

Implementing Residual-Based KPSS Tests for Cointegration with Data Subject to Temporal Aggregation and Mixed Sampling Frequencies*

J. Isaac Miller[†] Xi Wang[‡]

First Draft: December 15, 2014

This Draft: February 19, 2015

Abstract

We show how temporal aggregation affects the size and power of the DOLS residual-based KPSS test of the null of cointegration. Size is effectively controlled by setting the minimum number of leads equal to one – as opposed to zero – when selecting the lag/lead order of the DOLS regression, but at a cost to power in finite samples. If high-frequency data for one or more series are available, we show how to effectively utilize the high-frequency data to increase power while controlling size.

JEL Classification: C12, C22

Key words and phrases: temporal aggregation, mixed sampling frequencies, MIDAS, cointegration, KPSS test, residual-based cointegration test, dynamic OLS

*The authors are grateful to Eric Ghysels for comments.

[†]Corresponding author. Department of Economics, University of Missouri, 118 Professional Building, Columbia, Missouri 65211, USA. E-mail: millerjisaac@missouri.edu

[‡]Department of Economics, University of Missouri

1 Introduction

It is widely known – see Marcellino (1999) e.g. – that temporal aggregation and sampling frequency have no effect on the long-run relationship between two or more series. The effects on tests of such a relationship are less well known. A number of authors have used Monte Carlo methods to study the effects of sampling frequency on the power of trace and residual-based Dickey-Fuller test statistics, including Hooker (1993), Lahiri and Mamingi (1995), Hu (1996), Otero and Smith (2000), and Haug (2002). The consensus seems to be that sample length has more of an impact than sampling frequency on the power of these tests.¹

More recently, Ghysels and Miller (2015) consider the effects of the realistic scenario of mismatched temporal aggregation schemes. They find that size may be seriously distorted when the test statistics are calculated using data that have been aggregated differently, and they propose modified test statistics based on mixed-frequency approaches when some of the data are available at a higher frequency.

In this paper, we complement the existing studies by switching the null and alternative hypotheses. Instead of the null of no cointegration posited by the residual-based Dickey-Fuller test statistic, we consider the null of cointegration against an alternative of no cointegration. Specifically, we examine Shin’s (1994) adaptation of the widely popular KPSS test statistic (Kwiatkowski *et al.*, 1992) constructed from the residuals of a dynamic ordinary least squares (DOLS) regression. We consider Shin’s (1994) single-frequency case to be an infeasible high-frequency (denoted HF) benchmark, in the sense that at least one series is observed only at a lower frequency (low-frequency denoted LF).

We first explore the size and power properties of single-frequency test statistics based on data subject to possibly mismatched aggregation schemes. Our main findings with respect to LF data are that adding a single lead to a DOLS regression that is otherwise devoid of serial correlation effectively controls size – regardless of how the data are aggregated.

Since the aggregation schemes may be unknown, we introduce two mixed-frequency (MF) versions of the residual-based KPSS test in case some data are observed at the HF. We find that a parsimonious nonlinear mixed-frequency (MIDAS) specification has the best power properties of the specifications with effective size control that we consider.

In the following section, we briefly review Shin’s (1994) DOLS residual-based KPSS test statistic, we show an analogous test statistic using aggregated LF data, and we propose two new test statistics using MF data. In Section 3, we present Monte Carlo results on size and power of the tests, and Section 4 concludes with some practical guidance for implementing the tests with aggregated or mixed-frequency data.

¹Along similar lines, Chambers (2011) showed analytically that least squares estimation of a cointegrating relationship is not even consistent when the span is fixed but the sampling frequency increases.

2 Framework for Testing

2.1 High-Frequency and Low-Frequency Models and Tests

We assume that the data are generated at a higher frequency than the lowest frequency at which they are observed. In particular, the HF data-generating process (DGP) is given by

$$y_{t-i/m} = c'_{t-i/m}\alpha + x'_{t-i/m}\beta + \varepsilon_{t-i/m}, \quad (1)$$

where $c'_{t-i/m} = (1, t - i/m)$ with either $\alpha = (0, 0)'$ (no intercept), $\alpha = (\alpha_1, 0)'$ (intercept only), or $\alpha = (\alpha_1, \alpha_2)'$ (intercept and linear trend), $(x_{t-i/m})$ is an I(1) p -vector series that is not mutually cointegrated, and $(\varepsilon_{t-i/m})$ is I(0) under the null but I(1) under the alternative. We assume T LF observations, indexed by t , and M HF observations, such that $m = M/T < \infty$ and $i = 0, \dots, m - 1$ indexes the m HF observations per LF observation.

If the regressand $y_{t-i/m}$ is subject to aggregation or systematic sampling, it is infeasible to use the HF DGP in (1). Instead, a LF regressand (y_t^a) is observed such that $y_t^a \equiv \sum_{i=0}^{m-1} \varpi_{i+1} y_{t-i/m}$, where (ϖ_{i+1}) with $i = 0, \dots, m - 1$ is a series of aggregation weights such that $\varpi_{i+1} \geq 0$ (non-negativity) and $\sum_{i=0}^{m-1} \varpi_{i+1} = 1$ (unit sum). The most common aggregation schemes are end-of-period sampling for stocks ($\varpi_1 = 1$ and $\varpi_{i+1} = 0$ for $i \geq 1$) and flat sampling (or average sampling or period average) for flows ($\varpi_i = 1/m$ for all i).

A LF model that is equivalent to the HF DGP in (1) is

$$y_t^a = c'_t \alpha^a + x_t^a \beta + \varepsilon_t^a, \quad (2)$$

where either $\alpha^a = (0, 0)'$, $\alpha^a = (\alpha_1, 0)'$, or $\alpha^a = (\alpha_1 - \alpha_2 \sum_{i=0}^{m-1} (i/m) \varpi_{i+1}, \alpha_2)$, $x_t^a = \sum_{i=0}^{m-1} \varpi_{i+1} x_{t-i/m}$, and $\varepsilon_t^a = \sum_{i=0}^{m-1} \varpi_{i+1} \varepsilon_{t-i/m}$.²

When the regressors are aggregated *differently* from the regressand and possibly from each other, then $x_t^b = \sum_{i=0}^{m-1} \Upsilon_{i+1} x_{t-i/m}$ with

$$\Upsilon_{i+1} = \begin{bmatrix} \varpi_{1,i+1} & & 0 \\ & \ddots & \\ 0 & & \varpi_{p,i+1} \end{bmatrix}$$

replaces x_t^a , so that

$$y_t^a = c'_t \alpha^a + x_t^b \beta + \eta_t^b \quad (3)$$

is equivalent to the DGP in (1) by defining $\eta_t^b = \varepsilon_t^a - (x_t^b - x_t^a)' \beta$.

Single-frequency DOLS regressions are given by

$$y_{t-i/m} = c'_{t-i/m}\alpha + x'_{t-i/m}\beta + \sum_{j=-K_H}^{K_H} \Delta^{(1/m)} x'_{t-i/m-j/m} \pi_j^H + \varepsilon_{t-i/m}^* \quad (4)$$

(HF-DOLS) and

$$y_t^a = c'_t \alpha^a + x_t^b \beta + \sum_{j=-K_L}^{K_L} \Delta x_{t-j}^b \pi_j^L + \eta_t^{b*} \quad (5)$$

²Note that aggregating a linear trend alters the intercept, but does not affect the slope, in direct analogy to aggregating a series with a stochastic trend, which affects the short-run properties of the series but does not affect the stochastic trend.

(LF-DOLS), where $\varepsilon_{t-i/m}^* = \varepsilon_{t-i/m} - \sum_{j=-K_H}^{K_H} \Delta^{(1/m)} x'_{t-i/m-j/m} \pi_j^H$, $\eta_t^{b*} = \eta_t^b - \sum_{j=-K_L}^{K_L} \Delta x_{t-j}^{bL} \pi_j^L$, and K_H and K_L denote the number of leads and lags of the HF and LF differences included to allow for endogeneity from a non-diagonal long-run error variance matrix. The regression in (4) augments that in (1) by adding $(2K_H + 1)p$ regressors, while that in (5) augments that in (3) by adding $(2K_L + 1)p$ regressors.

In our notation, Shin's (1994) residual-based KPSS test statistic may be written as

$$CI_H = M^{-2} \sum_{i=1}^M \left(\sum_{j=1}^i \hat{\varepsilon}_{j/m}^* \right)^2 / \hat{\omega}_\varepsilon^2 \quad \text{and} \quad CI_L = T^{-2} \sum_{t=1}^T \left(\sum_{s=1}^t \hat{\eta}_s^{b*} \right)^2 / \hat{\omega}_\eta^2 \quad (6)$$

for the HF and LF cases, respectively. $\hat{\omega}_\varepsilon^2$ is a consistent estimator of the long-run variance of $(\varepsilon_{t-i/m}^*)$ using the fitted residuals $(\hat{\varepsilon}_{t-i/m}^*)$ from (4), and $\hat{\omega}_\eta^2$ is the same estimator applied to the fitted residuals $(\hat{\eta}_t^{b*})$ from (5). The only difference between these statistics is that CI_H is applied to a longer sample of disaggregated – but possibly latent – data, the infeasible HF benchmark, while CI_L is applied to a much shorter sample of aggregated data.

2.2 Two Mixed-Frequency Models and Test Statistics

Now suppose that the regressors are available at the HF.³ Creating a MF test using HF regressors has the potential to control size and improve power in finite samples by using the additional information to estimate the aggregation scheme of the regressors rather than fixing it. On the other hand, doing so increases the number of regressors, in turn increasing the possibilities for estimation error, which may decrease power.

The MF model may be written exactly as in (3), except now the weights are unspecified and may be estimated. Specifically, there are mp coefficients in $(\Upsilon'_{i+1}\beta)$ instead of just p coefficients in β . If the aggregation weights are estimated subject to the restrictions that they sum to unity across the aggregation period and that they are the same across all regressors – i.e., $\Upsilon_{i+1} = v_{i+1}I_p$ – then there are $p + (m - 1)$ unknown parameters in $(\Upsilon'_{i+1}\beta)$.

We specify a mixed-frequency DOLS (MF-DOLS) regression as

$$y_t^a = c_t' \alpha^a + \sum_{i=0}^{m-1} x'_{t-i/m} v_{i+1} \beta + \sum_{j=-K_L}^{K_L} \left(\sum_{i=0}^{m-1} \Delta x'_{t-i/m-j} v_{i+1} \right) \pi_j^L + \eta_t^{b*}, \quad (7)$$

which is derived from (5) by substituting $x_t^b = \sum_{i=0}^{m-1} v_{i+1} x_{t-i/m}$. Now, there are $(2K_L + 1)mp$ coefficients rather than $(2K_L + 1)p$, but the restrictions reduce the number of parameters to be estimated to $(2K_L + 1)p + (m - 1)$.⁴ The KPSS test statistic in this case is CI_L , identical to the LF case, except of course the fitted residuals used in both the numerator and denominator are from (7) instead of (5).

³Since our goal is simply to test for cointegration, there is no loss of generality from setting the regressand to be a LF series. If the data contain more than one LF series, obvious modifications may be made to accommodate regressors observed at different frequencies.

⁴We note that this already complicated representation of the error term understates its complexity when the data are generated continuously (Chambers, 2014). In that case, ML estimation of the underlying continuous-time parameters might improve efficiency under the null and thus control size even better.

Finally, we propose a fourth model – a second MF model – that is a parsimonious nonlinear parameterization of v_{i+1} : $v_{i+1} = v(i, \gamma)$, where γ is a low-dimensional vector. Referred to as MIDAS (MIXed DATA Sampling, Ghysels *et al.*, 2006; Andreou *et al.*, 2010), such parameterizations are common in the mixed-frequency literature. Under our null hypothesis of a cointegrating regression, cointegrating MIDAS (CoMIDAS) regressions have been studied by Götz *et al.* (2014) and Miller (2014).

The idea of the MIDAS parameterization is to estimate the aggregation scheme by minimizing the regression error but subject to a flexible nonlinear restriction to preserve degrees of freedom in comparison with an unrestricted MF regression. This approach is especially useful when $\dim(\gamma)$ is much smaller than $m - 1$.⁵ A leading example of such a parameterization is the second-order exponential Almon lag given by $\kappa(i, \gamma) = \exp(\gamma_1(i + 1) + \gamma_2(i + 1)^2)$. To impose the unit sum restriction, the weights are typically normalized: $v(i, \gamma) = \kappa(i, \gamma) / \sum_{j=0}^{m-1} \kappa(j, \gamma)$.

We propose a CoMIDAS version (CoMIDAS-DOLS) of the MF extension of Shin's DOLS regression model in (7) given by

$$y_t^a = c_t' \alpha^a + \sum_{i=0}^{m-1} x_{t-i/m}' v(i, \gamma) \beta + \sum_{j=-K_L}^{K_L} \left(\sum_{i=0}^{m-1} \Delta x_{t-i/m-j}' v(i, \gamma) \right) \pi_j^L + \eta_t^{b*}, \quad (8)$$

reducing the number of unknown parameters to $(2K_L + 1)p + \dim(\gamma)$ but requiring nonlinear least squares instead of least squares for estimation. The KPSS test statistic in the CoMIDAS-DOLS case is again CI_L , the same as that in the LF-DOLS and MF-DOLS cases, except the fitted residuals come from (8) instead of (5) or (7).

2.3 Brief Comparison

In order to compare the three approaches in (5), (7), and (8), we examine their error terms, given by

$$\varepsilon_t^a - \sum_{j=-K_L}^{K_L} \Delta x_{t-j}^{b'} \pi_j^L - (x_t^b - x_t^a)' \beta, \quad (9)$$

$$\varepsilon_t^a - \sum_{j=-K_L}^{K_L} \left(\sum_{i=0}^{m-1} \Delta x_{t-i/m-j}' v_{i+1} \right) \pi_j^L - \sum_{i=0}^{m-1} x_{t-i/m}' (v_{i+1} - \varpi_{i+1}) \beta, \quad \text{and} \quad (10)$$

$$\varepsilon_t^a - \sum_{j=-K_L}^{K_L} \left(\sum_{i=0}^{m-1} \Delta x_{t-i/m-j}' v(i, \gamma) \right) \pi_j^L - \sum_{i=0}^{m-1} x_{t-i/m}' (v(i, \gamma) - \varpi_{i+1}) \beta, \quad (11)$$

respectively.

Notice that the first terms of each are identical and result simply from the fact that the regressand is aggregated. Chambers (2003) and Miller (2015) showed that efficient estimation of the cointegrating vector, which affects the sizes of the tests above, very much depends on (ϖ_{i+1}) , even when we ignore the remaining terms. Miller (2015) defined aggregation-conditional and aggregation-unconditional efficiency bounds and showed that these are the same with a flat-sampled regressand and obtained when the regressors are also flat sampled.

⁵In contrast, Forni *et al.* (2011) advise against using nonlinear MIDAS specifications with quarterly and monthly data ($m = 3$).

The second terms of (9), (10), and (11) result from including the “dynamic” regressors in the respective DOLS models. Their purposes are to orthogonalize the error with respect to the first difference of the regressors in each model.

The main differences between the errors of each DOLS model lie in the third terms of (9), (10), and (11), which drive differences in both size and power of the respective tests. In (9), this term is determined by the aggregation schemes of the regressand and regressors and is zero only when all have been aggregated in exactly the same way.

The third term in (10) is the same as that in (9), but v_{i+1} is now estimated to minimize the regression error, more effectively controlling size.⁶ However, estimating v_{i+1} introduces $m - 1$ estimation errors that may substantially sap power in finite samples – especially when m is large relative to T .

The third term of the CoMIDAS-DOLS error term in (11) presents a compromise. The weights are less restrictively specified than in (9), but in a way that nests the most common sampling schemes, potentially controlling size better than the LF case. However, the parsimoniousness on the MIDAS specification allows less estimation error than in the MF-DOLS case in (10), potentially improving power. The decrease in estimation error is twofold: there are $m - 1 - \dim(\gamma)$ fewer unknown parameters, and the nonlinear function mutes the variation in the fitted residuals that results from small-sample estimation errors of γ .

3 Monte Carlo Results on Size and Power

We consider a univariate I(1) regressor series $(x_{t-i/m})$, and let $b_{0,t-i/m} = (\varepsilon_{t-i/m}, \Delta^{(1/m)}x_{t-i/m})'$ under the null, but $b_{1,t-i/m} = (\Delta^{(1/m)}\varepsilon_{t-i/m}, \Delta^{(1/m)}x_{t-i/m})'$ under the alternative. We employ HF DGP's given by the model in (1) such that $b_{j,t-i/m} \sim \mathbf{iidN}(0, \Sigma)$ with $\sigma_{11} = \sigma_{22} = 1$ and $\sigma_{21} = \sigma_{12} \in \{1/10, 9/10\}$ for $j = 0, 1$. The choices of σ_{21} reflect weak (1/10) and strong (9/10) endogeneity of the regressors.

We set $\beta = 1$ for simulations, but as may be expected from (9)-(11), varying β affects the results when $x_t^b \neq x_t^a$ – i.e., when the regressor is aggregated differently from the regressand. We set $\alpha_1 = 100$ and $\alpha_2 = 1/10$ in considering the cases with deterministic components.

We consider $m = 12$ and $T \in \{200, 500\}$. Following Kwiatkowski *et al.* (1992) and Lee and Schmidt (1996), the long-run variances in the denominators of the test statistics are estimated using a Bartlett (triangular) window with a window width of integer $(12(M/100))^{1/4}$ for the HF case and integer $(12(T/100))^{1/4}$ for the LF and MF cases.⁷

We conduct 1,000 simulations for each HF DGP, and our critical values are those that yield a rejection rate of 0.05 (nominal size) for the null HF DGP using the HF-DOLS model in (4) with K_H correctly set to zero.

⁶The value of v_{i+1} that minimizes the regression error may or may not be ϖ_{i+1} , depending on the regressand aggregation scheme and endogeneity of the regressors (Miller, 2015).

⁷We also tried the quadratic spectral kernel with Andrews plug-in bandwidth, but do not report the results. The sizes of the test statistics appeared to be similar, but the powers were generally weaker using the plug-in bandwidth.

3.1 Size

Wang (2015) derived the asymptotic distributions of the LF test statistic under the three trend specifications. These distributions contain nuisance parameters that are zero only under strict exogeneity of the regressors and identical aggregation schemes. As a result of these nuisance parameters, we may generally expect some size distortion that is exacerbated when the aggregation schemes are different.

Table 1 shows the deviation of empirical rejection rates from 0.05 for tests based on various aggregation schemes. Specifically, we consider three aggregation schemes: flat/average sampling (or a period average) denoted by F , end-of-period (EOP) sampling denoted by E , and beginning-of-period (BOP) sampling denoted by B .

The rows labeled $K_L = 0$ and $K_L = 1$ reflect sizes of LF tests from the LF-DOLS model in (5) with the two series aggregated in the nine combinations of these three aggregation schemes using regressions with $K_L = 0$ and $K_L = 1$ respectively. For example, the column labeled $B-F$ refers to the case in which the regressand is BOP sampled, but the regressor is flat sampled. The rows labeled $MF-OLS$ and $MIDAS$ correspond to the MF-DOLS and CoMIDAS-DOLS models in (7) and (8) respectively. For these two mixed-frequency models, we set $K_L = 0$. The columns labeled $F-$, $E-$, and $B-$ use models in which only the regressand is flat sampled, EOP sampled, and BOP sampled respectively. We consider an acceptable range for size distortion to be rejection rates less than 0.07 – i.e., deviations in the table less than 0.02. Anything greater than 0.02 is deemed unacceptable and noted in bold.

First, compare cases with both series aggregated and a smaller sample ($T = 200$). Note that the rejection rates almost always decrease as K_L increases. In other words, increasing the number of lags/leads in the LF-DOLS model better controls positive size distortion. In some cases, an unacceptable rejection rate is lowered to the acceptable range by doing so. When $K_L = 1$, 50 out of 54 rejection rates are acceptable, compared to only 34 for $K_L = 0$. There are some differences between the results under strong and weak endogeneity, but there are no salient patterns in the differences between these, other than the fact that the remaining 4 unacceptable rejection rates when $K_L = 1$ correspond to cases of weak endogeneity.

Looking now at the MF cases, with only one series aggregated, we clearly see that the rejection rates are all acceptable. The CoMIDAS-DOLS model has a smaller absolute deviation than the MF-DOLS model in 5 out of 18 cases, but the deviations are close in most cases.

Moving to the larger sample size of $T = 500$, we can see a similar pattern for the results using the LF-DOLS model. Increasing K_L generally decreases positive size distortion and yields acceptable rejection rates in all 54 cases, compared to just 48 when $K_L = 0$. The results are also qualitatively similar for the MF models: all rejection rates are acceptable. Further, with the additional precision from the larger sample, we see 6 cases where the rejection rates of the CoMIDAS-DOLS and MF-DOLS models are the same.

Additional simulations using the LF-DOLS with $K_L > 1$ did not show very much improvement over the case of $K_L = 1$. Most of the improvement in the rejection rates occurs as we move from $K_L = 0$ to $K_L = 1$. Why do we obtain a threshold at $K_L = 1$? In our simulations, we do not introduce any serial correlation in $(b_{0,t-i/m})$. Thus, in the HF-DOLS

Table 1: Deviations of the rejection rates from a nominal size of 0.05 based on residuals of HF-DOIS, LF-DOIS, MF-DOIS, and CoMIDAS-DOIS. Results from 1,000 simulations with $m = 12$, $T \in \{200, 500\}$, and $\sigma_{21} \in \{1/10, 9/10\}$. F , E , and B denote flat sampling (period average), end-of-period sampling, and beginning-of-period sampling, respectively. **Bold** denotes a positive deviation exceeding 0.02.

Strong Endogeneity, T = 200										
	F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B	
No Intercept	$K_L = 0$	0.020	-0.016	0.009	0.012	-0.009	0.012	-0.008	-0.005	-0.009
	$K_L = 1$	-0.008	-0.017	-0.013	-0.011	-0.009	-0.010	-0.008	-0.008	-0.012
		F-			E-			B-		
MF-OLS	0.010			0.001			0.001			
MIDAS	0.001			-0.010			-0.007			
Intercept	$K_L = 0$	0.081	0.011	0.089	0.057	0.012	0.130	0.016	-0.001	0.004
	$K_L = 1$	-0.008	0.003	-0.008	0.005	0.010	-0.008	-0.006	-0.001	-0.005
		F-			E-			B-		
MF-OLS	0.002			0.008			0.015			
MIDAS	0.020			0.013			0.005			
Intercept + Trend	$K_L = 0$	0.063	0.000	0.123	0.033	-0.009	0.127	0.014	0.001	-0.006
	$K_L = 1$	0.000	-0.011	-0.019	-0.007	-0.008	-0.018	-0.007	-0.009	-0.011
		F-			E-			B-		
MF-OLS	0.008			-0.006			-0.016			
MIDAS	0.016			-0.009			-0.004			
Strong Endogeneity, T = 500										
	F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B	
No Intercept	$K_L = 0$	0.010	-0.017	-0.005	-0.002	-0.014	-0.013	-0.006	-0.006	-0.005
	$K_L = 1$	-0.013	-0.024	-0.025	-0.019	-0.015	-0.005	-0.019	-0.009	-0.008
		F-			E-			B-		
MF-OLS	-0.004			-0.013			-0.013			
MIDAS	-0.003			-0.014			-0.004			
Intercept	$K_L = 0$	0.085	-0.005	0.101	0.043	-0.008	0.099	0.004	-0.001	-0.005
	$K_L = 1$	0.002	-0.003	0.005	-0.004	-0.008	-0.003	0.001	-0.004	-0.004
		F-			E-			B-		
MF-OLS	0.011			-0.004			0.000			
MIDAS	0.005			-0.008			-0.005			
Intercept + Trend	$K_L = 0$	0.109	0.012	0.175	0.076	0.017	0.191	0.011	0.003	0.003
	$K_L = 1$	-0.002	0.004	-0.005	0.015	0.013	0.005	-0.011	0.001	0.000
		F-			E-			B-		
MF-OLS	0.010			0.017			0.017			
MIDAS	0.010			0.017			0.017			

Weak Endogeneity, T = 200										
	F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B	
No Intercept	$K_L = 0$	0.009	0.011	0.033	0.039	-0.004	0.045	0.026	0.026	0.018
	$K_L = 1$	0.004	0.007	0.010	0.013	0.003	0.022	0.024	0.025	0.014
		F-			E-			B-		
MF-OLS	0.006			0.003			0.012			
MIDAS	0.003			-0.004			0.018			
Intercept	$K_L = 0$	0.002	0.003	0.107	0.049	0.021	0.114	0.016	0.000	0.001
	$K_L = 1$	0.006	0.002	-0.006	0.014	0.021	0.000	-0.004	-0.001	-0.005
		F-			E-			B-		
MF-OLS	0.003			0.018			-0.004			
MIDAS	0.011			0.020			0.004			
Intercept + Trend	$K_L = 0$	-0.008	0.003	0.127	0.050	-0.009	0.145	0.008	0.004	-0.014
	$K_L = 1$	-0.011	-0.007	-0.013	-0.012	-0.012	-0.017	-0.017	-0.011	-0.014
		F-			E-			B-		
MF-OLS	-0.004			-0.007			-0.008			
MIDAS	0.000			-0.009			-0.017			
Weak Endogeneity, T = 500										
	F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B	
No Intercept	$K_L = 0$	-0.005	-0.011	0.020	0.012	0.008	0.040	0.016	-0.002	-0.007
	$K_L = 1$	0.001	-0.014	-0.003	-0.002	0.011	0.007	0.004	-0.002	-0.004
		F-			E-			B-		
MF-OLS	0.002			0.008			-0.007			
MIDAS	-0.001			0.008			-0.007			
Intercept	$K_L = 0$	0.005	0.013	0.109	0.069	0.012	0.128	0.017	0.012	0.009
	$K_L = 1$	0.000	0.014	0.002	0.012	0.009	0.003	0.001	0.008	0.006
		F-			E-			B-		
MF-OLS	0.009			0.013			0.004			
MIDAS	0.006			0.013			0.007			
Intercept + Trend	$K_L = 0$	0.006	-0.016	0.121	0.048	-0.004	0.156	0.024	-0.002	0.008
	$K_L = 1$	0.001	-0.015	-0.007	-0.006	-0.003	-0.009	-0.006	-0.007	0.006
		F-			E-			B-		
MF-OLS	0.007			-0.002			0.006			
MIDAS	0.005			-0.004			0.007			

benchmark model, setting $K_H = 0$ in (4) is sufficient to control for endogeneity, because the error $(\varepsilon_{t-i/m} - \Delta^{(1/m)}x'_{t-i/m}\pi_0^H)$ in (4) is estimated by least squares to be orthogonal to $(\Delta^{(1/m)}x_{t-i/m})$.

However, aggregation generally creates a first-order moving average, MA(1), in (Δx_t^a) . This result follows from generalizing Working's (1960) well-known result for a single series subject to flat sampling to multivariate series subject to general and possibly different aggregation schemes (Ghysels and Miller, 2015).⁸ For expositional simplicity, denote the MA(1) structure by $\Delta x_t^a = v_t + \Pi v_{t-1}$.

To gauge the effects of the MA(1) serial correlation, first suppose that the aggregation schemes are the same. The error term ε_t^a of (3) is correlated with Δx_t^a and Δx_{t+1}^a due to the MA(1) structure of the latter terms. Adding these as regressors, as in (5), implicitly subtracts them from the error term, as in (9). The new error term, $\varepsilon_t^a - \Delta x_t^{a'}\pi_0^L - \Delta x_{t+1}^a\pi_{-1}^L$, is estimated to be orthogonal to Δx_t^a and Δx_{t+1}^a . This procedure eliminates the correlation between ε_t^a and Δx_t^a , Δx_{t+1}^a , but the inclusion of the MA(1) Δx_t^a introduces Πv_{t-1} . The problem is that Πv_{t-1} is correlated with Δx_{t-1}^a , which is a component of the I(1) regressor x_t^a .

Adding a lag as well as a lead does not eliminate the problem, since there will be correlation between Πv_{t-2} and x_t^a in that case. Increasing K_L does not eliminate the problem either, since there will always be a term of the form $\Pi v_{t-(K_L+1)}$ that is included in the error and is correlated with x_t^a . Fortunately, our simulations show that the size distortion from any such correlation is quite small. As long as the aggregation schemes are the same, we should not expect substantial size distortion from aggregation.

If the aggregation schemes are different, on the other hand, there is an additional term $(x_t^b - x_t^a)'\beta$ in (9). To see why this term is problematic, consider the special case of a HF univariate series $(x_{t-i/m})$ and a LF m -variate series (z_t) constructed by stacking across i : $z_t = (x_t, x_{t-1/m}, \dots, x_{t-(m-1)/m})'$. Stack the aggregation weights by letting $\varpi = (\varpi_1, \dots, \varpi_m)'$ and $v = (v_1, \dots, v_m)'$, so that the aggregated series may be written as $x_t^a = \varpi'z_t$ and $x_t^b = v'z_t$. Now, let ι denote an $m \times 1$ vector of ones and let H 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. z_t may be written as $\iota x_t - H\Delta^{(1/m)}z_t$ by decomposing z_t into a vector ιx_t of the common stochastic trend x_t of z_t less deviations $H\Delta^{(1/m)}z_t$ from that trend. Since the elements of the weight vectors ϖ and v must sum to one, the stochastic trends cancel from the difference of the aggregated series, so that $x_t^b - x_t^a = (\varpi - v)'H\Delta^{(1/m)}z_t$.

Consequently, the additional term $(x_t^b - x_t^a)'\beta$ is correlated with Δx_t^b and Δx_{t+1}^b – but only with those, as long as $(b_{0,t-i/m})$ is iid. The strength of this correlation is determined both by the aggregation schemes and by the magnitude of the cointegrating coefficient β , and our simulations with $\beta = 1$ show that the size distortions may be quite high. Fortunately, orthogonalizing the error with respect to Δx_t^b and Δx_{t+1}^b by adding them as regressors eliminates this correlation. This explains the observed threshold at $K_L = 1$.

Because only one LF lead is required, we could use $\sum_{j=-1}^0 \Delta x_{t-j}^{b'}\pi_j$ in place of $\sum_{j=-1}^1 \Delta x_{t-j}^{b'}\pi_j$ in (5), but the extra lag seems innocuous, so we just set $K_L = 1$ rather than differentiating

⁸The uncorrelatedness is preserved only if all series are skip sampled in the same way.

the endpoints of the summation.

3.2 Power

Since temporal aggregation does not change the long-run properties of a time series, it is reasonable to expect that it would not affect the consistency of the residual based KPSS test that Shin (1994) showed for the disaggregated case. Wang (2015) derived consistency in the LF case, and a proof for the MF case would follow along the same lines. Our concern in this paper is not with an asymptotic result, but rather with a small-sample comparison of the power of the tests in each of the LF and MF cases already considered for size comparisons. For brevity, we present only the results of simulations with $T = 200$. As expected, the results with $T = 500$ are qualitatively similar to those with $T = 200$ but with higher powers reflecting a larger sample.

Table 2 shows the results for power. For the nine LF cases discussed above, we show both $K_L = 0$ and $K_L = 1$. It is clear from our discussion above that the case of $K_L = 0$ may generate substantial size distortion, while that of $K_L = 1$ does not. The rows labeled *vs. $K_L = 0$* show the percentage changes in power from increasing the lag/lead order to 1. Power is lost in 52 out of 54 cases, and the power loss is as high as 5.7%.

The MF cases have $K_L = 0$, since this choice does not lead to any serious size distortions. We wish to know if power can be improved while controlling size by using the HF series in the MF testing environment. In other words, we would like to control size at least as well as the LF case with $K_L = 1$, but we would like to increase power relative to that case. As discussed above, we expect CoMIDAS-DOLS to be more powerful than MF-DOLS, but it is less clear whether or not MF-DOLS should be more powerful than LF-DOLS with $K_L = 1$. The rows labeled *vs. $K_L = 1$* show the percentage changes in power from moving from the LF case with $K_L = 1$ to the respective MF models. MF-DOLS improves power in 52 out of 54 cases, while CoMIDAS-DOLS improves power in *all* cases. The improvements using CoMIDAS-DOLS range from 1.1% to 6.7%. Furthermore, CoMIDAS-DOLS shows an improvement that is larger or the same as that for MF-DOLS in *all* cases.

4 Practical Guidance for Implementation

A few insights emerge from this research that may inform empirical work. First, when testing for cointegration with the KPSS test using data that are observed at the same frequency but known to be temporally aggregated, we should not consider $K_L = 0$ in the DOLS regression. The recommendation holds even in contexts where a reasonable model may feature serially uncorrelated increments, as the efficient market hypothesis suggests would be the case with prices of tradeable assets. In our HF DGP with no serial correlation in the error, $K_L = 1$ was sufficient to eliminate the size distortion from aggregation.

In practice, an information criterion is often used to select K_L over a range $\min(K_L) = 0 \leq K_L \leq \max(K_L)$. Our results suggest that $K_L = 0$ should not be considered when the data are known to be temporally aggregated. Generalizing our results to serially correlated HF increments, the threshold at $K_L = 1$ noted above may be greater. If we observe a lag/lead length, say K_L^* , at which the rejection rate drops off more substantially than at

Table 2: Powers of the KPSS tests based on residuals of HF-DOLS, LF-DOLS, MF-DOLS, and CoMIDAS-DOLS. Results from 1,000 simulations with $m = 12$, $T = 200$, and $\sigma_{21} \in \{1/10, 9/10\}$. F , E , and B denote flat sampling (period average), end-of-period sampling, and beginning-of-period sampling, respectively.

		Strong Endogeneity, T = 200								
		F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B
No Intercept	$K_L = 0$	0.445	0.441	0.439	0.444	0.444	0.445	0.442	0.439	0.442
	$K_L = 1$	0.434	0.435	0.436	0.434	0.434	0.440	0.429	0.429	0.435
	vs. $K_L = 0$	-2.5%	-1.4%	-0.7%	-2.3%	-2.3%	-1.1%	-2.9%	-2.3%	-1.6%
		F-			E-			B-		
	MF-OLS	0.437			0.437			0.434		
	vs. $K_L = 1$	0.7%	0.5%	0.2%	0.7%	0.7%	-0.7%	1.2%	1.2%	-0.2%
	MIDAS	0.445			0.445			0.441		
	vs. $K_L = 1$	2.5%	2.3%	2.1%	2.5%	2.5%	1.1%	2.8%	2.8%	1.4%
		F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B
	$K_L = 0$	0.669	0.672	0.654	0.665	0.672	0.640	0.655	0.663	0.670
$K_L = 1$	0.647	0.645	0.647	0.643	0.647	0.644	0.638	0.641	0.643	
vs. $K_L = 0$	-3.3%	-4.0%	-1.1%	-3.3%	-3.7%	0.6%	-2.6%	-3.3%	-4.0%	
	F-			E-			B-			
MF-OLS	0.656			0.655			0.653			
vs. $K_L = 1$	1.4%	1.7%	1.4%	1.9%	1.2%	1.7%	2.4%	1.9%	1.6%	
MIDAS	0.671			0.671			0.671			
vs. $K_L = 1$	3.7%	4.0%	3.7%	4.4%	3.7%	4.2%	5.2%	4.7%	4.4%	
	F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B	
$K_L = 0$	0.615	0.617	0.618	0.613	0.615	0.570	0.616	0.614	0.619	
$K_L = 1$	0.587	0.586	0.583	0.591	0.591	0.581	0.581	0.590	0.588	
vs. $K_L = 0$	-4.6%	-5.0%	-5.7%	-3.6%	-3.9%	1.9%	-5.7%	-3.9%	-5.0%	
	F-			E-			B-			
MF-OLS	0.612			0.612			0.611			
vs. $K_L = 1$	4.3%	4.4%	5.0%	3.6%	3.6%	5.3%	5.2%	3.6%	3.9%	
MIDAS	0.612			0.616			0.620			
vs. $K_L = 1$	4.3%	4.4%	5.0%	4.2%	4.2%	6.0%	6.7%	5.1%	5.4%	

		Weak Endogeneity, T = 200								
		F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B
No Intercept	$K_L = 0$	0.531	0.527	0.534	0.534	0.534	0.533	0.524	0.524	0.528
	$K_L = 1$	0.513	0.510	0.515	0.518	0.519	0.521	0.507	0.506	0.516
	vs. $K_L = 0$	-3.4%	-3.2%	-3.6%	-3.0%	-2.8%	-2.3%	-3.2%	-3.4%	-2.3%
		F-			E-			B-		
	MF-OLS	0.527			0.530			0.524		
	vs. $K_L = 1$	2.7%	3.3%	2.3%	2.3%	2.1%	1.7%	3.4%	3.6%	1.6%
	MIDAS	0.528			0.532			0.529		
	vs. $K_L = 1$	2.9%	3.5%	2.5%	2.7%	2.5%	2.1%	4.3%	4.5%	2.5%
		F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B
	$K_L = 0$	0.628	0.632	0.630	0.629	0.632	0.626	0.629	0.632	0.634
$K_L = 1$	0.607	0.611	0.609	0.604	0.606	0.603	0.603	0.612	0.611	
vs. $K_L = 0$	-3.3%	-3.3%	-3.3%	-4.0%	-4.1%	-3.7%	-4.1%	-3.2%	-3.6%	
	F-			E-			B-			
MF-OLS	0.619			0.622			0.618			
vs. $K_L = 1$	2.0%	1.3%	1.6%	3.0%	2.6%	3.2%	2.5%	1.0%	1.1%	
MIDAS	0.631			0.628			0.629			
vs. $K_L = 1$	4.0%	3.3%	3.6%	4.0%	3.6%	4.1%	4.3%	2.8%	2.9%	
	F-F	F-E	F-B	E-F	E-E	E-B	B-F	B-E	B-B	
$K_L = 0$	0.602	0.605	0.616	0.607	0.608	0.613	0.598	0.603	0.602	
$K_L = 1$	0.576	0.579	0.580	0.574	0.577	0.578	0.574	0.578	0.575	
vs. $K_L = 0$	-4.3%	-4.3%	-5.8%	-5.4%	-5.1%	-5.7%	-4.0%	-4.1%	-4.5%	
	F-			E-			B-			
MF-OLS	0.602			0.598			0.599			
vs. $K_L = 1$	4.5%	4.0%	3.8%	4.2%	3.6%	3.5%	4.4%	3.6%	4.2%	
MIDAS	0.605			0.607			0.601			
vs. $K_L = 1$	5.0%	4.5%	4.3%	5.7%	5.2%	5.0%	4.7%	4.0%	4.5%	

others, this may be an indication that the lag/lead length that provides the best approximation in the absence of aggregation-induced serial correlation would be $K_L^* - 1$. In that case, we should set $\min(K_L) = K_L^*$ for the selection process to avoid size distortions from aggregation – especially when we know that not all series have been aggregated in the same way.⁹

Second, available HF data may be useful to increase the power of the test, and we suggest the CoMIDAS-DOLS residual-based KPSS test statistic for this purpose. This approach increases power the most relative to all other techniques that have reasonable size properties among those we consider. In the setup that we consider, fully disaggregated regressors with serially uncorrelated increments, setting $K_L = 0$ works well. With possibly serially correlated increments, K_L should be selected with no restriction on the minimum. If, on the other hand, some of the regressors are subject to aggregation, the potential exists for mismatched aggregation schemes and we should restrict the minimum lag/lead as above.

⁹Simulations results with strong first-order autoregressive serial correlation in the difference of the regressor (not shown) also support this approach. In that case, almost all distortions are eliminated by setting $K_L^* = 2$ instead of $K_L^* = 1$.

5 References

- Andreou, E., E. Ghysels, and A. Kourtellis (2010). “Regression Models with Mixed Sampling Frequencies,” *Journal of Econometrics* 158, 246-261.
- Chambers, M.J. (2003). “The Asymptotic Efficiency of Cointegration Estimators under Temporal Aggregation,” *Econometric Theory* 19, 49-77.
- Chambers, M.J. (2011). “Cointegration and Sampling Frequency,” *Econometrics Journal* 14, 156-185.
- Chambers, M.J. (2014). “The Estimation of Continuous Time Models with Mixed Frequency Data,” mimeo, Department of Economics, University of Essex.
- Froni, C., M. Marcellino, and C. Schumacher (2011). “U-MIDAS: MIDAS regressions with unrestricted lag polynomials,” Discussion Paper Series 1: Economic Studies 35, Deutsche Bundesbank, Research Centre.
- Ghysels, E. and J.I. Miller (2015). “Testing for Cointegration with Temporally Aggregated and Mixed-frequency Time Series,” *Journal of Time Series Analysis*, forthcoming.
- 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 (2014). “Forecasting Mixed Frequency Time Series with ECM-MIDAS Models,” *Journal of Forecasting* 33, 198-213.
- 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.
- Hu, W. (1996). “Time Aggregation and Skip Sampling in Cointegration Tests,” *Statistical Papers* 37, 225-234.
- Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992), “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root,” *Journal of Econometrics*, 54, 159-78.
- Lahiri, K. and N. Mamingi (1995) “Testing for Cointegration: Power Versus Frequency of Observation – Another View,” *Economics Letters* 49, 121-124.
- Lee, D. and P. Schmidt (1996). “On the power of the KPSS test of stationarity against fractionally integrated alternatives,” *Journal of Econometrics* 73, 285-302.
- Marcellino, M. (1999). “Some Consequences of Temporal Aggregation in Empirical Analysis,” *Journal of Business and Economic Statistics* 17, 129-136.

- Miller, J.I. (2014). "Mixed-frequency Cointegrating Regressions with Parsimonious Distributed Lag Structures," *Journal of Financial Econometrics* 12, 584-614.
- Miller, J.I. (2015). "Conditionally Efficient Estimation of Long-run Relationships Using Mixed-Frequency Time Series," *Econometric Reviews*, forthcoming.
- Otero, J. and J. Smith (2000) "Testing for Cointegration: Power Versus Frequency of Observation – Further Monte Carlo Results," *Economics Letters* 67, 5-9.
- Shin, Y. (1994). "A Residual-Based Test of the Null of Cointegration Against the Alternative of No Cointegration," *Econometric Theory* 10, 91-115.
- Wang, X. (2015). "The Size Robustness of the Residual-Based KPSS Test with Temporally Aggregated Series," mimeo, Department of Economics, University of Missouri.
- Working, H. (1960). "Note on the Correlation of First Differences of Averages in Random Chain," *Econometrica* 28, 916-918.