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

Eric Ghysels† J. Isaac Miller‡

First Draft: February 24, 2013
This Draft: November 19, 2014

Abstract

We examine the effects of mixed sampling frequencies and temporal aggregation on the size of commonly used tests for cointegration, and we find that these effects may be severe. Matching sampling schemes of all series generally reduces size, and the nominal size is obtained asymptotically only when all series are skip-sampled in the same way—e.g., end-of-period sampling. We propose and analyze mixed-frequency versions of the cointegration tests in order to control the size when some high-frequency data are available. Otherwise, when no high-frequency data are available, we discuss controlling size using bootstrapped critical values. We test stock prices and dividends for cointegration as an empirical demonstration.

JEL Classification: C12, C32

Key words and phrases: temporal aggregation, mixed sampling frequencies, cointegration, trace test, residual-based cointegration tests

*We thank Thomas Götz, Peter Phillips, Peter Zadrozny, and two anonymous referees for their helpful comments. We are also grateful to participants of the 2013 NBER-NSF Time Series Conference, the Advances in Econometrics Conference in Honor of Peter Phillips, the 24th EC² Conference (The Econometric Analysis of Mixed Frequency Data), the 2014 Kansas Workshop on Advanced Econometrics, seminars at Sungkyunkwan University and the University of Maryland, and a research colloquium at the University of Missouri for useful comments. Ghysels acknowledges support of a Marie Curie FP7-PEOPLE-2010-IIF grant.

†CEPR Research Fellow and Department of Economics and Department of Finance, Kenan-Flagler Business School, University of North Carolina at Chapel Hill. E-mail: eghysels@unc.edu

‡Corresponding author. Department of Economics, University of Missouri, 118 Professional Building, Columbia, Missouri 65211, USA. E-mail: millerjisac@missouri.edu
1 Introduction

Economic data are sampled at different frequencies, mostly because the cost of collecting or measuring variables can vary considerably. Prices (or price indices) are relatively easy to collect. At one extreme of the spectrum we have the prices of financial assets, such as stocks, commodities, etc. which 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).

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 quarterly via GDP) across different countries we would end up with quarterly CPI (consumer price index, a stock) and quarterly GDP (a flow) for each country. Note that in this case, CPI data are available monthly but are aligned with quarterly GDP observations.

The most common method for collecting same-frequency series is temporal aggregation, and we use “temporal aggregation” in a very general sense that includes periodic sampling. To operationalize temporal aggregation, we consider a vector \( \omega \) of \( m \) aggregation weights, ordered from most recent to least recent. Quarterly CPI may be created by averaging monthly CPI. Referred to as flat sampling or average sampling, such a scheme would have equal aggregation weights of \( 1/3 \) (more generally, \( 1/m \)). Alternatively, quarterly CPI may be created by sampling one month per quarter, referred to as skip sampling. The most common type of skip sampling, end-of-period (EOP) sampling, features a unit weight on the most recent high-frequency observation in each low-frequency period, and zero weights on the remaining high-frequency observations.

The analysis in this paper extends a line of research on cointegration under the assumptions of temporal aggregation or mixed sampling frequencies. In particular, Stock (1987),
Granger (1990), and Phillips (1991) were among the first to analyze cointegration under temporal aggregation, and important contributions have been made by Marcellino (1999) and Chambers (2003, 2011), with simulation studies focusing on the power of cointegration tests conducted by Hooker (1993), Lahiri and Mamingi (1995), Hu (1996), Otero and Smith (2000), and Haug (2002). In the setting of cointegrated series observed at different frequencies, a number of recent contributions have been made by Miller (2014a, 2014b), Götz et al. (2013, 2014), and Seong et al. (2013).

In this paper, we consider testing for cointegration, and, in particular, 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 is allowed). We show that, even though temporal aggregation does not change the cointegrating vector, aggregation can cause size distortion – i.e., deviation from the nominal size when critical values from the standard textbook distributions are used – in cointegration tests.\footnote{In contrast, Hooker (1993) \textit{inter alia} 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. For example, 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.}

Size distortions can either be absent, mild, or severe. Consider a first example of inflation sampled quarterly in several countries and the analyst’s interest is the possible cointegration of prices. The sampling scheme involved in this example is the same – i.e., all series are skip sampled at a (low) quarterly frequency. In such cases when all series are skip sampled in the same way, we show that there is no size distortion.

Now consider a second case, in which cointegration of GDP from the different countries is of interest. In this case, we expect size distortions, but these distortions appear to be acceptable. Third, consider testing cointegration within a given country between GDP and

\footnote{We acknowledge that some existing cointegration tests – especially those less reliant on parametric form – may be more robust to aggregation. We focus here on the three most widely used cointegration tests in empirical work, and leave analysis of other cointegration tests to future research.}
prices. Here the series are aggregated/sampled differently – namely a flow versus a stock with the latter being available at a higher frequency. We show that with series subject to different aggregation schemes, size distortion of these tests can be quite severe.

As a fourth example, suppose that at least one series of interest is subject to unknown aggregation weights. Such a situation may arise with aggregated pre-filtered data, such as seasonally adjusted data. Even if the aggregation weights on the filtered data are known, the implied weights on the unfiltered high-frequency series might not be known.

This paper is also part of a growing literature on mixed frequency data models. Early contributions include Phillips (1976) who studied structural estimation of continuous models and provided a number of empirical applications involving aggregation and mixed data sampling issues in a discrete time VARX model corresponding to a linear stochastic differential equation with exogenous variable inputs. This work by Phillips was a precursor to later work by Zadrozny (1988) who studied Gaussian MLE of continuous time ARMAX models when data are stocks and flows observed at different frequencies.

In order to solve the size distortion problems, we propose novel ways to use available high-frequency (henceforth HF) information or to bootstrap low-frequency (henceforth LF) tests when no HF information is available. Recall that price and GDP series are actually available at different frequencies, and the HF price information may be used to our advantage. Instead of running a LF trace test, we propose a mixed-frequency (henceforth MF) trace test, which has less size distortion, because fewer series are subject to aggregation.

To this end, we rely on a MF vector autoregressive (henceforth VAR) model and, in particular, a vector error correction model (henceforth VECM) to implement the new class of trace tests. VAR models for MF data have been studied by Anderson et al. (2012) and Ghysels (2014). The latter also discusses the fact that MF VAR models can be viewed as

---

3Our paper has some relationship with the impact of seasonal adjustment filters on cointegration tests. First, seasonal adjustment filters affect the power of unit root tests, as noted by Ghysels (1990), among others, whereas Ericsson, Hendry, and Tran (1987) study the consequences of such filters on tests for cointegration.

4State 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. Seong et al. (2013) analyzed a cointegrated VAR in this context.
a multivariate extension of MIDAS regressions.\(^5\)

We also propose MF residual-based tests, as opposed to LF residual-based tests involving aggregate HF data. Residual-based tests using least squares estimation of a mixed-frequency regression suffer from severe size distortion resulting from the inconsistency of the aggregation weight estimators under the null. We show that nonlinear cointegrating MIDAS (CoMIDAS) regressions studied by Miller (2014b) work surprisingly well to mitigate size distortion.

When a MF approach is not possible, we show that a sieve bootstrap similar to that of Palm et al. (2010) but with either the LF residual-based tests considered in this paper or the trace test procedure of Cavaliere et al. (2012) applied to LF data controls size well.

The rest of the paper is organized as follows. In Section 2, we analyze VECMs and the associated trace tests for cointegration in HF (i.e., single-frequency), LF (i.e., single-frequency but subject to aggregation), and MF environments. We analyze the single-equation residual-based tests in these environments in Section 3. Because size distortion may be particularly bad in the LF environment, we present alternative LF and MF testing strategies in Section 4. Section 5 contains simulation results and comparisons, Section 6 contains an empirical application to stock prices and dividends, and Section 7 concludes. An online appendix contains proofs of the theoretical results.

## 2 Testing Vector Error Correction Models

Consider a \( p \)-variate HF data-generating process (DGP) given by the VECM

\[
\Delta^{(1/m)} y_{t-i/m} = \Gamma A'y_{t-(i+1)/m} + \varepsilon_{t-i/m},
\]

\(^5\)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. Foroni et al. (2013) provide a survey of mixed frequency VAR models and related literature. Anderson et al. (2014) provide a survey of the literature on estimating parameters of high frequency models with mixed frequency data.
where $\Delta^{(1/m)}$ is the HF difference operator, $\Gamma$ and $A$ are both $p \times r$ matrices of rank $r$ such that $0 \leq r \leq p$, and $\varepsilon_{t-i/m} | F_{t-(i+1)/m} \sim (0, \Sigma)$. We consider samples with $T$ LF observations, indexed by $t$, but we assume that the data are generated $m < \infty$ times more often, so that $i = 0, ..., m - 1$. We let $M$ denote the HF sample size and assume that $m = M/T$.\(^6\)

If $r = p$, then $\Gamma A'$ is full rank and $(y_{t-i/m})$ is I(0). On the other hand, if $r = 0$, then $\Gamma A' = 0$ and $(y_{t-i/m})$ is I(1) with $p$ distinct stochastic trends. In the intermediate case of $0 < r < p$, $(y_{t-i/m})$ is I(1) and cointegrated by $r$ linearly independent cointegrating relationships with $p - r$ stochastic trends remaining. Following the lead of Cheung and Lai (1993) and other authors, because our goal of examining size distortions in a relatively simple model is similar to theirs, we consider the null $H_0 : r = 0$ against the full rank alternative. The null model is therefore simply $\Delta^{(1/m)} y_{t-i/m} = \varepsilon_{t-i/m}$.

We assume an invariance principle of the form $M^{-1/2} \sum_{j=1}^{[rM]} \varepsilon_{j/m} \rightarrow_d B(r)$ as $M \rightarrow \infty$, where $B$ is a vector Brownian motion with variance $\Sigma$. Thus, $B = \Sigma^{1/2} W$, where $W$ is a vector of independent standard Brownian motions. Under the null model, $M^{-1/2} y_{t-i/m} \rightarrow_d B(r)$ if $y_0 = 0$, as is often assumed.

Analyses of VECMs typically allow two additional characteristics: deterministic trends and higher-order autoregressions. It should be emphasized that the aim of this paper is to show the effects of aggregation and mixed frequencies on an otherwise simple model. We fully expect that including deterministic trends would affect the limiting distribution, as it does in a single-frequency VECM. However, because deterministic trends may be “observed” at any frequency, because adding such trends should not change any of the results qualitatively, and because the additional notation becomes quite cumbersome, we do not explicitly consider them here.\(^7\)

\(^6\)If $M/T$ is not an integer, then $m$ is the greatest integer not exceeding $M/T$, and we may expect all of our asymptotic results to hold since $m < \infty$. See Chambers (2011) for analysis of single-equation cointegration models as $m \rightarrow \infty$.

\(^7\)We conducted a full set simulations (including the bootstrap and MIDAS approaches discussed below) on a DGP with an intercept and on a DGP with an intercept and a linear trend. In order to save space, we do not report the numerical results, but we note that they were indeed qualitatively similar to those we
Higher-order autoregressions are typically allowed by including lags of $\Delta^{(1/m)}y_{t-i/m}$ on the right-hand side. In that case, $\Delta^{(1/m)}y_{t-i/m}$ and $y_{t-(i+1)/m}$ may be replaced by residuals from regressing these onto the lagged differences, but the limiting distribution of the trace test statistic does not change. There is an additional reason for considering a higher-order VECM: aggregation itself may create first-order moving average (MA) serial correlation, as Working (1960) showed for the special case of flat sampling a univariate series.

After we analyze the size distortion resulting from this serial correlation, we subsequently consider increasing the lag order from zero in the tests in order to ameliorate that distortion. We expect that similar results would hold if we started with a non-zero number of lags in the DGP and increased the lag order in the tests from that number, but leaving out lags in the DGP avoids additional notation.

### 2.1 A Motivating Example

To understand more clearly the source of the serial correlation and thus of the size distortion, consider a univariate ($p = 1$) series $(y_{t-i/m})$ with $m = 2$. If $\Delta^{(1/m)}y_{t-i/2} = \varepsilon_{t-i/2}$ with $\varepsilon_{t-i/2} | \mathcal{F}_{t-(i+1)/2} \sim (0, \sigma^2)$, a random walk at the HF, we may write

$$
\Delta z_t \equiv \begin{bmatrix}
\Delta y_t \\
\Delta y_{t-1/2}
\end{bmatrix} = \begin{bmatrix}
\varepsilon_t \\
\varepsilon_{t-1/2}
\end{bmatrix} + \begin{bmatrix}
\varepsilon_{t-1/2} \\
\varepsilon_{t-1}
\end{bmatrix} \equiv u_t + u_{t-1/2},
$$

or, equivalently,

$$
\Delta z_t = u^*_{t} - \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} u^*_{t-1} \quad \text{with} \quad u^*_t \equiv \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} u_t,
$$

to see that the LF difference induces first-order MA serial correlation.
Aggregating the series \((y_{t-i/2})\) to the LF yields

\[
\Delta y_t^a = \begin{bmatrix}
\varpi_1 \\
\varpi_1 + \varpi_2
\end{bmatrix}' \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1/2} 
\end{bmatrix} + \begin{bmatrix} \varpi_2 \\ 0
\end{bmatrix}' \begin{bmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-3/2} 
\end{bmatrix} = \begin{bmatrix}
\varpi_1 \\
\varpi_1 + \varpi_2
\end{bmatrix}' u_t + \begin{bmatrix} \varpi_2 \\ 0
\end{bmatrix}' u_{t-1}
\]

where \(\varpi\) is an aggregation weight vector for a general aggregation scheme.

If the aggregation scheme is EOP (end-of-period) sampling, so that \(\varpi_1 = 1\) and \(\varpi_2 = 0\), then \(\Delta y_t^a = \varepsilon_t + \varepsilon_{t-1/2}\), which is clearly not serially correlated at the LF. Likewise, if the aggregation scheme is BOP (beginning-of-period) sampling, such that \(\varpi_1 = 0\) and \(\varpi_2 = 1\), then \(\Delta y_t^a = \varepsilon_{t-1/2} + \varepsilon_{t-1}\), which is not serially correlated at the LF either.

On the other hand, if the aggregation scheme is flat sampling, such that \(\varpi_1 = \varpi_2 = 1/2\), then \(\Delta y_t^a = 1/2\varepsilon_t + \varepsilon_{t-1/2} + 1/2\varepsilon_{t-1}\), which is serially correlated at the LF – one of Working’s (1960) well-known results. Specifically, the first-order autocovariance is \(\mathbf{E}\Delta y_t^a \Delta y_{t-1}^a = \sigma^2/4\). Of course, such serial correlation extends to more general and possibly unknown weighting schemes.

Now consider an extension of this simple example to a second series, so that \(\Delta (1/m)y_{t-i/2} = \varepsilon_{t-i/2}\) is bivariate \((p = 2)\) with \(\varepsilon_{t-i/2} | \mathcal{F}_{t-(i+1)/2} \sim (0, \Sigma)\). Suppose both series are aggregated using flat sampling. It is straightforward to see that \(\mathbf{E}\Delta y_t^a \Delta y_{t-1}^a = (1/4)\Sigma\) in that case.

Suppose instead that the first series is EOP sampled, but the second is BOP sampled. In that case, \(\Delta y_t^a = (\varepsilon_t + \varepsilon_{1, t-1/2}, \varepsilon_{2, t-1/2} + \varepsilon_{2, t-1})'\), so that \(\mathbf{E}\Delta y_t^a \Delta y_{t-1}^a\) is a \(2 \times 2\) matrix with \(\sigma_{21}\) in the lower left corner and zeros elsewhere. Note that the serial correlation results from cross-correlation \(\sigma_{21}\) of the errors. Unless \(\sigma_{21}\) happens to be zero, serial correlation is present in the bivariate series \((\Delta y_t^a)\) even though the aggregation of the individual series creates no serial correlation.

In light of the preceding two examples, we distinguish two types of size distortion in the following set of definitions.
**Definition:** Type A size distortion results from running tests on multiple series subject to identical aggregation schemes.

**Definition:** Type B size distortion results from running tests on multiple series subject to different aggregation schemes.

In the above examples, the covariance \((1/4)\Sigma\) results in Type A size distortion, while \(\sigma_{21}\) results in Type B size distortion. Our numerical results below suggest that Type B size distortion is generally worse than Type A when \(\sigma_{21} \neq 0\).

We expect that increasing or decreasing \(m\) will have some effect on size distortion. Indeed, Working (1960) derives an exact expression for this serial correlation as a function of \(m\) in the special case of a univariate flat-sampled series. Working’s expression, \((m^2 - 1)/(2 (2m^2 + 1))\), for the autocorrelation in that case suggests that there is a limiting autocorrelation of 1/4 as \(m\) increases, which suggests an upper bound on size distortion. Generalizing this expression for the autocorrelation as a function of \(m\) to multiple series subject to different sampling schemes does not appear to be as straightforward, and we leave this topic for future research.

### 2.2 Low-Frequency VECM

Now suppose that the analyst observes \(y^{i}_{st} = \omega^{'}_s z_{st}\) instead of \(y_{st}\), where \(\omega_s = (\omega_{s1}, ..., \omega_{sm})^{'}\) is a vector of \(m\) non-negative deterministic aggregation weights such that \(\omega^{'}_s = 1\) (the weights sum to one) for \(s = 1, \ldots, p\). (Throughout the paper, we use \(\iota\) to denote a unit vector of length given by the context in which it is used.)

As in the example above, we combine the \(p\)-vectors \(y_{t-i/m}\) for each \(i = 0, ..., m-1\) representing HF observations into a single \(mp\)-vector \(z_t\) observed every LF period. To this end, define \(z_{st} \equiv (y_{st}, y_{s,t-1/m}, ..., y_{s,t-(m-1)/m})^{'}\), \(u_{st} \equiv (\epsilon_{st}, \epsilon_{s,t-1/m}, ..., \epsilon_{s,t-(m-1)/m})^{'}\), \(z_t \equiv (z_{1t}, ..., z_{pt})^{'}\), and \(u_t \equiv (u'_{1t}, ..., u'_{pt})^{'}\). We then temporally aggregate \(z_t\) to obtain a series \((y'^{i}_t)\) of \(p\)-vectors. In order to do this, define \(\Pi_a \equiv (\omega_1^{'} \oplus \cdot \cdot \cdot \oplus \omega_p^{'})\) to be a \(p \times mp\)
aggregation matrix, where the operator $\oplus$ (direct sum) is such that $C \oplus D$ creates a block diagonal matrix with diagonal blocks given by $C$ and $D$.

**Proposition 1.** The LF difference of the mp-vector $z_t$ may be rewritten as

$$\Delta z_t = (\Gamma A' \otimes I_m) z_{t-1} + \eta_t,$$

where $\eta_t \equiv (I + (\Gamma A' \otimes I_m)) \sum_{i=1}^{m-1} \Delta^{(1/m)} z_{t-i/m} + u_t$ with $\text{var}(u_t) = \Sigma \otimes I_m$. Temporally aggregating the HF system in (1) yields a LF system given by

$$\Delta y^a_t = \Gamma A' y^a_{t-1} + \eta^a_t,$$

where $y^a_t \equiv \Pi_a z_t$ and $\eta^a_t \equiv \Pi_a \eta_t$.

The matrices $\Gamma$ and $A$ in the aggregated system in (5) are the same as those in the VECM in (1). Temporal aggregation does not change the cointegrating relationships (columns of $A$), but it may substantially affect the short-run properties of the series, as noted by Marcellino (1999) *inter alia* and seen in the example above.

Under the null, recall that $\Gamma A' = 0$ so that (4) reduces to $\Delta z_t = \sum_{i=0}^{m-1} u_{t-i/m}$, which simply extends the case of $m = 2$ in (2) and suffers from first-order serial correlation at the LF as in (3).

### 2.3 High-Frequency and Low-Frequency Trace Tests

If we could observe the HF series directly, Johansen’s (1988) trace test would have the expected limiting distribution in (6) below, since the HF data are not subject to any aggregation. We consider this to be an infeasible benchmark against which we measure the LF and later MF trace tests.

We define $S_{gh} \equiv M^{-1} \sum_t \sum_{i=0}^{m-1} s_{g,t-i/m}s'_{h,t-i/m}$ for $g, h = 0, 1$ with $s_{0,t-i/m} \equiv \Delta^{(1/m)} y_{t-i/m}$ and $s_{1,t-i/m} \equiv y_{t-(i+1)/m}$. For observable HF series, Johansen’s (1988) trace test of
$H_0: r = r_0$ against $H_A: r = p$ (all series $I(0)$, no cointegration) is implemented as

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

where $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ are ordered from largest to smallest and may be found by solving the determinantal equation $|\lambda I - S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}| = 0$.

Using properties of the trace and an expansion of the logarithm around 1,

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

where $o_p(1)$ represents higher-order terms in the expansion of the logarithm that collapse to zero faster than the first term as $M \to \infty$. See Johansen (1995) for a very detailed discussion. Further references to the trace test statistic may be understood to refer either to $-2 \log Q(H_{r_0|p})$ or to its asymptotic approximation $M \text{tr}\{\bullet\}$.

Under the null of $H_0: r = 0$, 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 $\Sigma$, the test has a limiting null distribution given by

$$M \text{tr}\{S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}\} \to_d \text{tr}\{\int (dW)'W\left(\int WW'\right)^{-1}\int WdW'\}$$

as $M \to \infty$. The number of stochastic trends $p$ under the null is the only nuisance parameter.\footnote{Extending to high-order nulls with $p - r$ stochastic trends 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 and is analogous to the null in the single-equation case (no cointegration), even though the tests have different alternatives.}

The aggregated LF system in (5) contains the same number $p$ of series as the HF VECM. The LF trace test statistic becomes $T \text{tr}\{(R_{11}^a)^{-1}R_{10}^a (R_{00}^a)^{-1}R_{01}^a\}$, where $R_{gh}^a \equiv \Pi_{a} R_{gh} \Pi_{a}'$ for $R_{gh} \equiv T^{-1} \sum r_{gt}r_{ht}'$ with $r_{0t} \equiv \Delta z_t$ and $r_{1t} \equiv z_{t-1}$ for $g, h = 0, 1$, analogously to $S_{gh}$ above. Its implementation using $|\lambda I - (R_{11}^a)^{-1}R_{10}^a (R_{00}^a)^{-1}R_{01}^a| = 0$ is completely analogous.
to the HF case. No modification to the procedure is necessary, because the procedure is
run on \((\triangle y_t^a)\) in place of \((\triangle^{(1/m)}y_{t-i/m})\). The moments \(R^{a}_{00} = T^{-1} \sum \triangle y_t^a \triangle y_t^{a'}\), etc., are
defined in terms of the aggregated series and use the smaller sample size associated with
this series.

We assumed an invariance principle, \(M^{-1/2} \sum_{j=1}^{[rM]} \varepsilon_{j/m} \rightarrow_d B(r)\), to hold at the HF
and that \(\triangle^{(1/m)}y_{t-i/m} = \varepsilon_{t-i/m}\) under the null. At the LF, we observe \(\varepsilon_t^\prime \triangle z_t\) instead of
\(\triangle^{(1/m)}y_{t-i/m}\), and \(R^{a}_{gh}\) is defined in terms of \(\triangle z_t\). In turn, \(\triangle z_t\) is defined in terms of sums
of HF error of the form \(\sum_{i=0}^{m-1} \varepsilon_{t-(i+j)/m}\). To get a LF invariance principle, we may write
\[
T^{-1/2} \sum_{t=1}^{[rT]} \left( \sum_{i=0}^{m-1} \varepsilon_{t-(i+j)/m} \right) = T^{-1/2} \sum_{t=1}^{[rT]} \left( \sum_{i=0}^{m-1} \varepsilon_{t-i/m} \right) + o_p(1) \rightarrow_d m^{1/2} B(r),
\]
where the \(o_p(1)\) follows because \((\varepsilon_{t-i/m} - \varepsilon_{t-(i+j)/m})\) has a long-run variance of zero and
the limit follows from the HF invariance principle. (See also Lemma 2 of Miller, 2014a.)

The limiting distribution of the LF trace test statistic is given in the following theorem.

We first define \(H\) to be an \(m \times m\) matrix with all elements below the main diagonal set to be
ones and all those on and above the main diagonal set to be zeros. From \(H\), we define \(H_{00} \equiv (I_m + H') (I_m + H) + HH', H_{01} \equiv H (I_m + H), H_{10} = H_{01}', \) and \(Z^{a}_{00} \equiv \Pi \Pi (I_p \otimes m^{-1} H_{00}) \Pi\)
for \(i = 0, 1, \ldots, p\).

\textbf{Theorem 2.} \textit{Consider the null hypothesis of no cointegrating relationships: }\(r_0 = 0\) \textit{in the}
LF VECM in \((5)\). The trace test statistic }\(\text{tr}\left\{ (R^{a}_{00})^{-1} R^{a}_{10} (R^{a}_{00})^{-1} R^{a}_{01} \right\} \text{ has an asymptotic}
null distribution coinciding with
\[
\text{tr}\left\{ (\int (dW)^\prime W + Z^{a}_{10}) (\int W W'^{-1})(\int W dW' + Z^{a}_{10}) (Z^{a}_{00})^{-1} \right\}
\]
as \(T \rightarrow \infty.\)

\footnote{Specifically, the elements of }\(\varepsilon_{t-(i+j)/m}\) \textit{are the elements }\(j + 1, m + j + 1, 2m + j + 1, \ldots, (p-1)m + j + 1\)
of the \(mp \times 1\) vector \(\triangle^{(1/m)}z_{t-i/m}\), and \(\triangle z_t = \sum_{i=0}^{m-1} \triangle^{(1/m)}z_{t-i/m}\).
Note that the benchmark distribution in (6) is generally not obtained, and some size distortion may be expected. It is obtained as a special case when \( m^{-1} \varpi'_{s} H_{00} \varpi_{u} = 1 \) and \( m^{-1} \varpi'_{s} H_{10} \varpi_{u} = 0 \) for all \( s, u = 1, \ldots, p \), so that \( Z^0_{00} = I_m \) and \( Z^a_{10} = 0_m \). Some algebra reveals that the diagonal elements of \( H_{00} \) defined above are given by \( m \) and that those of \( H_{10} \) are given by 0. Hence, \( \varpi_{s} \) must be a binary vector with a unit in any element and zeros elsewhere for all \( s = 1, \ldots, p \) in order for \( Z^0_{00} = I_m \) and \( Z^a_{10} = 0_m \) to hold. In practical terms, this means that the standard critical values are valid if all series are skip sampled in the same way.

Suppose that all series are skip sampled, but that some are skip sampled differently from others – e.g., some are EOP sampled and some are BOP sampled. Although there is no serial correlation in any of the individual series (\( \Delta y^a_{st} \)), there is serial correlation in the vector (\( \Delta y^a_t \)) causing type B size distortion by its definition, as in the bivariate example in Section 2.1.

### 2.4 Mixed-Frequency VECM

Moving to a system of equations with series observed at different sampling frequencies, we assume that \( p_l \) series in (\( \Delta^{(1/m)} y_{t-i/m} \)) are aggregated and observed at the LF, while \( p_h \) series are observed at the HF, with \( p_l + p_h = p \). Because the implementation of the test is invariant with respect to the order of the series, we let the first \( p_l \) series in (\( \Delta^{(1/m)} y_{t-i/m} \)) be those observed only at the LF without any loss of generality.

The matrix \( \Pi_m \) defined by

\[
\Pi_m \equiv \begin{bmatrix}
\varpi'_{1} \oplus \cdots \oplus \varpi'_{p_l} & 0 \\
0 & I_{mp_h}
\end{bmatrix}
\]

is a \( (p_l + mp_h) \times mp \) partial aggregation matrix. Premultiplying \( z_t \) by \( \Pi_m \) temporally aggregates the HF observations of the first \( p_l \) series in (\( \Delta^{(1/m)} y_{t-i/m} \)) but leaves the remaining series unaggregated. The resulting series (\( z^m_t \)), defined by \( z^m_t \equiv \Pi_m z_t \), contains \( p_l \) LF series.
The remaining \( mp_h \) “low-frequency” series in \((z^m_t)\) are just \( p_h \) distinct HF series observed at \( m \) different LF intervals. The “low-frequency” series created in this way from the same HF series are mutually cointegrated by construction, so that these \( mp_h \) series have no fewer than \((m-1)p_h\) cointegrating relationships – i.e., no more than \( p_h \) stochastic trends. 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 for distinct months. The last 12 series may not have more than one stochastic trend.

**Proposition 3.** A MF VECM equivalent to the HF system in (1) is

\[
\Delta z^m_t = \Gamma^m A^m z^m_{t-1} + \eta^m_t, \quad (8)
\]

where \( \eta^m_t \equiv \Pi_m \eta_t \) and where \( \Gamma^m \) and \( A^m \) are \((p_l + mp_h) \times r\) matrices. The set of eigenvalues of \( \Gamma^m A^m' \) include those of \( \Gamma A' \) and \((m-1)p_h\) additional zeros.

The coefficient matrix \((\Gamma A' \otimes I_m)\) in (4) is \( mp \times mp \) with a rank of \( mr \). The fact that its rank must be a multiple of \( m \) requires \((m-1)p\) restrictions to be imposed. Premultiplying \((\Gamma A' \otimes I_m)\) by the \((p_l + mp_h) \times mp\) matrix \( \Pi_m \) to create \( \Delta z^m_t \) imposes \((m-1)p_l\) restrictions, but \((m-1)p_h\) remain to be imposed. The \((m-1)p_h\) known cointegrating relationships in \( (z^m_{t-1}) \) may be identified similarly by postmultiplying the cointegrating matrix \( A' \), and the matrix \( A^m \) in the MF VECM in (8) reflects this identification. See the proof of the proposition in the online appendix for more details.

### 2.5 Mixed-Frequency Trace Test

The MF model contains \( p_l + mp_h \) series, but \((m-1)p_h\) cointegrating relationships are already known. A hypothesis of \( r_0 = 0 \) cointegrating relationships in the HF VECM corresponds to a hypothesis of \((m-1)p_h\) cointegrating relationships in the MF VECM, and both imply \( p \) stochastic trends. More generally, a hypothesis of \( r_0 \) corresponds to \( r_0 + (m-1)p_h \)
cointegrating relationships in the MF VECM, implying \( p - r_0 \) stochastic trends.

The MF trace test is therefore given by

\[
-2 \log Q(H_{r_0+(m-1)p_h}|p_t+m_{ph}) = -T \sum_{i=r_0+(m-1)p_h+1}^{p_t+m_{ph}} \log(1 - \hat{\lambda}_i) \\
= T \ \text{tr}\{(R_{11}^m)^{-1}R_{10}^m(R_{00}^m)^{-1}R_{01}^m\} + o_p(1)
\]

where \( R_{gh}^m \equiv \Pi_m R_{gh} \Pi'_m \) for \( g, h = 0, 1 \) and \( \hat{\lambda}_1, \ldots, \hat{\lambda}_{p_t+m_{ph}} \) (ordered from largest to smallest) solve the determinantal equation \( |\lambda I - (R_{11}^m)^{-1}R_{10}^m(R_{00}^m)^{-1}R_{01}^m| = 0 \).

The null of no cointegrating vectors in the HF VECM is equivalent to the hypothesis that the smallest \( p \) roots are zero in all three (HF, LF, MF) cases. These are the only roots in the HF and LF cases. In the MF case, the determinantal equation has additional \((m-1)p_h\) (larger) roots that are not used in the test statistic.

**Theorem 4.** Consider the null hypothesis of \((m-1)p_h\) cointegrating relationships in the MF system, which is equivalent to \( r_0 = 0 \) cointegrating relationships in the HF and LF systems. The trace test statistic \( T \ \text{tr}\{(R_{11}^m)^{-1}R_{10}^m(R_{00}^m)^{-1}R_{01}^m\} \) based on the MF VECM in (8) has an asymptotic null distribution coinciding with \( \text{tr}\{\Xi_{10}^m(\Xi_{11}^m)^{-1}\Xi_{00}^m(\Xi_{00}^m)^{-1}\} \), where

\[
\Xi_{00}^m \equiv \Pi_m (\Sigma \otimes m^{-1}H_{00}) \Pi'_m \\
\Xi_{11}^m \equiv \Pi_m (\Sigma^{1/2} \int WW'\Sigma^{1/2} \otimes \mu') \Pi'_m \\
\Xi_{10}^m \equiv \Pi_m ((\Sigma^{1/2} \int WdW'\Sigma^{1/2} \otimes \mu') + (\Sigma \otimes m^{-1}H_{10})) \Pi'_m
\]

as \( T \to \infty \).

The limiting distribution does not easily simplify in a way that lends itself to a direct comparison with the benchmark distribution in (6). Instead, we rely on simulations to compare it with the benchmark distribution and the LF distribution in equation (7).
3 Testing Single-Equation Models

Suppose now that \( r \leq 1 \), so that either \( r = 1 \) and there is only one cointegrating relationship, or \( r = 0 \) and \( (y_{t-i/m}) \) has \( p \) linearly independent stochastic trends (no cointegration). The DGP may be written as \( \Delta^{(1/m)}y_{t-i/m} = \gamma \alpha' y_{t-(i+1)/m} + \varepsilon_{t-i/m} \) with \( p \)-vectors \( \gamma \) and \( \alpha \), and a residual-based test for cointegration along the lines of Engle and Granger (1987) and Phillips and Ouliaris (1990) is simply a unit root test of the fitted residuals \( \hat{\varepsilon}_{t-i/m} = \hat{\alpha}' y_{t-i/m} \), where one element of \( \alpha \) is normalized to unity, so that \( \alpha = (1, -\beta')' \). \( \hat{\alpha} = (1, -\hat{\beta}')' \) and \( \hat{\beta} \) is the least squares estimator of \( \beta \) from regressing the first series in \( (y_{t-i/m}) \) onto the remaining \( p-1 \) series.

Under the null, \( \gamma = 0 \), so that \( \Delta^{(1/m)}y_{t-i/m} = \varepsilon_{t-i/m} \) and \( \varepsilon_{t-i/m} = \alpha' \sum_{j=1}^{m-t-i} \varepsilon_{j/m} \) is I(1). Recall from the assumed invariance principle that \( M^{-1/2} \sum_{j=1}^{[rM]} \varepsilon_{j/m} \to_d B(r) \), so it will be useful to partition \( B(r) = (B_1(r), B_2(r))' \) in the same way as \( \alpha \). Specifically, \( B_1(r) \) is a scalar Brownian motion and \( B_2(r) \) is a \((p-1)\)-vector of Brownian motion. It follows from the invariance principle and continuous mapping theorem that \( \hat{\beta} \to_d (\int B_2 B_2')^{-1} \int B_2 B_1 \).

In the absence of serial correlation in \((\varepsilon_{t-i/m})\), the Dickey-Fuller (DF) unit root tests utilize a regression of \( \Delta^{(1/m)} \hat{\varepsilon}_{t-i/m} \) onto \( \hat{\varepsilon}_{t-(i+1)/m} \) with a null that the coefficient is zero (no cointegration). The DF coefficient test and \( t \)-test may be written as \( \rho_M = M(\hat{\alpha}' S_{11} \hat{\alpha})^{-1} \hat{\alpha}' S_{10} \hat{\alpha} \) and \( \tau_M = (\hat{\alpha}' S_{00} \hat{\alpha} M^{-1} S_{11} \hat{\alpha})^{-1/2} \hat{\alpha}' S_{10} \hat{\alpha} \) up to an \( o_p(1) \) term in \( \tau_M \).

Phillips and Ouliaris (1990) show that the coefficient and \( t \) tests have limiting distributions given by

\[
\rho_M \to_d \left( \int Q^2 \right)^{-1} \int QdQ \quad \text{and} \quad \tau_M \to_d \left( \kappa' \kappa \int Q^2 \right)^{-1/2} \int QdQ, \tag{9}
\]

where \( Q(r) \equiv \kappa' W(r) \) and \( \kappa' \equiv (1, -\int W_1 W_2 (\int W_2 W_2')^{-1}) \) with \( W = (W_1, W_2)' \) partitioned like \( B \). The distributions in (9) are the HF benchmarks in the sense that the textbook of this term comes from the fact that \( \hat{\alpha}' S_{00} \hat{\alpha} - \hat{\alpha}' S_{01} \hat{\alpha} (\hat{\alpha}' S_{11} \hat{\alpha})^{-1} \hat{\alpha}' S_{10} \hat{\alpha} = \hat{\alpha}' S_{00} \hat{\alpha} + o_p(1) \), and we suppress the remainder term in this and subsequent \( t \) tests. Ignoring this term is equivalent to imposing the null coefficient in the variance of the estimator instead of estimating it.
critical values for these tests are drawn from them.

3.1 Low-Frequency Model and Residual-Based Tests

The aggregated LF regression may be written as

\[ y_{1t}^a = (y_{2t}^a, ..., y_{pt}^a) \beta + e_t^a, \]  

(10)

where \( y_{st}^a = \omega_s' z_{st} \) is a scalar aggregate of the vector \( z_{st} \) for series \( s = 1, ..., p \), as above. The residual-based testing strategy is adapted to be a unit root test of the LF fitted residuals \( \hat{e}_t^a \) instead of \( \hat{e}_{t-i/m} \). 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 \( \tau_T^a \equiv (\hat{\alpha}' R_{00}^a \hat{\alpha}' T^{-1} R_1^a \hat{\alpha})^{-1/2} \hat{\alpha}' R_{10}^a \hat{\alpha} \), using the same notation \( R_{00}^a \), etc., defined above for the VECM. Calculating these statistics poses no additional computation complications – they are simply unit root tests on the fitted residuals of (10).

The following theorem shows the limiting distributions of these test statistics.

**Theorem 5.** Consider the null hypothesis of no cointegrating relationships: \( r_0 = 0 \). The coefficient test and \( t \)-test calculated from the fitted residuals of the LF regression in (10) have limiting null distributions given by

\[
\begin{align*}
\rho_T^a \to_d & \left( \int Q^2 \right)^{-1} \left( \int QdQ + \kappa' Z_{10}^a \kappa \right) \\
\tau_T^a \to_d & \left( \kappa' Z_{00}^a \kappa \int Q^2 \right)^{-1/2} \left( \int QdQ + \kappa' Z_{10}^a \kappa \right)
\end{align*}
\]

as \( T \to \infty \).

The notations \( Z_{00}^a \) and \( Z_{10}^a \) are the same as those used in (7).

As in the case of an aggregated VECM, the benchmark distributions in the single-equation case are obtained when \( m^{-1} \omega_s' H_{00} \omega_u = 1 \) and \( m^{-1} \omega_s' H_{10} \omega_u = 0 \) for all \( s, u = \)
...p. In other words, size is not distorted when all series are skip sampled in the same way.\textsuperscript{11}

### 3.2 Mixed-Frequency Model and Residual-Based Tests

The MF model may be expressed as

\[ y_{1t}^a = (y_{2t}^a, \ldots, y_{p_{l+1}^t}, \varpi_{p_{l+1}^t+1}^t; \ldots, \varpi_{p_{l}^t}^t z_{p_{l}^t}) \beta + \epsilon_t^a \]  

(11)

where \( \varpi_s \) for \( s = p_{l} + 1, \ldots, p \) are weights to be estimated in the MF model. The model is similar to the MF cointegrating regression considered by Miller (2014a), except that the regression is spurious under the null hypothesis of the tests.

It should be clear that \( \varpi_s^t z_{st} \) is a sum of \( m \) regressors with a common stochastic trend for each \( s = p_{l} + 1, \ldots, p \). In other words, the MF model has \( p_{l} + mp_{h} - 1 \) regressors that share no more than \( p - 1 \) stochastic trends. We isolate the stochastic trends by imposing the \( m - 1 \) known cointegrating relationships in the \( m \)-vector \( z_{st} \).

Recall that \( H \) is an \( m \times m \) matrix with ones below the main diagonal and zeros on and above the main diagonal. We may write \( z_{st} = y_{st} - H \Delta^{(1/m)} z_{st} \) so that \( z_{st} \) is expressed as a linear combination of an \( m \)-vector \( y_{st} \) of a single common trend \( y_{st} \) and deviations from that trend given by sums of HF differences \( H \Delta^{(1/m)} z_{st} \). Since the aggregation weights sum to one, the weighted series \( \varpi_s^t z_{st} \) may be written as \( y_{st} - \varpi_s^t H \Delta^{(1/m)} z_{st} \). Now, instead of \( m \) cointegrated terms in \( \varpi_s^t z_{st} \), \( y_{st} - \varpi_s^t H \Delta^{(1/m)} z_{st} \) is an equivalent expression with a single I(1) term and \( m - 1 \) I(0) terms created by imposing the \( m - 1 \) known cointegrating relationships.

\textsuperscript{11}These results are the opposite of those expected from the literature on efficient estimation of the cointegrating vector, in which Chambers (2003) and Miller (2014a) suggest flat sampling all series, if possible.
Rewriting $\varpi_s' z_{st}$ for each $s = p_l + 1, \ldots, p$ in this way in (11) yields a model given by

$$y_{1t}^a = (y_{2t}^a, \ldots, y_{p_l+1,t}, \ldots, y_{pt}) \beta - (0, \ldots, 0, \varpi_{p_l+1} H \Delta^{(1/m)} z_{p_l+1,t}, \ldots, \varpi_p H \Delta^{(1/m)} z_{pt}) \beta + e_t^a,$$

such that the vector $(y_{2t}^a, \ldots, y_{p_l+1,t}, \ldots, y_{pt})$ captures all of the stochastic trends in the model, and the remaining terms are I(0) and contain all of the aggregation weights to be estimated.

In practice, the weight vectors $\varpi_s$ for $s = p_l + 1, \ldots, p$ on the HF observations may be fixed or estimated. If fixed, then the analyst simply aggregates the remaining HF observations to the LF, resulting in the LF model in (10), with the same possibilities for size distortion.

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 $\tau_T^m \equiv (\hat{\alpha}' R_{00}^m \hat{\alpha} \hat{\alpha}' T^{-1} R_{11}^m \hat{\alpha})^{-1/2} \hat{\alpha}' R_{10}^m \hat{\alpha}$ with $R_{00}^m$, etc., defined as above for the MF trace test.

**Theorem 6.** Consider the null hypothesis of no cointegrating relationships: $r_0 = 0$. The coefficient test and $t$-test calculated from the fitted residuals of the MF model in (12) have limiting distributions null given by $\rho_T^m \to_d O_p(1)$ and $\tau_T^m \to_d O_p(1)$ as $T \to \infty$. The point of the $O_p(1)$ terms is that they are not easily expressed in terms of the benchmark limits in (9). In fact, tests based on the residuals from a MF model generate additional size distortion that cannot be zero. 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 (12) is spurious and these weights are inconsistently estimated.

The only restriction imposed on the aggregation weights by estimating the model in (12) with least squares is that they sum to one. The lack of any further restriction means that the fitted residuals in (10) used in the test statistics are sensitive to inconsistent estimates of the aggregation weights. Although it seems counterintuitive, it may be preferable to impose
arbitrary aggregation weights using the LF regression in (10) to inconsistently estimating them without restriction.

4 Alternative Testing Procedures

4.1 Alternatives to the Trace Tests

Up to this point, we have presented two options for the trace test in the presence of MF data: either ignore the HF data and run a LF trace test or construct a MF trace test. Our theoretical results (and simulations below) suggest that the MF trace test dominates the LF trace test in terms of size, simply because fewer series are subject to aggregation in a MF model. However, size control may also be achieved using the LF data without relying on the MF data.

In principle, when we know precisely how all of the LF data were aggregated, we may derive asymptotically valid corrections to the LF trace test. Ghysels and Miller (2014) derived such corrections for HF data subject to linear interpolation from end-of-period sampled LF series. They found that the rejection rates were quite sensitive to the exact aggregation and interpolation schemes, which may be unknown. Size could be more generally controlled simply by allowing for additional lags in the VECM to control for the first-order MA serial correlation in the residuals.

As a further option, a bootstrapping scheme that allows for serial correlation in the LF residuals could be used to generate more appropriate critical values. Cavaliere et al. (2012) propose an iid bootstrap for the trace test, but a sieve bootstrap may be used to control for serial correlation of possibly unknown form. We consider a sieve bootstrap procedure similar to that of Palm et al. (2010), but substituting the steps relevant to the trace test from the procedure of Cavaliere et al. (2012) in place of the single-equation test that the latter authors analyze.\(^\text{12}\)

\(^{12}\text{In particular, we use residuals and parameter estimates generated under the null, as Cavaliere et al. (2012) do.}\)
We may expect that all three alternatives to the LF trace test – a MF trace test, a trace test with higher-order VECM, and a sieve bootstrap trace test – will improve on the size distortion of the LF trace test.

4.2 Alternatives to the Residual-Based Tests

We have thus far discussed two options for each of the two residual-based tests when MF data are available: either standard LF tests or residual-based tests based on MF regressions. Our theoretical results (and simulations below) show that tests based on a MF regression may have worse size properties than those based on a LF regression with aggregated data, because the aggregation weights cannot be estimated consistently.

Of course, we may also consider increasing the lag order of the test: augmented Dickey-Fuller (ADF) instead of DF tests. This increases the allowance for serial correlation in the residuals of the cointegrating relationship, but note that it is not exactly the same as allowing for serial correlation in the first difference of the series, as in the case of a VECM.

A bootstrapping scheme is natural in this case, too. Specifically, we use the residual-based sieve bootstrap algorithm of Palm et al. (2010) modified by using the DF tests as above in place of the Wald test considered by those authors. Our purpose here is to compare different ways to implement the same test using LF and MF series, so our substitution of the DF test for their test is just for comparisons with our other DF tests.

In addition, we propose a MF alternative to the residual-based MF tests estimated with least squares: a restricted CoMIDAS (cointegrating mixed data sampling) regression, analyzed by Miller (2014b) and based on the MIDAS regression of Ghysels et al. (2004). Instead of least squares, the coefficients are estimated in a way that imposes restrictions on the magnitude of the inconsistency. 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), restrict the weights to sum to unity and to be non-negative, as we have assumed.
An \( i \)-th order exponential Almon lag generates weights of the form

\[
\omega_{sk}(\gamma_1, \ldots, \gamma_i) = \frac{\exp(\gamma_1 k + \cdots + \gamma_i k^i)}{\sum_{j=1}^{m} \exp(\gamma_1 j + \cdots + \gamma_i j^i)},
\]

for \( s = 1, \ldots, p \) and \( k = 1, \ldots, m \). This function is chosen for its flexibility in mimicking empirically useful weighting schemes, including flat, EOP, or BOP sampling.

In order to limit size distortion, an exponential Almon lag may be used to estimate the model in (12), because the fitted residuals are less sensitive to inconsistent estimates of the parameters \( \gamma_1, \ldots, \gamma_i \) underlying the aggregation weights due to the nonlinearity than they are to inconsistent estimates of the aggregation weights using least squares. The tests we propose are simply the DF tests run on the residuals of the CoMIDAS regression.

5 Finite-Sample Comparisons

We examine both the size and the power of the LF tests, the MF tests, and the proposed alternatives by way of Monte Carlo simulations. Recall that type A size distortion is defined as the size distortion resulting from aggregating all series identically, while type B size distortion results from aggregating two or more series differently.

5.1 Size

Similarly to Cheung and Lai (1993), we use a bivariate \( (p = 2) \) HF DGP given by \( \Delta^{(1/m)} y_{t-i/m} = \varepsilon_{t-i/m} \) under the null, with \( \text{var}(\varepsilon_{t-i/m}) = \Sigma \) such that \( \sigma_{11} = \sigma_{22} = 1 \) and \( \sigma_{21} = \sigma_{12} = 0.9 \). Consequently, \( \Delta z_1^a = \Pi_a (I \otimes (I + H' + HL)) u_t \) and \( \Delta z_1^m = \Pi_m (I \otimes (I + H' + HL)) u_t \) are the LF and MF null models.

We consider \( m = 12 \) and \( T = 200 \), suggestive of monthly series aggregated to an annual frequency. We vary the aggregation weight vectors \( \omega_1, \omega_2 \), and we conduct 1,000 simulations for each model. We use E-E to denote that both series have been EOP sampled to the LF, while E- denotes that the first series has been EOP sampled, while the remaining
series is observed at the HF. We use similar notations for beginning-of-period sampling (B), and flat sampling (F). MF-OLS denotes a single MF regression estimated using least squares, while CoMIDAS(2) denotes a CoMIDAS regression with a 2\textsuperscript{nd}-order exponential Almon lag estimated using nonlinear least squares.

For the tests that do not use bootstrapped critical values, we calculate critical values based on those that yield a rejection rate of 0.05 for 1,000 simulations of the HF null DGP. By comparing the test statistics from the MF and LF models – i.e., those with data subject to aggregation – to these critical values, we more accurately assess finite-sample deviations from the nominal size due to aggregation than if we were to use critical values from the literature generated using different seeds, sample sizes, etc.

As discussed above, bootstrapped critical values are generated using the algorithm of Palm et al. (2010) with 99 bootstrap replications, but substituting in the test statistics considered here. We fix the lag order of the test statistics to be the same in the bootstrap replications as for that of the original test statistic (zero), but Palm et al. (2010) do not require these to be the same. Palm et al. (2010) discuss conditions for the lag order of the sieve relative to the lag order of the test statistic. Their theoretical assumptions require that the sieve lag order be smaller, but they employ an information criterion to choose it.

Since our goal is to pick up first-order MA serial correlation, we set the sieve order to be 2.

Table 1 and Figure 1 show empirical sizes (rejection rates). Looking first at Table 1, the first three rows and three columns show Type A size distortion – that is, size distortion that results when we are lucky enough to be able to use the same aggregation scheme on all series. Clearly skip sampling (either BOP or EOP) of all series provides the least size distortion: Type A size distortion for these cases is asymptotically zero. The worst Type A size distortions are for tests applied to flat-sampled series, but even these seem acceptable against a nominal size of 0.05.

The worst size distortion is caused by mixing different skip sampling techniques (the last three columns of the first three rows), which suggests the use of extreme caution when
<table>
<thead>
<tr>
<th>Case</th>
<th>Test</th>
<th>E-E</th>
<th>B-B</th>
<th>F-F</th>
<th>E-B</th>
<th>F-E</th>
<th>F-B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-frequency Trace</td>
<td></td>
<td>0.057</td>
<td>0.056</td>
<td>0.078</td>
<td>0.821</td>
<td>0.489</td>
<td>0.475</td>
</tr>
<tr>
<td>Rho</td>
<td></td>
<td>0.048</td>
<td>0.050</td>
<td>0.011</td>
<td>0.743</td>
<td>0.397</td>
<td>0.400</td>
</tr>
<tr>
<td>Tau</td>
<td></td>
<td>0.057</td>
<td>0.052</td>
<td>0.013</td>
<td>0.737</td>
<td>0.393</td>
<td>0.395</td>
</tr>
<tr>
<td>Low-frequency (bootstrapped) Trace</td>
<td></td>
<td>0.030</td>
<td>0.033</td>
<td>0.045</td>
<td>0.065</td>
<td>0.042</td>
<td>0.047</td>
</tr>
<tr>
<td>Rho</td>
<td></td>
<td>0.030</td>
<td>0.028</td>
<td>0.032</td>
<td>0.065</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td>Tau</td>
<td></td>
<td>0.033</td>
<td>0.033</td>
<td>0.029</td>
<td>0.063</td>
<td>0.046</td>
<td>0.050</td>
</tr>
<tr>
<td>MF-VAR Trace</td>
<td></td>
<td>0.052</td>
<td>0.051</td>
<td>0.068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF-OLS Rho</td>
<td></td>
<td>0.508</td>
<td>0.519</td>
<td>0.410</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tau</td>
<td></td>
<td>0.497</td>
<td>0.510</td>
<td>0.406</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CoMIDAS(2) Rho</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.053</td>
<td>0.057</td>
<td>0.021</td>
</tr>
<tr>
<td>Tau</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.060</td>
<td>0.059</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Table 1: **Empirical sizes compared to a nominal size of** 0.05. Results from 1,000 simulations with $m = 12$ and $T = 200$. E-E, both series EOP sampled; E-, first series EOP sampled and remaining series observed at the HF. Similarly, B, beginning-of-period sampling; F, flat sampling. MF-OLS, MF regression estimated using least squares; CoMIDAS(2), MF regression with 2nd-order exponential Almon lag. **Bold** denotes $|\text{size} - 0.05| < 0.02$.

![Figure 1](image-url)  

Figure 1: **Empirical sizes compared to a nominal size of** 0.05 with an increasing lag order. Results from 1,000 simulations with $m = 12$ and $T = 200$. Lag order of the tests denoted by $p$. 
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! Recall from Section 2.1 that in the case of mixing EOP and BOP sampling the serial correlation that induces Type B size distortion is a function of the cross-correlation $\sigma_{21}$, which is generally non-zero. This relationship between cross-correlations and serial correlation holds when mixing any different aggregation schemes. Flat sampling is more conservative, but mixing flat and skip sampling (the last two columns) still leads to an unacceptably large size. Clearly, aggregating a MF model to the lowest frequency is risky.

In the previous section, we discussed two alternatives to the LF tests: bootstrapping the critical values or increasing the lag order of the tests. The second three rows of Table 1 shows the sizes of the LF tests using bootstrapped critical values. The first three columns suggest a systematic under-rejection using all three tests – negative type A size distortion, while the next three columns suggest positive or negative type B size distortion. Most importantly, we observe much better size control in most cases with the bootstrap. The only exceptions are when both series are either EOP and BOP sampled, in which case our theoretical results show that there is no size distortion to begin with, so that the bootstrap is unnecessary.

Figure 1 shows the empirical sizes of the LF tests when their lag orders are allowed to increase for combinations of two representative schemes: flat and EOP sampling. Most notable is the dramatic decrease in the size distortion in the Flat-EOP case across all three statistics. Including just two lags beyond the lag order of the DGP brings the rejections rates from about 40%-50% to less than 10%. The danger of this approach is that size distortion may also increase, as seen most clearly in the case of flat sampling using the ADF coefficient test, where the rejection rate jumps to 20% after adding one lag. The other tests show less dramatic increases.

We now turn to the results of the tests using MF series. The remaining rows of Table 1 show results from the simulations using the MF models. As our theoretical results suggest,
the MF trace test exhibits an acceptable size regardless of how the LF series is aggregated.

Size distortion is expected using the remaining methods. Using unrestricted least squares on a MF regression performs the worst of the MF methods, but not necessarily worse than arbitrarily aggregating every series to the low frequency. CoMIDAS does reasonably well, which is not surprising since the weight restrictions limit the size distortion from inconsistent aggregation weight estimates.

Finally, revisiting Figure 1, since the size distortion using least squares is so bad, we consider increasing the LF lag order of the MF tests. We generally come to the same conclusions about increasing the lag order as we did for the LF case. Specifically, increasing the lag order to 1 or 2 can ameliorate the most extreme size distortions, but increasing it too much can increase distortions.

5.2 Power

Power against local alternatives is an important consideration, and we now focus on the powers of the tests and procedures that achieve the best size control. We are particularly interested in determining whether any of the methods that improve size sacrifice power in doing so. We size-adjust power by using the critical values that give 5% rejection rates under the null. The size-adjusted powers of the base LF tests are therefore the same as their bootstrap versions, so we deem them to be interchangeable for power comparisons.

We consider a power function generated by

\[
\Delta^{(1/m)} y_{t-i/m} = \frac{c}{T} \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} y_{t-(i+1)/m} + \varepsilon_{t-i/m},
\]

where \(c = [0.4, 0.8, \ldots, 7.6, 8] \) corresponds to local alternatives and \(c = 0 \) corresponds to the null. The parameters of the simulations are otherwise identical to those under the null discussed above.

Simulated power functions are displayed in Figure 2. The left panels show the size-
Figure 2: **Size-adjusted power against local alternatives.** Results from 1,000 simulations with $m = 12$ and $T = 200$. Left panels show models with (at least) one series subject to flat sampling. Right panels show models with (at least) one series subject to EOP sampling.
adjusted powers of the three sets of tests when (at least) one series is subject to flat sampling. The top left panel shows that of the trace test. Clearly, increasing the lag order in order to achieve size control substantially lowers the power of the test against local alternatives. Flat sampling the second series and running a LF trace test has nearly the same power as using a MF trace test. In contrast, higher power can be achieved by using different aggregation schemes, but recall that the bootstrap correction is essential to control size in this case.

The middle and bottom left panels of the figure show the powers of the residual-based tests. Again, increasing the lag order sacrifices power. The residual-based tests using least squares also have lower power, once the size is adjusted. The remaining three options: aggregating the HF series using flat sampling, aggregating the HF series using EOP sampling, or using a MIDAS regression, exhibit about the same power – but, again, with the caveat that the size adjustment is critical when the aggregation schemes are different.

The right panels of Figure 2 differ from the left in that (at least) one series is EOP sampled rather than flat sampled. The results are qualitatively the same as with flat sampling, with the exception that the MIDAS regression gives a small power advantage.

6 Empirical Application

We study data made available by Robert Shiller and pertaining to his 2000 book entitled *Irrational Exuberance*. This data set consists of monthly US stock prices, annual dividends, and earnings data, all starting January 1871.\(^\text{13}\) The dividend and earnings series represent flows, whereas stock prices are point sampled stocks. Annual and monthly data run through 2012, hence \(T = 142\) and \(m = 12\). We use \(P\) and \(D\) to denote the log of nominal price and the log of nominal dividends.

We compare testing cointegration with annual versus MF data using annual dividends versus annual and monthly stock price data. Shiller sets annual price equal to January daily

\(^\text{13}\)Data from Shiller’s website are available at: [http://www.econ.yale.edu/~shiller/data.htm](http://www.econ.yale.edu/~shiller/data.htm). Monthly dividends are available too, but they are linear interpolations of the annual data. Testing with interpolated data brings about other issues not covered here. (See Ghysels and Miller, 2014.)
average (BOP sampling).

We expect that each series follows a random walk with drift, and unit root tests (not shown) are strongly supportive. 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. Although each series has a drift, we expect them to have the same drifts. In that case, the drifts cancel out and we do not need to include a linear trend.

There is a simple reason why we do not believe that cointegration holds between aggregate prices and dividends: many tech companies do not pay dividends, and the increasing weight of tech companies in the US economy beginning in the mid-1990’s means the possibility of a structural break in or the dissolution of the long-run relationship between these series. A rejection of the null of no cointegration discussed above using any of the three tests should indicate positive size distortion.

Estimation is a two step procedure. We first demean $P$ and $D$, and then run cointegration tests on demeaned data. The test results appear in Table 2, where Annual-$F$ uses annual $D$ with average (flat) $P$, Annual-$B$ uses annual $D$ with January $P$ (Shiller’s BOP sampling scheme), Annual-$E$ uses annual $D$ with December $P$ (EOP, or December average daily price). Finally, $MF$ denotes the MF VECM for the trace test and the unrestricted MF single-equation regression for the residual-based tests, while CoMIDAS(2) uses nonlinear least squares to estimate a $2^{nd}$-order exponential Almon lag (single-equation case only). We show both asymptotic and bootstrapped critical values.

Our theoretical results suggest that the “safest” LF strategy (in terms of possible size distortion) would be to flat sample monthly $P$. Indeed, if we do this, we fail to reject the no cointegration null with any of the tests, as expected. However, the two residual-based tests show statistics of $-19.60$ and $-3.24$ that are close to the asymptotic critical values of $-20.50$ and $-3.37$.

Our theoretical results also suggest that we can lower size distortion by using the “right” skip sampling for $P$. The test statistics are slightly smaller using BOP (Annual-$B$) sampling
<table>
<thead>
<tr>
<th></th>
<th>Trace Test Statistic</th>
<th>Coefficient Test Statistic</th>
<th>t-test Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BS CV</td>
<td>BS CV</td>
<td>BS CV</td>
</tr>
<tr>
<td>Annual-F</td>
<td>12.56</td>
<td>-19.60</td>
<td>-3.24</td>
</tr>
<tr>
<td>Annual-B</td>
<td>10.47</td>
<td>-17.54</td>
<td>-3.05</td>
</tr>
<tr>
<td>Annual-E</td>
<td><strong>28.49</strong></td>
<td><strong>-36.90</strong></td>
<td><strong>-4.61</strong></td>
</tr>
<tr>
<td>MF</td>
<td>11.03</td>
<td><strong>-23.43</strong></td>
<td><strong>-3.57</strong></td>
</tr>
<tr>
<td>CoMIDAS(2)</td>
<td>–</td>
<td>-16.31</td>
<td>-2.94</td>
</tr>
<tr>
<td>Asymptotic CV</td>
<td>19.96</td>
<td>-20.50</td>
<td>-3.37</td>
</tr>
</tbody>
</table>

Table 2: Cointegration Tests: Log Stock Price and Log Dividend. \(T = 142, m = 12, M = mT = 1704\). Annual-F, annual \(D\) with average \(P\) (flat); Annual-B, annual \(D\) with January \(P\) (BOP); Annual-E, annual \(D\) with December \(P\) (EOP); MF, unrestricted MF VAR or single-equation model with OLS; CoMIDAS(2), restricted with a 2nd-order exponential Almon lag (single-equation case only). Size 0.05 asymptotic critical values from Osterwald-Lenum (1992, case 1*, restricted constant) and Phillips and Ouliaris (1990, demeaned). **Bold** denotes that the statistic exceeds the respective asymptotic critical value. BS CV denotes bootstrapped critical values.

of monthly \(P\) rather than using flat sampling and much smaller using BOP versus EOP sampling. Using BOP sampling of prices, we fail to reject any of the nulls, as expected, but we reject all of the nulls using EOP sampling. These findings are indeed consistent with the theory if in fact annual \(D\) is sampled using something close to BOP sampling.

As shown in the previous section, bootstrapped critical values ameliorate size distortion, and this is most notable in the case of EOP sampled \(P\) (Annual-E), where the rejections using the trace test statistic and \(t\)-test statistic are reversed and that using the coefficient test statistic becomes quite marginal.

What happens if we use the MF data directly? The trace test statistic 11.03 is close to the Annual-B case (10.47), which is also consistent with the theory if in fact annual \(D\) is sampled using something close to BOP sampling. The residual-based test statistics are not close to the Annual-B case, which is again consistent with the theory, since these tests cannot avoid size distortion. They reject the null. In contrast, the residual-based CoMIDAS tests control size, and these test statistics are close to those using the BOP sampling. As we expect and as with the BOP case, we fail to reject the null using either residual-based test applied to CoMIDAS regression residuals.
Of course, we do not know exactly how $D$ was aggregated. In such cases, we should use the MF trace test and CoMIDAS for single-equation tests, or otherwise we should use a bootstrap for size control. Since these gave us test statistics similar to the $Annual-B$ in this case, using annual dividends and sampling January prices from each year, it seems reasonable to use the $Annual-B$ strategy, which is what Shiller (1989) actually did to create the annual data set.

7 Concluding Remarks and Practical Recommendations

Standard tests for cointegration are affected by temporal aggregation. 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. The severity of the size distortion results directly from serial correlation induced by the aggregation. That flat sampling causes serial correlation in a univariate series has been known for a long time (Working, 1960). However, the effects on cointegration tests using more general aggregation schemes and possibly different aggregation schemes on each series has not been studied previously.

Although the cointegration tests considered here are routinely modified to accommodate mild serial correlation, our empirical results suggest that such a modification may have the opposite of the intended effect – an increase in size – and generally sacrifices power. Without additional modification, however, the standard tests cannot adequately capture the serial correlation from aggregation.

We find that type A size distortion, or size distortion resulting from using multiple series that have been aggregated in the same way, is nil at best and mild at worst. The largest rejection rate that we get from simulations is an acceptable 0.078 when both series have been flat sampled compared to a nominal size of 0.05. When all series are skip sampled in the same way, there is no size distortion.

In contrast, type B size distortion, when series have been aggregated differently, may
be very severe. The most severe case appears to be when different skip sampling methods (EOP and BOP sampling) are used, even though univariate series sampled in this way have no serial correlation. The serial correlation that drives size distortion of the tests in this case arises from the covariance of the differently aggregated series.

Our results for the trace test unambiguously show that the MF VECM suffers the least size distortion, except in the case in which all series happen to be skip sampled in the same way. As long as it is feasible to do so, there appears to be no disadvantage to constructing the trace test on a MF VECM rather than aggregating the HF to get a LF VECM – especially when the aggregation schemes for the existing LF series are unknown.

The results for the residual-based cointegration tests are less clear, because all residual-based tests analyzed suffer from size distortion. The source of the size distortion using LF series is augmented by inconsistent estimation of the aggregation weights under the null. Least squares performs particularly poorly in this regard. Although such size distortion is unavoidable, a nonlinear cointegrating MIDAS strategy may be used to mitigate the inconsistency and therefore the size distortion.

As an alternative to the LF tests and the MF versions of those tests considered here, the critical values of the LF tests may be chosen using a sieve bootstrap. Such a procedure is particularly useful when no HF data are available. Our numerical results suggest that bootstrapping the LF tests works about as well (in terms of size and power) as the MF trace test or the residual-based tests based on a nonlinear CoMIDAS regression with a second-order exponential Almon lag.

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, and we address the effects of interpolation on these tests in a separate analysis (Ghysels and Miller, 2014).
References


Supplemental Appendix to “Testing for Cointegration with Temporally Aggregated and Mixed-frequency Time Series”

Eric Ghysels and J. Isaac Miller

A Proofs of the Main Results

The following two ancillary lemmas simplify the subsequent proofs of the main results. Specifically, Lemma A1 collects limiting distributions of the moments $R_{00}$, $R_{11}$, and $R_{10}$, which are common to the four theorems (and all six LF and MF tests), and Lemma A2 presents the common limiting distribution of the least squares estimators of the single-equation LF and MF models used in the proofs of the latter two theorems.

Lemma A1. Suppose that the null of no cointegrating vectors is true.

(a) $R_{00} \to_d m(\Sigma^{1/2} \otimes I_m)(I_p \otimes m^{-1}H_{00})(\Sigma^{1/2} \otimes I_m)$

(b) $T^{-1}R_{11} \to_d m(\Sigma^{1/2} \otimes I_m)(\int WW' \otimes u')(\Sigma^{1/2} \otimes I_m)$

(c) $R_{10} \to_d m(\Sigma^{1/2} \otimes I_m)((\int WdW' \otimes u') + (I \otimes m^{-1}H_{10}))(\Sigma^{1/2} \otimes I_m)$

as $T \to \infty$.

Proof of Lemma A1. (a) Under the null hypothesis, the HF increments are $\Delta(1/m)z_t = u_t$, so that the LF increments are $\Delta z_t = \sum_{i=0}^{m-1} u_{t-i/m} = (I_p \otimes (I_m + H' + HL))u_t$. Since $(u_t)$ is iid,

$$\text{plim } T^{-1} \sum u_t u'_t = \text{plim } T^{-1} \sum u_{t-1} u'_{t-1} = \mathbb{E}u_t u'_t = \Sigma \otimes I_m,$$

which simplifies to $\Sigma \otimes H_{00}$ in our notation. The stated result follows.

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

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

$$= (I_p \otimes u') \sum_{j=1}^{t-1} u_j - (I_p \otimes H)u_{t-1}$$
since \( I + H' + H = u' \) by construction. Thus,

\[
T^{-1}R_{11} = (I_p \otimes u')T^{-2} \sum_{j=1}^{t} \sum_{i=1}^{t} u_j u_i (I_p \otimes u') + o_p(1)
\]

where the \( o_p(1) \) terms follow from an LLN and the rate of convergence in part (c) of this lemma.

Now, we may write

\[
T^{-1}R_{11} = m \left( T^{-1} \sum \left( \frac{1}{\sqrt{M}} \sum_{j=1}^{t} \sum_{k=0}^{m-1} \varepsilon_{j-k/m} \right) \left( \frac{1}{\sqrt{M}} \sum_{i=1}^{t} \sum_{l=0}^{m-1} \varepsilon_{l-t/m} \right) \otimes u' \right) + o_p(1)
\]

using the fact that \((I_p \otimes \cdot')u_j = \sum_{k=0}^{m-1} \varepsilon_{j-k/m}\). By appealing to the assumed HF invariance principle \( T^{-1}R_{11} \to_d m(\int BB' \otimes u') \) and the stated result follows.

(c) The sample moment \( R_{10} = T^{-1} \sum z_{t-1} \Delta z'_t \) may be rewritten as

\[
R_{10} = T^{-1} \sum \left[ (I_p \otimes u') \sum_{j=1}^{t-1} u_j - (I_p \otimes H)u_{t-1} \right] u'_t (I_p \otimes (I_m + H)) + T^{-1} \sum \left[ (I_p \otimes u') \sum_{j=1}^{t-2} u_j + (I_p \otimes (I_m + H))u_{t-1} \right] u'_{t-1} (I_p \otimes H')
\]

by substitution as above and again recognizing that \( I + H' + H = u' \). The first term of (13) may be written as

\[
T^{-1} \sum (I_p \otimes u') \sum_{j=1}^{t-1} u_j u'_t (I_p \otimes (I_m + H)) + o_p(1),
\]

because \((u_t)\) is serially uncorrelated. The second term of (13) may be written as

\[
T^{-1} \sum (I_p \otimes u') \sum_{j=1}^{t-1} u_j u'_t (I_p \otimes H') + (\Sigma \otimes H_{10}) + o_p(1),
\]

similarly.

Adding the first terms of (14) and (15) yields

\[
T^{-1} \sum \sum_{l=0}^{m-1} \left( \sum_{j=1}^{t-1} \sum_{k=0}^{m-1} \varepsilon_{j-k/m} \varepsilon'_{l-t/m} \right) \otimes u'
\]

using the facts that \( I + H' + H = u' \) and \((I_p \otimes \cdot')u_t = \sum_{k=0}^{m-1} \varepsilon_{t-k/m}\). The expression in (16) may be rewritten as

\[
m \left[ \frac{1}{M} \sum_{l=1}^{M} \sum_{j=1}^{m(t-1)} \varepsilon_{j/m} \varepsilon'_{l/m} \right] \otimes u',
\]

and the expression in square brackets has a limit given by \( \int BdB' \) using standard unit root covariance asymptotics with the HF invariance principle assumed here. Adding the remaining non-negligible (second) term in (15), the stated result follows. \( \square \)
Lemma A2. Define \( e_1 \) and \( E_2 \) to be the vector and matrix that select the first column and all but the first column, respectively, of the preceding matrix. Define the partition \( B = (B_1, B_2)' \) using \( B_1 = e_1'B \) and \( B_2 = E_2'B \). Under the null hypothesis of no cointegrating relationships, the limiting distributions of least squares estimators of \( \beta \) in the LF model in (10) and the MF model in (12) both coincide with \((\int B_2B_2')^{-1} \int B_2B_1\).

Proof of Lemma A2. The least squares estimator \( \hat{\beta} \) of \( \beta \) in the LF model in (10) may be written as

\[
\hat{\beta} = \left(T^{-2} \sum_t E_2\Pi_a z_t z_t' \Pi_a' E_2\right)^{-1} T^{-2} \sum_t E_2' \Pi_a z_t z_t' \Pi_a E_1.
\]

From Lemma A1, it follows that

\[
T^{-2} \sum_t z_t z_t' = T^{-1} R_{11} + o_p(1) \rightarrow_d m \left( \int B'B \otimes u' \right)
\]

so that \( T^{-2} \sum_t \Pi_a z_t z_t' \Pi_a' \rightarrow_d m \int B'B \) due to the block diagonal structure of \( \Pi_a \) and the restriction that the aggregation weights must sum to one. The resulting distribution follows immediately.

The MF model in (12) may be rewritten as

\[
y_{1t} = \left(y_{2t}', ..., y_{pt}'\right) \beta + w_t' \gamma + e_{1t}'
\]

where \( y_{kt} = y_{kt} \) for \( k = p_l + 1, ..., p \), \( w_t \equiv (\Delta^{(1/m)} z_{p_l+1,t}' H', ..., \Delta^{(1/m)} z_{p,t}' H')' \), and \( \gamma \equiv (\beta_{p_l+1} \omega_{p_l+1}, ..., \beta_p \omega_p)' \). In other words, imposing end-of-period sampling \( (y_{kt} = y_{kt}) \) on the MF model in (12) results in the LF model in (10) with the addition of the I(0) term \( w_t' \gamma \). Even though the second term contains some elements of \( \beta \), \( \beta \) is identified by the first term. The addition of the I(0) term does not change the limiting distribution of the least squares estimator of \( \beta \). \( \square \)

Proof of Proposition 1. Transposing and then stacking the original DGP in (1) across \( i \) allows

\[
(\Delta^{(1/m)} z_{1t}, \ldots, \Delta^{(1/m)} z_{pt}) = (z_{1,t-1/m}, \ldots, z_{p,t-1/m}) A_1'H' + (u_{1t}, \ldots, u_{pt}),
\]

and vectorizing both sides allows \( \Delta^{(1/m)} z_t = (\Gamma A' \otimes I_m) z_{t-1/m} + u_t \). This expression represents a system of \( mp \) equations with an error variance given by \( \text{var}(u_t) = \Sigma \otimes I_m \). The result in (4) is obtained by adding and subtracting \( z_{t-1} \) and \( (\Gamma A' \otimes I_m) z_{t-1} \) from this expression. The result in (5) subsequently follows by premultiplying both sides by \( \Pi_a \) and noting that \( \Pi_a(\Gamma A' \otimes I_m) = \Gamma A' \Pi_a \). \( \square \)

Proof of Proposition 3. Premultiply both sides of the VECM in (4) by \( \Pi_m \) to get

\[
\Delta z_t^m = \Pi_m(\Gamma A' \otimes I) z_{t-1} + \eta_t^m.
\]
Now, $\Pi_m(\Gamma A' \otimes I_m)$ does not simplify directly as it did in the LF case, and the model in (17) is not a VECM, because both HF and LF series in $\langle \triangle z_t^m \rangle$ are regressed on $(z_{t-1})$ rather than $(z_{t-1}^{m-1})$. Restricting $A$ both identifies the cointegrating space and simplifies the model in (17). Defining $\varpi_*$ such that $\varpi_*^t = 1$, and defining $(p_l + mp_n) \times r$ matrices

$$\Gamma^m \equiv \begin{bmatrix} I_{p_l} & 0 \\ 0 & I_{p_n} \otimes \varpi_* \end{bmatrix} \Gamma$$

and

$$A^m \equiv \begin{bmatrix} I_{p_l} & 0 \\ 0 & I_{p_n} \otimes \varpi_* \end{bmatrix} A,$$

some algebra reveals that $\Pi_m(\Gamma A' \otimes I_m) = \Gamma^m A^m \Pi_m$ and the result is obtained. □

**Proofs of Theorems 2 and 4.** The proof of Theorem 4 follows immediately from the results of Lemma A1 by noting the continuity of the trace and matrix multiplication operators. The proof of Theorem 2 proceeds in the same way, but aggregation of all series allows simplification. For a generic $p \times p$ matrix $C$, the block diagonal structure of $\Pi_a$ allows $\Pi_a(C \otimes I_m) = C \Pi_a$ and, along with the restriction that the weights in each block of $\Pi_a$ sum to one, $\Pi_a(C \otimes \mu')\Pi_a = C$. The first of these equalities, along with the multiplicative structure of the test statistic, allows complete cancellation of $\Sigma^{1/2}$ in the limits in Lemma A1 from the limiting distribution of the test statistic, which employs these limits. The second equality allows simplification of the limits involving $\mu'$ in parts (b) and (c) of Lemma A1, once the aggregation matrix $\Pi_a$ is applied to these limits. The stated result then follows from the definitions of $Z_{00}^a$ and $Z_{10}^a$.

**Proof of Theorem 5.** The series on which the unit root tests are conducted is $\hat{e}_t^a = z_t^a \Pi_u'(1, -\hat{\beta}')'$, and the tests may be written as $T^{-2} \sum_{t}(\hat{e}_t^{a-1})^2 T^{-1} \sum_{t}(\hat{e}_t^{a}) \Delta \hat{e}_t^a$ or $(\hat{\alpha}' T^{-1} R_{11}^a \hat{\alpha})^{-1} \hat{\alpha}' R_{10}^a \hat{\alpha}$ and $(\hat{\sigma}_T^2)^{-1} \sum_{t}(\hat{e}_t^{a-1})^2 T^{-1} \sum_{t}(\hat{e}_t^{a}) \Delta \hat{e}_t^a$ with $\hat{\sigma}_T^2 = T^{-1} \sum_{t}(\Delta \hat{e}_t^a)^2$ or $(\hat{\alpha}' R_{00}^a \hat{\alpha}')^{-1} T^{-1} R_{11}^a \hat{\alpha}^{-1/2} \hat{\alpha}' R_{10}^a \hat{\alpha}$.

We employ the decomposition of Phillips and Ouliaris (1990): $\Sigma = L' L$, where

$$L = \begin{bmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{bmatrix} = \begin{bmatrix} \sqrt{\sigma_{11} - \sigma_{12} \Sigma^{-1}_{22} \sigma_{21}} & 0 \\ \Sigma^{-1/2}_{22} \sigma_{21} & \Sigma^{1/2}_{22} \end{bmatrix} \quad \text{with} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma'_2 \\ \sigma_2 & \Sigma_{22} \end{bmatrix},$$

such that $\sigma_1^2$ is a scalar, $\Sigma_{22}$ is a $(p - 1) \times (p - 1)$ matrix and $\sigma_{21}$ is a $(p - 1)$-vector. The rotation $B = L' W$ allows expressions of $B_1$ and $B_2$ explicitly in terms of $W_1$ and $W_2$: $B_1 = W_1 l_{11} + W_2' l_{21}$ and $B_2 = L_{22}' W_2 = \Sigma_{22}/2 W_2$.

Denote the limit of $\hat{\alpha}$ as $\eta$ and define a generic matrix $\Xi$ such that $m\eta' L' \Xi L \eta$ denotes the limits of $\hat{\alpha}' R_{00}^a \hat{\alpha}$, $\hat{\alpha}' T^{-1} R_{11}^a \hat{\alpha}$, and $\hat{\alpha}' R_{10}^a \hat{\alpha}$ implied by applying the aggregation matrix $\Pi_a$ to the three limits in Lemma A1. Specifically, $\Xi = Z_{00}^a$ corresponds to part (a), $\Xi = \int W W'$ corresponds to part (b), and $\Xi = \int W dW'$ + $Z_{10}^a$ corresponds to part (c).

By Lemma A2,

$$\hat{\beta} \rightarrow_d L_{22}^{-1} \left( \int W_2 W_2' \right)^{-1} \int W_2 (W_1 l_{11} + W_2' l_{21}) = L_{22}^{-1} (l_{11} \kappa_2 + l_{21})$$

where $\kappa_2 \equiv (\int W_2 W_2')^{-1} \int W_2 W_1$ such that $\kappa = (1, -\kappa_2)'$ with $\kappa$ defined in (9). Conse-
quenty, \( \eta = (1, -(l_{11} \kappa_2' + l_{21}^2)L_{22}^{-1} \lambda)' \). For convenience, we may write \( L e_1 = (l_{11}, l_{21}^2)' \) and \( L E_2 = (0, L_{22})' \), with \( e_1 \) and \( E_2 \) defined as in Lemma A2, and note that

\[
Le_1 - LE_2 L_{22}^{-1} (l_{11} \kappa_2 + l_{21}) = l_{11} \kappa
\]

holds.

We are interested in the limit given by \( m \eta' L \Xi L \eta \) for each of the three moments. Generically,

\[
\eta' L \Xi L \eta = \begin{bmatrix}
1 & 0 \\
L_{22}^{-1}(l_{11} \kappa_2 - l_{21}) & L_{22}^{-1}(l_{11} \kappa_2 - l_{21})
\end{bmatrix}
\begin{bmatrix}
e_1' L \Xi L e_1 & e_1' L \Xi L E_2 \\
E_2' L \Xi L e_1 & E_2' L \Xi L E_2
\end{bmatrix}
\begin{bmatrix}
1 \\
L_{22}^{-1}(l_{11} \kappa_2 - l_{21})
\end{bmatrix}
\]

which simplifies to \( l_{11} \kappa \Xi \kappa \) using (18).

Now, substituting in \( \Xi = \int WW' \), we get \( \alpha' T^{-1} R_{11} \alpha \rightarrow_d ml_{11}^2 \kappa' \int WW' \kappa = ml_{11}^2 \int Q^2 \). Substituting in \( \Xi = Z_{00}^a \), we get \( \alpha' R_{00} \alpha \rightarrow_p ml_{11}^2 \kappa' Z_{00}^a \). And, substituting in \( \Xi = \int WdW' + Z_{10}^a \), we get \( \alpha' R_{10} \alpha \rightarrow_d ml_{11}^2 \kappa' (\int WdW' + Z_{10}^a) \kappa = ml_{11}^2 (\int QdQ + \kappa' Z_{10}^a \kappa) \). The stated results follow directly from these.

**Proof of Theorem 6.** The MF model in (12) may be rewritten as

\[
z_i' \Pi_\alpha' (1, -\beta)' = w_i' \gamma + e_i'
\]

using the notation defined in the proof of Lemma A2. The proof of Theorem 6 proceeds similarly to that of Theorem 5, with one critical difference: the fitted residuals are \( \hat{e}_t^a = z_i' \Pi_\alpha' (1, -\hat{\beta}') = e_t^a - z_i' \Pi_\alpha' E_2 (\hat{\beta} - \beta) \) in the LF case, while an addition term, \( w_i' (\hat{\gamma} - \gamma) \), is subtracted in the MF case. Because \( \hat{\gamma} \) does not estimate \( \gamma \) consistently under the null of a spurious regression (no cointegration), this term generates non-negligible \( O_p(1) \) remainders in the limits of the two test statistics. \( \square \)