. While it is well known that aggregation and sampling frequency do not affect the long-run properties of time series, we find that the effects of aggregation on the size of the tests may be severe. Faced with this fact, we propose novel ways to solve the size distortion problems exploiting mixed frequency time series techniques which are of recent date. The issues we cover in the paper can be extended to the reverse of aggregation, namely interpolation – another approach to the creation of same frequency series. The impact of interpolation on inference is a topic we leave for future research.

Technical Appendices

A Proofs

Lemma A1. Suppose that the null of no cointegrating vectors is true. Letting $\Gamma_\eta(k) = \mathbf{E}\eta_t\eta_{t-k}'$ denote the autocovariance function of (η_t),

$$\Gamma_\eta(0) = \Sigma \otimes H_{00}, \quad \Gamma_\eta(1) = \Sigma \otimes H_{01}, \quad \Gamma_\eta(-1) = \Sigma \otimes H_{10}$$

and $\Gamma_\eta(k) = 0$ for $|k| > 1$, with $H_{00} \equiv (I + H')(I + H) + HH'$, $H_{01} \equiv H(I + H)$, and $H_{10} = H_{01}'$. The long-run variance of (η_t) is

$$\sum\nolimits_{k=-\infty}^{\infty} \Gamma_\eta(k) = \Sigma \otimes (H_{00} + H_{01} + H_{10}) = \Sigma \otimes \iota\iota'$$

under the null.

Proof of Lemma A1. Note that $\sum_{i=1}^{m-1} \triangle^{(1/m)} z_{k,t-i/m} = (H' + HL)\triangle^{(1/m)} z_{kt}$, so that

$$\eta_t = (I + (\Gamma A' \otimes I))(I \otimes (H' + HL))\triangle^{(1/m)} z_t + u_t \tag{A.1}$$

using this notation. Under the null of no cointegrating vectors, $\triangle z_t = \eta_t = (I \otimes (I + H' + HL))u_t$. The autocovariances and long-run variance follow in a straightforward way from this expression. $\square$

Lemma A2. Letting "a, m" denote either a or m on both sides,

(a) $R_{00}^{a,m} \to_d m\Xi_{00}^{a,m} \equiv m\Pi_{a,m}(\Sigma^{1/2} \otimes I)(I \otimes m^{-1}H_{00})(\Sigma^{1/2\prime} \otimes I)\Pi_{a,m}'$

(b) $T^{-1}R_{11}^{a,m} \to_d m\Xi_{11}^{a,m} \equiv m\Pi_{a,m}(\Sigma^{1/2} \otimes I)(\int WW' \otimes \iota\iota')(\Sigma^{1/2\prime} \otimes I)\Pi_{a,m}'$

(c) $R_{10}^{a,m} \to_d m\Xi_{10}^{a,m} \equiv m\Pi_{a,m}(\Sigma^{1/2} \otimes I)((\int W dW' \otimes \iota\iota') + (I \otimes m^{-1}H_{10}))(\Sigma^{1/2\prime} \otimes I)\Pi_{a,m}'$

Proof of Lemma A2. (a) R_{00}^{a} *and* R_{00}^{m}. Under the null hypothesis, the HF increments are $\triangle^{(1/m)}z_t = u_t$, so that the LF increments are $\triangle z_t = \sum_{i=0}^{m-1} u_{t-i/m} = (I \otimes (I + H' + HL))u_t$. Since (u_t) is iid,

$$\text{plim } T^{-1} \sum u_t u_t' = \text{plim } T^{-1} \sum u_{t-1} u_{t-1}' = \mathbf{E} u_t u_t' = \Sigma \otimes I,$$

while

$$\text{plim } T^{-1} \sum u_t u_{t-1}' = \text{plim } T^{-1} \sum u_{t-1} u_t' = 0$$

by an LLN. Thus,

$$R_{00} \to_p (I \otimes (I + H'))(\Sigma \otimes I)(I \otimes (I + H)) + (I \otimes H)(\Sigma \otimes I)(I \otimes H'),$$

which simplifies to $\Sigma \otimes H_{00}$ in our notation.

(b) R_{11}^{a} *and* R_{11}^{m}. We may write the LF series (z_{t-1}) in terms of (u_t) by noting that $z_{t-1} = \sum_{j=1}^{t-1} \triangle z_j$ and then proceeding as above. Specifically, we have

$$z_{t-1} = (I \otimes (I + H')) \sum_{j=1}^{t-1} u_j + (I \otimes H) \sum_{j=1}^{t-2} u_j.$$

The outer product $z_{t-1} z_{t-1}'$ thus involves four terms containing $\sum_{j=1}^{t-a} \sum_{i=1}^{t-b} u_j u_i$ for $a, b = 1, 2$. For finite integers a, b,

$$T^{-2} \sum \sum_{j=1}^{t-a} \sum_{i=1}^{t-b} u_j u_i = T^{-2} \sum \sum_{j=1}^{t} \sum_{i=1}^{t} u_j u_i + o_p(1)$$

by applying standard covariance asymptotics to the $O_p(T^{-1})$ remainder terms. Thus,

$$\begin{aligned} T^{-1} R_{11} &= (I \otimes (I + H' + H)) T^{-2} \sum \sum_{j=1}^{t} \sum_{i=1}^{t} u_j u_i (I \otimes (I + H' + H)) + o_p(1) \\ &= (I \otimes \iota\iota') T^{-2} \sum \sum_{j=1}^{t} \sum_{i=1}^{t} u_j u_i (I \otimes \iota\iota') + o_p(1) \end{aligned}$$

since $I + H' + H = \iota\iota'$ by construction.

Since the sum $\sum_{i=1}^{t} u_j$ is not iid, the increment by which the sample moment above is defined affects the asymptotic distributions. Specifically, a LF moment does not have the same dependence structure – and therefore not the same limit – as a HF moment. It is convenient to note that $(I \otimes \iota')u_j = (\sum_{k=0}^{m-1} \varepsilon_{j-k/m} \otimes 1)$, so that

$$T^{-1}R_{11} = m\left(T^{-1}\sum\left(\frac{1}{\sqrt{mT}}\sum_{j=1}^{t}\sum_{k=0}^{m-1}\varepsilon_{j-k/m}\right)\left(\frac{1}{\sqrt{mT}}\sum_{i=1}^{t}\sum_{l=0}^{m-1}\varepsilon'_{i-l/m}\right)\otimes\iota\iota'\right) + o_p(1)$$

Now, $\frac{1}{\sqrt{mT}}\sum_{j=1}^{t}\sum_{k=0}^{m-1}\varepsilon_{j-k/m} \to_d B(r)$ by appealing to an invariance principle (see Miller, 2011), so that $T^{-1}R_{11} \to_d m(\int BB' \otimes \iota\iota')$.

(c) R_{10}^{a} *and* R_{10}^{m}. The sample moment $T^{-1}\sum z_{t-1}\triangle z_t'$ may be rewritten as

$$\begin{aligned} T^{-1}\sum z_{t-1}\triangle z_t' &= T^{-1}\sum\left[(I\otimes(I+H'))\sum_{j=1}^{t-1}u_j + (I\otimes H)\sum_{j=1}^{t-1}u_{j-1}\right]u_t'(I\otimes(I+H)) \\ &+ T^{-1}\sum\left[(I\otimes(I+H'))\sum_{j=1}^{t-1}u_j + (I\otimes H)\sum_{j=1}^{t-1}u_{j-1}\right]u_{t-1}'(I\otimes H') \end{aligned} \tag{A.2}$$

using arguments similar to those above. The square-bracketed factor in (A.2) may be written as

$$\left(\sum_{j=1}^{t-1}\sum_{k=0}^{m-1}\varepsilon_{j-k/m}\otimes\iota\right) + (I\otimes H)(u_0 - u_{t-1})$$

by noting that $\sum_{j=1}^{t-1} u_{j-1} = \sum_{j=1}^{t-1} u_j + (u_0 - u_{t-1})$ and recognizing that $I + H' + H = \iota\iota'$. Since (u_t) is an iid series, the whole first term of (A.2) may be written as

$$T^{-1}\sum\left(\sum_{j=1}^{t-1}\sum_{k=0}^{m-1}\varepsilon_{j-k/m}\otimes\iota\right)u_t'(I\otimes(I+H)) + o_p(1).$$

The square-bracketed factor may be written as

$$\left(\sum_{j=1}^{t-2}\sum_{k=0}^{m-1}\varepsilon_{j-k/m}\otimes\iota\right)+(I\otimes(I+H'))u_{t-1}+(I\otimes H)u_0$$

by noting that $\sum_{j=1}^{t-1}u_j=\sum_{j=1}^{t-2}u_j+u_{t-1}$ and $\sum_{j=1}^{t-1}u_{j-1}=\sum_{j=1}^{t-2}u_j+u_0$. The second term of (A.2) is thus

$$T^{-1}\sum\left(\sum_{j=1}^{t-2}\sum_{k=0}^{m-1}\varepsilon_{j-k/m}\otimes\iota\right)u_{t-1}'(I\otimes H')+(I\otimes(I+H'))T^{-1}\sum u_{t-1}u_{t-1}'(I\otimes H') \tag{A.3}$$

$$+\,o_p\left(1\right),$$

with the key difference between the two terms of (A.2) being the nondegenerate limit of $T^{-1}\sum u_{t-1}u_{t-1}'$ in the second. The iid assumption on (u_t) allows (A.3) to be rewritten as

$$T^{-1}\sum\left(\sum_{j=1}^{t-1}\sum_{k=0}^{m-1}\varepsilon_{j-k/m}\otimes\iota\right)u_t'(I\otimes H')+(\Sigma\otimes H_{10})+o_p\left(1\right),$$

so that the whole expression in (A.2) is

$$\begin{aligned}R_{10}&=T^{-1}\sum\left(\sum_{j=1}^{t-1}\sum_{k=0}^{m-1}\varepsilon_{j-k/m}\otimes\iota\right)u_t'(I\otimes\iota\iota')+(\Sigma\otimes H_{10})+o_p\left(1\right)\\&=\sum_{l=0}^{m-1}\left(T^{-1}\sum\left(\sum_{j=1}^{t-1}\sum_{k=0}^{m-1}\varepsilon_{j-k/m}\right)\varepsilon_{t-l/m}'\otimes\iota\iota'\right)+(\Sigma\otimes H_{10})+o_p\left(1\right),\end{aligned}$$

which has a limiting distribution given by $m(\int BdB'\otimes\iota\iota')+(\Sigma\otimes H_{10})$. The stated results follow by again considering the weight matrices Π_a and Π_m. □

Proof of Theorem 1. Writing $(\int WW'\otimes\iota\iota')=(I\otimes\iota)\int WW'(I\otimes\iota')$, note that the weights ϖ_s in Π_a must sum to one and thus $\Pi_a(I\otimes\iota)=I$. The results for the fully aggregated system then follow from continuity of the trace and matrix multiplication using the results of Lemma A2. The results for the partially aggregated system follow in the same way, but do not simplify without imposing additional assumptions. □

Proof of Theorem 2. *Low-frequency Case.* The least squares estimator $\hat{\beta}$ of β using the aggregated low-frequency model in (10) may be written as

$$\hat{\beta} = \left(T^{-2}\sum\nolimits_t E_2'\Pi_a z_t z_t'\Pi_a' E_2\right)^{-1} T^{-2}\sum\nolimits_t E_2'\Pi_a z_t z_t'\Pi_a' E_1$$

where E_1 and E_2 are the vector and matrix that select the first column and all but the first column, respectively, of the preceding matrix. From the proof of the previous theorem, it follows that

$$T^{-2}\sum\nolimits_t z_t z_t' = T^{-1}R_{11} + o_p(1) \to_d m\left(\int B(r)B(r)'dr \otimes \iota\iota'\right)$$

so that $T^{-2}\sum_t \Pi_a z_t z_t'\Pi_a' \to_d m\int BB'$, similarly to the result in that theorem. Hence, $\hat{\beta} \to_d (\int B_2 B_2')^{-1}\int B_2 B_1$ using the partition defined above: $B_1 = E_1'B$ and $B_2 = E_2'B$.

The series on which the unit root tests are conducted is $\hat{e}_t^a = z_t'\Pi_a'(1, -\hat{\beta}')'$, and the tests may be written as $T\hat{\rho}_T = (T^{-2}\sum_t(\hat{e}_{t-1}^a)^2)^{-1}T^{-1}\sum_t \hat{e}_{t-1}^a \triangle \hat{e}_t^a$ or $T\hat{\rho}_T = (\hat{\alpha}'T^{-1}R_{11}^a\hat{\alpha})^{-1}\hat{\alpha}'R_{10}^a\hat{\alpha}$ and $\hat{\tau}_T = (\hat{\sigma}^2 T^{-2}\sum_t(\hat{e}_{t-1}^a)^2)^{-1/2}T^{-1}\sum_t \hat{e}_{t-1}^a\triangle\hat{e}_t^a$ with $\hat{\sigma}^2 = T^{-1}\sum_t(\triangle\hat{e}_t^a)^2$ or $\hat{\tau}_T = (\hat{\alpha}'R_{00}^a\hat{\alpha}$ $\hat{\alpha}'T^{-1}R_{11}^a\hat{\alpha})^{-1/2}\hat{\alpha}'R_{10}^a\hat{\alpha}$. (Using algebra along the lines of Phillips and Ouliaris, 1990, along with the limiting distributions established in Lemma A2 yields the stated results.)

Mixed-frequency Case. The MF model in (11) may be rewritten as

$$\begin{aligned} z_{1t}^a &= (z_{2t}^a, ..., z_{pt}^a)\beta + w_t'\gamma + e_t^a \\ z_t'\Pi_a'(1, -\beta')' &= w_t'\gamma + e_t^a \end{aligned}$$

where $z_{kt}^a \equiv y_{kt}$ for $k = p_l + 1, ..., p$, $w_t \equiv (\triangle^{(1/m)}z_{p_l+1,t}'H', ..., \triangle^{(1/m)}z_{pt}'H')'$, and $\gamma \equiv (\beta_{p_l+1}\varpi_{p_l+1}', ..., \beta_p\varpi_p')'$. Note that the first term now contains no weights to be estimated, and that the second term contains only I(0) series, and that while $\beta_k\varpi_k$ is estimated jointly in the second term, β is identified by the first term. The choice of z_{kt}^{a} is for simplification but not relevant for the results of the proof. If instead $z_{kt}^{a} \equiv y_{kt} - \varpi_{k}^{*} H \triangle^{(1/m)} z_{kt}$ for some other weight vector ϖ_{k}^{*}, additional terms of the type $(\varpi_k - \varpi_k^{*})' \triangle^{(1/m)} z_{kt}' H'$ will not qualitatively affect the results.

It straightforward to establish that the limit of $\hat{\beta}$ is the same as that which omits the stationary regressors (w_t). A fundamental difference arises from comparing the fitted residuals. In the LF case, these are $\hat{e}_t^a = z_t' \Pi_a' (1, -\hat{\beta}')' = e_t^a - z_t' \Pi_a' E_2 (\hat{\beta} - \beta)$, while they are $\hat{e}_t^a = z_t' \Pi_a' (1, -\hat{\beta}')' - w_t' (T^{-1} \sum_t w_t w_t')^{-1} T^{-1} \sum_t w_t z_t' \Pi_a' (1, -\hat{\beta}')' = e_t^a - z_t' \Pi_a' E_2 (\hat{\beta} - \beta) - w_t' (\hat{\gamma} - \gamma)$ in the MF case. The reason for the additional $O_p(1)$ terms in the results is that $\hat{\gamma}$ does not estimate γ consistently under the null of a spurious regression (no cointegration). The details are tedious but follow using the same logic as that used by Phillips and Ouliaris (1990) and above. □

B Type B Size Distortion

General Weights, Diagonal Variance

Because Σ is diagonal, $\Pi(\Sigma \otimes m^{-1} H_{i0}) \Pi' = \Sigma H_{i0}^{*}(\varpi)$ with $H_{i0}^{*}(\varpi) \equiv m^{-1}(\varpi_1' H_{i0} \varpi_1 \oplus \cdots \oplus \varpi_p' H_{i0} \varpi_p)$. The inverse of $\Pi(\Sigma \otimes m^{-1} H_{00}) \Pi'$ is thus easy to obtain: $(H_{00}^{*}(\varpi))^{-1} \Sigma^{-1}$. The trace may therefore be rewritten as

$$
\begin{aligned}
\operatorname{tr}\left\{(\Xi_{11}^{*})^{-1} \Xi_{10}^{*} (\Xi_{00}^{*})^{-1} \Xi_{01}^{*}\right\} &= \operatorname{tr}\{\Xi_{11}^{-1} \Xi_{10} (H_{00}^{*}(\varpi))^{-1} \Sigma^{-1} \Xi_{01}\} \\
&+ \operatorname{tr}\{2 \Xi_{11}^{-1} \Xi_{10} (H_{00}^{*}(\varpi))^{-1} H_{01}^{*}(\varpi)\} \\
&+ \operatorname{tr}\{\Xi_{11}^{-1} \Sigma H_{10}^{*}(\varpi) (H_{00}^{*}(\varpi))^{-1} H_{01}^{*}(\varpi)\} \\
&= V_1' b_1(\varpi) + V_2' b_2(\varpi) + V_3' b_3(\varpi)
\end{aligned}
$$

where

$$
\begin{aligned}
&V_1 \equiv vec(\Sigma^{-1}\Xi_{01}\Xi_{11}^{-1}\Xi_{10}) && b_1(\varpi) \equiv vec\left[(H_{00}^*(\varpi))^{-1}\right] \\
&V_2 \equiv vec(2\Xi_{11}^{-1}\Xi_{10}) && b_2(\varpi) \equiv vec\left[(H_{00}^*(\varpi))^{-1}H_{01}^*(\varpi)\right] \\
&V_3 \equiv vec(\Xi_{11}^{-1}\Sigma) && b_3(\varpi) \equiv vec\left[H_{10}^*(\varpi)(H_{00}^*(\varpi))^{-1}H_{01}^*(\varpi)\right]
\end{aligned}
$$

by the property of the *vec* operator.

For expositional clarity, consider the case of $p = 2$ and let V_{j1} and V_{j4} denote the first and last element of the 4×1 vector V_j for $j = 1, 2, 3$. Note that $U_j = V_{j1} + V_{j4}$ and that V_{j1} and V_{j4} are statistically independent due to the diagonal structure of Σ. The diagonal structure multiplies V_{12}, V_{13}, etc., by zero, so that the test size is now equal to

$$
\begin{aligned}
&\mathbf{P}\left\{V_{11}/a_{0\varpi_1} + V_{14}/a_{0\varpi_2} + V_{21}a_{1\varpi_1}/a_{0\varpi_1} + V_{24}a_{1\varpi_2}/a_{0\varpi_2} + V_{31}a_{1\varpi_1}^2/a_{0\varpi_1} + V_{34}a_{1\varpi_2}^2/a_{0\varpi_2} \geq k^*\right\} \\
&= \mathbf{P}\left\{V_{11} + V_{14}a_{0\varpi_1}/a_{0\varpi_2} + V_{21}a_{1\varpi_1} + V_{24}a_{0\varpi_1}a_{1\varpi_2}/a_{0\varpi_2} + V_{31}a_{1\varpi_1}^2 + V_{34}a_{0\varpi_1}a_{1\varpi_2}^2/a_{0\varpi_2} \geq k^*a_{0\varpi_1}\right\} \\
&= \mathbf{P}\left\{U_1 + U_2a_{1\varpi_1} + U_3a_{1\varpi_1}^2 + V_{14}c_{1\varpi} + V_{24}c_{2\varpi} + V_{34}c_{3\varpi} \geq k^*a_{0\varpi_1}\right\},
\end{aligned}
$$

where $c_{i\varpi} \equiv a_{1\varpi_2}^{i-1}a_{0\varpi_1}/a_{0\varpi_2} - a_{1\varpi_1}^{i-1}$.

We now recycle the notation

$$
\tau_\varpi(k^*, u_2, u_3) \equiv k^*a_{0\varpi_1} - u_2a_{1\varpi_1} - u_3a_{1\varpi_1}^2 - v_{14}c_{1\varpi} - v_{24}c_{2\varpi} - v_{34}c_{3\varpi},
$$

and again we wish to set this equal to $\tau_\varpi = k^*$ to minimize size distortion. This is accomplished by any weights that make $a_{0\varpi_1}, a_{0\varpi_2} = 1$ and $a_{1\varpi_1}, a_{1\varpi_2} = 0$.

We wish to minimize τ_ϖ to maximize positive size distortion. The first three terms are already minimized at $\varpi_1 = \iota(\iota'\iota)^{-1}$, but can the entire expression be decreased by varying ϖ_2? Note that the last quadratic term is

$$
a_{1\varpi_2}^2a_{0\varpi_1}/a_{0\varpi_2} - a_{1\varpi_1}^2 = a_{1\varpi_1}^2\left(\frac{a_{1\varpi_2}^2/a_{0\varpi_2}}{a_{1\varpi_1}^2/a_{0\varpi_1}} - 1\right),
$$

so we must increase $a^2_{1\varpi_2}$ and decrease $a_{0\varpi_2}$ to increase size. However, we already know that the $a^2_{1\varpi_2}/a_{0\varpi_2}$ is maximized at $\varpi_2 = \iota(\iota'\iota)^{-1}$.

Let $\varpi \equiv (\varpi_1', \varpi_2')'$, a $2m \times 1$ vector containing possibly distinct weight subvectors. Size may be maximized with a new Lagrangian of the form

$$\mathcal{L}_1(\varpi, \lambda_1, \lambda_0) = \epsilon_\varpi + \lambda_1'((I \otimes \iota')\varpi - \iota_2) + \lambda_0'\varpi$$

where $\iota_2 = (1,1)'$. λ_1 is now a bivariate vector of multipliers reflecting the constraint that each weight subvector must sum to unity, and λ_0 is now a $2m$-vector of non-negativity constraints. Similarly to the case of equal weights, we substitute a first-order condition, $(I \otimes \iota)\lambda_1 = -\partial\epsilon_\varpi/\partial\varpi - \lambda_0$, into the Lagrangian:

$$\begin{aligned}\mathcal{L}_2(\varpi, \lambda_1, \lambda_0) &= \epsilon_\varpi + \lambda_1'(I \otimes \iota')(\varpi - (I \otimes \iota(\iota'\iota)^{-1})\iota_2) + \lambda_0'\varpi \\ &= \epsilon_\varpi - (\varpi' - \iota_2'(I \otimes (\iota'\iota)^{-1}\iota'))\partial\epsilon_\varpi/\partial\varpi + \iota_2'(I \otimes (\iota'\iota)^{-1}\iota'))\lambda_0\end{aligned}$$

This Lagrangian has a first-order condition given by

$$-\frac{\partial^2_\varpi \epsilon_\varpi}{\partial\varpi\partial\varpi'}(\varpi - (I \otimes \iota(\iota'\iota)^{-1})\iota_2) = 0,$$

which is almost identical to the previous case and gives the same result: $\varpi_1 = \varpi_2 = \iota(\iota'\iota)^{-1}$.

Generalizing the result to $p > 2$ creates additional quadratic terms of the form $a^2_{1\varpi_s}a_{0\varpi_1}/a_{0\varpi_s} - a^2_{1\varpi_1}$, which are also maximized at $\varpi_s = \iota(\iota'\iota)^{-1}$.

In other words, size cannot be any worse than matching flat aggregation schemes when the variance matrix Σ is diagonal. Even choosing different aggregation schemes cannot make the test size any worse. This is still a special case.

General Weights, General Variance

More generally, we do not obtain the multiplicative separability above. Note that

$$\Pi(\Sigma \otimes m^{-1}H_{i0})\Pi' = \Sigma \odot \begin{bmatrix} m^{-1}\varpi_1' H_{i0}\varpi_1 & \cdots & m^{-1}\varpi_1' H_{i0}\varpi_p \\ \vdots & \ddots & \vdots \\ m^{-1}\varpi_p' H_{i0}\varpi_1 & \cdots & m^{-1}\varpi_p' H_{i0}\varpi_p \end{bmatrix}$$

which means that the minimum size is obtained when

$$\Pi(\Sigma \otimes m^{-1}H_{00})\Pi' = \Sigma \odot \iota_2\iota_2' \qquad \text{and} \qquad \Pi(\Sigma \otimes m^{-1}H_{10})\Pi' = 0$$

as above.

Maximization is much more complicated, and we take a more heuristic approach here. New potential for size distortion arises from the off-diagonal terms $\varpi_s' H_{i0}\varpi_u$ for $s \neq u$. Roughly speaking, size distortion increases as $m^{-1}\varpi_s' H_{00}\varpi_u$ decreases from unity (its maximum value) and as $m^{-1}\varpi_s' H_{10}\varpi_u$ increases from zero (its minimum value). Because of the structure of these matrices, the smallest elements of H_{00} are the corners furthest from the diagonal, which have unit elements. On the contrary, these corners have the largest elements of $H_{10} + H_{01}$, which are $m - 1$. Consequently, when Σ is not diagonal, the *worst* size distortion occurs when skip sampling opposite ends of the LF interval – i.e., mixing end-of-period with beginning-of-period sampling. The degree of size distortion depends on the magnitude of the off-diagonal elements of Σ, as our simulations illustrate.

References

Anderson, B. D. O., M. Deistler, E. Felsenstein, B. Funovits, P. Zadrozny, M. Eichler, W. Chen, and M. Zamani (2012). "Identifiability of Regular and Singular Multivariate Autoregressive Models from Mixed Frequency Data," In *51st Conference on Decision and Control*, 184-189, Maui, HI. IEEE Control Systems Society.

Ansley, C. F., and R. Kohn (1983). "Exact likelihood of vector autoregressive-moving average process with missing or aggregated data," *Biometrika* 70, 275-278.

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

Andreou, E., E. Ghysels, and A. Kourtellos (2011). "Forecasting with Mixed-Frequency Data," In *Oxford Handbook of Economic Forecasting*, ed. by M. Clements, and D. Hendry, 225-245.

Armesto, M., K. Engemann, and M. Owyang (2010). "Forecasting with Mixed Frequencies," *Federal Reserve Bank of St. Louis Review* 92, 521-536.

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

Chambers, M.J. (2011). "Cointegration and Sampling Frequency," *Econometrics Journal* 14, 156-185.

Cheung, Y.-W. and K.S. Lai (1993), "Finite-sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration," *Oxford Bulletin of Economics and Statistics* 55, 313-328.

Engle, R.F. and C.W.J. Granger (1987). "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55, 251-276.

Felsenstein, E., B. Funovits, M. Deistler, B. Anderson, M. Zamani, and W. Chen (2013). "Regular and Singular Multivariate Autoregressive Models and Mixed Frequency Data: Identifiability and Consistent Estimation," Working paper, Vienna University of Technology.

Foroni, C., E. Ghysels, and M. Marcellino (2013). "Mixed Frequency Approaches for Vector Autoregressions," In *VAR Models in Macroeconomics, Financial Econometrics, and Forecasting – Advances in Econometrics*, ed. by T. Fomby, and L. Kilian, vol. 31.

Friedman, M. (1962). "The Interpolation of Time Series by Related Series," *Journal of the American Statistical Association* 57, 729-757.

Ghysels, E. (2012). "Macroeconomics and the Reality of Mixed Frequency Data," Department of Economics, University of North Carolina, mimeo.

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 (2005). "There Is a Risk-return Tradeoff After All," *Journal of Financial Economics* 76, 509-548.

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.

Götz, T.B., A. Hecq, and J.-P. Urbain (2012). "Testing for Common Cycles in Non-Stationary VARs with Varied Frequency Data," Department of Quantitative Economics, Maastricht University, mimeo.

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.

Harvey, A. C., and R. G. Pierse (1984). "Estimating missing observations in economic time series," *Journal of the American Statistical Association* 79, 125-131.

Haug, A.A. (2002). "Temporal Aggregation and the Power of Cointegration Tests: a Monte Carlo Study," *Oxford Bulletin of Economics and Statistics* 64, 399-411.

Hooker, M.A. (1993). "Testing for Cointegration: Power versus Frequency of Observation" *Economics Letters* 41, 359-362.

Horvath, M.T.K. and M.W. Watson (1995). "Testing for Cointegration When Some of the Cointegrating Vectors Are Prespecified," *Econometric Theory* 11, 984-1014.

Hu, W. (1996). "Time Aggregation and Skip Sampling in Cointegration Tests," *Statistical Papers* 37, 225-234.

Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control* 12, 231-54.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Jones, R. H. (1980). "Maximum likelihood fitting of ARMA models to time series with missing observations," *Technometrics* 22, 389-395.

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

Mariano, R.S. and Y. Murasawa (2003). "A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series," *Journal of Applied Econometrics* 18, 427-43.

Mariano, R.S. and Y. Murasawa (2010). "A Coincident Index, Common Factors, and Monthly Real GDP," *Oxford Bulletin of Economics and Statistics* 72, 27-45.

McCracken, M., M. Owyang, and T. Sekhposyan (2013). "Real-Time Forecasting with a Large Bayesian Block Model," Discussion Paper, Federal Reserve Bank of St. Louis and Bank of Canada.

Miller, J.I. (2011). "Conditionally Efficient Estimation of Long-run Relationships Using Mixed-frequency Time Series," Department of Economics, University of Missouri, mimeo.

Miller, J.I. (2013). "Mixed-frequency Cointegrating Regressions with Parsimonious Distributed Lag Structures," *Journal of Financial Econometrics*, forthcoming.

Osterwald-Lenum, M. (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics* 54, 461-472.

Perron, P. (1991). "Test Consistency with Varying Sampling Frequency," *Econometric Theory* 7, 341-368.

Phillips, P.C.B. and S. Ouliaris (1990). "Asymptotic Properties of Residual-Based Tests for Cointegration," *Econometrica* 58, 165-193.

Pons, G. and A. Sansó (2005). "Estimation of Cointegrating Vectors with Time Series Measured at Different Periodicity," *Econometric Theory* 21, 735-756.

Seong, B., S.K. Ahn, and P.A. Zadorozny (2013). "Cointegration analysis with mixed-frequency data," *Journal of Time Series Analysis* 34, 194-205.

Shiller, R.J. (1989). *Market Volatility.* Cambridge, MA: MIT Press.

Shiller, R.J. (2000). *Irrational Exuberance.* Princeton: Princeton University Press.

Shiller, R.J. and P. Perron (1985). "Testing the Random Walk Hypothesis: Power versus Frequency of Observation" *Economics Letters* 18, 381-386.

Stock, J.H. (1987). "Temporal Aggregation and Structural Inference in Macroeconomics – a Comment," *Carnegie-Rochester Conference Series of Public Policy* 26, 131-139.

Zadrozny, P. A. (1988). "Gaussian Likelihood of Continuous-Time ARMAX Models When Data Are Stocks and Flows at Different Frequencies," *Econometric Theory* 4, 108-124.

Zadrozny, P. A. (1990). "Estimating a multivariate ARMA model with mixed frequency data: an application to forecasting US GNP at monthly intervals," *Economic Review - Federal Reserve Bank of Atlanta* 90-6, 2-15.

Figures and Tables

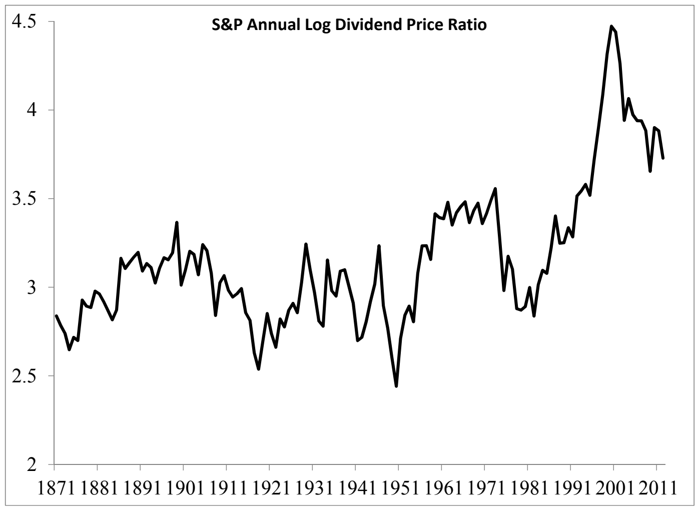

Figure 1: S&P Annual Log Dividend Price Ratio

Table 1: Simulated Size for Flat Weights

$m=$	1	2	3	...	6	...	12	...	18
$T = 600$	0.05	0.065	0.073	...	0.078	...	0.079	...	0.079
$T = 2400$	0.05	0.054	0.062	...	0.063	...	0.064	...	0.064

Table 2: Size: $m = 12$ $\varsigma = 0$

Case	Test	F-F	E-E	B-B	S-S	F-E	F-B	E-B
Low-freq	Trace	0.078	**0.057**	**0.056**	0.077	**0.068**	**0.061**	**0.055**
	Rho	0.006	**0.047**	**0.046**	0.006	0.018	0.017	**0.047**
	Tau	0.010	**0.046**	**0.046**	0.010	0.019	0.019	**0.046**
Case	Test	F-	E-	B-	S-			
Mixed-freq	Trace	0.072	**0.056**	**0.062**	0.071			
MF-OLS	Rho	0.018	**0.047**	**0.051**	0.018			
	Tau	0.019	**0.046**	**0.051**	0.020			
MIDAS(2)	Rho	0.011	**0.042**	**0.041**	0.012			
	Tau	0.013	**0.045**	**0.042**	0.013			
MIDAS(m)	Rho	0.012	**0.038**	**0.038**	0.013			
	Tau	0.015	**0.041**	**0.042**	0.014			

Bold denotes $|\text{size} - 0.05| < 0.02$.

Table 3: Size: $m = 12$ $\varsigma = 0.9$

Case	Test	F-F	E-E	B-B	S-S	F-E	F-B	E-B
Low-freq	Trace	0.078	**0.057**	**0.056**	0.077	0.489	0.475	0.821
	Rho	0.011	**0.048**	**0.050**	0.011	0.397	0.400	0.743
	Tau	0.012	**0.051**	**0.049**	0.012	0.383	0.393	0.735
Case	Test	F-	E-	B-	S-			
Mixed-freq	Trace	**0.068**	**0.052**	**0.051**	**0.069**			
MF-OLS	Rho	0.397	**0.048**	0.732	0.452			
	Tau	0.383	**0.051**	0.722	0.436			
MIDAS(2)	Rho	0.021	**0.053**	**0.056**	0.026			
	Tau	0.021	**0.053**	**0.054**	0.024			
MIDAS(m)	Rho	**0.064**	0.229	0.355	**0.068**			
	Tau	**0.065**	0.227	0.336	**0.067**			

Bold denotes $|\text{size} - 0.05| < 0.02$.

Table 4: Cointegration Tests: Log Stock Price and Log Dividend

	Trace Test		Coeff. Test	T-test
	$H_0(1)/H_A(2)$	$H_0(0)/H_A(2)$	$H_0(0)/H_A(1)$	$H_0(0)/H_A(1)$
Annual-F	1.41	12.47	-19.37	-3.11
Annual-B	1.17	10.39	-17.40	-2.95
Annual-E	1.41	28.29	-36.55	-4.27
Mixed-freq	1.05	10.95	-23.22	-3.41
MIDAS(2)	–	–	-16.16	-2.84
MIDAS(m)	–	–	-17.42	-2.95
Critical Values	9.24	19.96	-20.5	-3.37