

Simple Robust Tests for the Specification of High-Frequency Predictors of a Low-Frequency Series*

J. Isaac Miller[†]

Abstract

I propose two variable addition test statistics aimed at the specification of a high-frequency predictor of a series observed at a lower frequency. Under the null, the high-frequency predictor is aggregated to the low frequency versus mixed-frequency alternatives. The first test statistic is similar to those in the extant literature, but I show its robustness to conditionally biased forecast error and cointegrated and deterministically trending covariates. It is feasible and consistent even if estimation is not feasible under the alternative. However, its size is not robust to nuisance parameters when the high-frequency predictor is stochastically trending, and size distortion may be severe. The second test statistic is a simple modification of the first that sacrifices power in order to correct this distortion. An application to forecasting and nowcasting monthly state-level retail gasoline prices illustrates how the test statistics may be utilized when the presence of nuisance parameters and orders of integration are unknown.

This Version: January 18, 2016

JEL Classification: C12, C22, Q41

Key words and phrases: temporal aggregation, mixed-frequency model, MIDAS, variable addition test, forecasting model comparison, retail gasoline prices

*The author is grateful to participants of the 2014 NBER-NSF Time Series Conference, 2015 International Symposium on Forecasting, 2015 Econometric Society World Congress, 2015 Midwest Econometric Group Meeting, 2015 International Conference on Computational and Financial Econometrics, a mathematical sciences seminar at Michigan Tech, and Graham Elliott for useful comments. All remaining errors are mine.

[†]Department of Economics, University of Missouri, 118 Professional Building, Columbia, Missouri 65211, millerjisaac@missouri.edu.

1 Introduction

The possibility of deleterious effects from aggregation on the properties of time series and forecasting models have long been known (Working, 1960; Lütkepohl, 1987; e.g.). In order to circumvent such problems and exploit both higher frequency and real-time data in forecasting, a large number of recent papers have used a mixed-frequency (MF) approach, exploiting high-frequency (HF) data – usually financial data – to predict low-frequency (LF) data – usually macroeconomic data. These papers include Tay (2007), Clements and Galvão (2008, 2009), Högrefe (2008), Schumacher and Breitung (2008), Armesto *et al.* (2010), Andreou *et al.* (2011, 2013), Marcellino and Schumacher (2010), Kuzin *et al.* (2011), Götz *et al.* (2014), Miller (2014), and Breitung and Røling (2015).

Many of these papers employ a parsimonious distributed lag (DL) structure under the MIDAS moniker, for “MIXed DATA Sampling” (Ghysels *et al.*, 2004, 2006). At one extreme, the most parsimonious DL structure reduces to temporal aggregation or systematic sampling of the HF data. With only a single coefficient to estimate in that case, a researcher must necessarily decide how to choose the vector of aggregation weights. A specific aggregation weight vector is the “fixed” null of the tests proposed in this paper. At the other extreme, the least parsimonious, an unrestricted DL specification, has at least as many parameters to estimate as the ratio m of frequencies in the model. When this ratio is large – e.g., for daily predictors of an annual target – the unrestricted specification is infeasible.

In forecasting models, out-of-sample testing of predictive accuracy along the lines of West (1996) and Clark and McCracken (2001) provides one approach to this problem. In contrast, Diebold (2015) emphasizes the advantage of using in-sample tests to compare forecasting models in order to exploit the full sample available to a researcher. Diebold’s (2015) approach may be viewed as pretesting in the sense that the tests are conducted before making the forecasts, while the out-of-sample tests are conducted after making forecasts.

In the recent MF literature, Andreou *et al.* (2010) propose two statistics for the null of aggregation with equal weights against a MIDAS alternative, while Kvedaras and Zemlys (2012) and Miller (2014) propose statistics for a MIDAS null against the unrestricted DL alternative.

One of the statistics proposed by Andreou *et al.* (2010) and that proposed by Miller (2014) are variable addition test (VAT) statistics. The use of VAT statistics for specification testing is often attributed to Wu (1973) and is of course utilized in a wide variety of econometric settings, most commonly for testing instrumental variable and panel data specifications (Hausman, 1978), but also for testing cointegration (Park, 1990), for example.

A strong advantage of the VAT framework is its simplicity in implementation, requir-

ing only two steps involving least squares. Another particularly useful aspect of the VAT approach in the context of MF series is that the test is implemented without having to estimate the model under the alternative – an impossible task against an unrestricted alternative when m is large relative to the LF sample size T .

In this paper, I introduce two simple VAT statistics. I show that when the HF predictor is $I(0)$, the first VAT statistic, similar to that of Andreou *et al.* (2010) and Miller (2014), is robust to conditionally biased forecast error, infeasible alternatives, and deterministic and stochastic trends in the model covariates – as long as the model specification is such that the forecast error remains $I(0)$.

The possibility that the HF predictor is $I(1)$ – and that the limiting distribution of the VAT statistic contains nuisance parameters as a result – motivates a second, modified VAT statistic. The modified VAT statistic augments the regressand by a term that increases mildly with the sample size. This strategy corrects size distortion from nuisance parameters in the limiting distribution, but at the expense of power. Power generally decreases, but the test remains consistent against fixed and some local alternatives.

Using the second test, the order of integration and number and/or structure of cointegrating relationships do *not* need to be known. Pretesting for unit roots and cointegration is not required, although a number of recent papers have made advances in testing for these in MF and LF environments.¹

The rest of the paper is organized as follows. In Section 2, I detail a general MF forecasting model, in which (at least) one of the predictors is available at a higher frequency than the target, and I motivate and present the two test statistics. Analysis of the large-sample properties of the test statistics – a limiting χ -squared distribution under the null and consistency under the alternative – follows in Section 3. Section 4 explores the small-sample size and power of the statistics under a range of assumptions, while Section 5 introduces a sequential testing strategy and applies that strategy to a forecasting model of state-level retail gasoline prices. Section 6 concludes. A technical appendix contains proofs of useful lemmas and the main theoretical results.

2 Forecasting Model and VAT Statistics

Consider a very general single-frequency forecasting model that uses data available at different frequencies given by

$$y_{t+1} = x_t^p \beta + w_t' \gamma + \varepsilon_{t+1} \quad (1)$$

¹For example, Chambers (2015) analyzes tests for unit roots using LF data, and Ghysels and Miller (2015) analyze tests for cointegration using LF and MF data.

where $t = 1, \dots, T$, and (x_t^p) denotes a single series subject to aggregation or sampling from p observations of a series $(x_{t-i/m})$ with $i = 0, \dots, m - 1$ available $m < \infty$ times more often. For the purpose of testing the specification of (x_t^p) , the remaining observable series in the model (w_t) are treated as nuisance terms, even though these series may also be observable at a higher frequency.

In general, $\mathbf{E}[\varepsilon_{t+1}|\mathcal{F}_t] \neq 0$ – i.e., the forecast error may be conditionally biased. If all series in the model are $I(0)$, this inequality is not problematic. However, if unit roots are present in some of the series, nuisance parameters that contaminate the limiting distributions of standard test statistics may result from this bias (Elliott, 2006).

Denote by $x_t^{1:p} = (x_t, x_{t-1/m}, \dots, x_{t-(p-1)/m})'$ a vector of the p high-frequency observations aggregated to obtain x_t^p , and denote by $\pi = (\pi_1, \dots, \pi_p)'$ the vector of aggregation weights used, so that $x_t^p = x_t^{1:p'}\pi$. Assume that the weights are non-negative and sum to one. For example, stocks are commonly sampled by taking the last observation in each LF period (end-of-period sampling), in which case $\pi_1 = 1$ while $\pi_j = 0$ for $j > 1$. Flat sampling (or average sampling), such that $\pi_j = 1/m$ for $j = 1, \dots, m$ and $\pi_j = 0$ for $j > m$ is often employed when $(x_{t-i/m})$ is a flow. However, the LF series (x_t^p) may result from an unknown or more complicated filter of the HF series, such as seasonal adjustment, in which case it is sensible to estimate π rather than fix it.

The aggregation weight vector π implicitly reflects aggregation of *all* LF variables in the model if the data are generated at a higher frequency than we observe them. Otherwise, if the predictors are aggregated differently than the target is, ε_{t+1} will contain a term that is a function of this difference and is correlated with the predictors – further reason that $\mathbf{E}[\varepsilon_{t+1}|\mathcal{F}_t] = 0$ might not hold.²

2.1 Mixed-Frequency Forecasting Model

Replacing x_t^p with $x_t^{1:p'}\pi$ in (1) yields

$$y_{t+1} = x_t^{1:p'}\alpha + w_t'\gamma + \varepsilon_{t+1} \quad (2)$$

where $\alpha = \pi\beta$. Let $\alpha_m = \pi_m\beta_m$ be the value of α that minimizes mean squared forecast error (MSFE) of the model. The minimum MSFE coefficient vector α_m does not generally equal α – specifically when $\pi_m \neq \pi$ or more generally when $\mathbf{E}[\varepsilon_{t+1}|\mathcal{F}_t] \neq 0$.

So that the aggregation weight vector is identified, assume that $\beta_m \neq 0$. This assumption rules out the case that $(x_t^{1:p'})$ is the *only* $I(1)$ series in the model. More generally, the

²Eliminating this difference by setting the aggregation schemes to be equal might *not* be optimal, since Miller (2015) shows the possibility that β may be estimated more efficiently in some cases. The efficiency gain could correspond to forecasts with lower mean squared forecast errors in the present context.

predictor of interest is required to have predictive power. Otherwise, π is not identified under any null or alternative and becomes irrelevant. If β_m is known to be 0, no test of π is needed. If β_m is unknown, but is in fact 0, then the difference between forecasts using different aggregation weight vectors is a function of shrinking estimation error.

The weight vector π may be conceptualized as a function $\pi(\theta)$ of some vector of $\ell \times 1$ reduced form parameter θ . In the most general case, $\ell = p$ and $\pi(\theta) = \theta$, so that the MF model in (2) is an unrestricted distributed lag (DL) model. Predictions may be made using a least squares estimate of the p -vector $\alpha = \theta\beta$, as long as p is small relative to the LF sample size T . Because p is often quite large compared to T , the unrestricted approach may not be feasible. An alternative approach is to fix θ to some value θ_0 , so that $\pi(\theta) = \pi(\theta_0)$, denoted for short by π_0 . This means that a known, fixed aggregation scheme given by π_0 is imposed. The MF model in (2) reduces to the single-frequency model in (1), and predictions may be made using a least squares estimate of β with a scalar LF predictor given by $x_t^{1:p'}\pi_0$.

The problems of large p and unknown π motivate the MIDAS approach, a compromise in which a parsimonious structure is imposed on π so that ℓ is small, say 2, and $\alpha(\theta) = \pi(\theta)\beta$ is estimated using nonlinear least squares. The asymptotic analysis below requires $\pi(\theta)$ to be continuously differentiable in θ , which is satisfied by common lag functions employed in the MF literature, such as the exponential Almon lag.

The focus of this analysis is on testing the fixed null of a known aggregation scheme $\pi_m = \pi_0$. Under the alternative, either $\ell = p$ (unrestricted DL) or $\ell \ll p$ (parsimonious MIDAS). Local alternatives may take the form $\pi_A = \pi(\theta_A) = \pi(\theta_0 + cg_T^{-1})$, where c is a constant $\ell \times 1$ vector and $g_T > 0$ is a scalar-valued non-decreasing function. This set of hypotheses generalizes those of Andreou *et al.* (2010) in the sense that they consider a flat sampling null (a particular π_0) against a MIDAS alternative ($\ell \ll p$). Since the tests proposed here do not require the model to be estimated under the alternative, they also have power against the unrestricted DL case in which $\ell = p$.

The orders of integration of $(x_{t-i/m})$ and (y_{t+1}) are completely general, and (w_t) may contain I(0) series, I(1) series, and deterministic trends. Because of the allowance for an intercept, all other predictors are assumed to have zero mean. In order to avoid spurious forecasts, (ε_{t+1}) must be I(0), requiring the coefficients or linear combinations of the coefficients on I(1) series to be zero and superconsistently estimated to be so. The order of integration will be discussed in much more detail below.

2.2 Variable Addition Test Statistic

The main attraction of the variable addition strategy lies in its simplicity even while it remains robust to a wide variety complications under both the null and the alternative.

The implementation of the basic VAT statistic proposed here for a fixed null follows that proposed by Miller (2014) for a MIDAS null. Specifically, the fitted residual from estimating (2) under the null, denoted by $\hat{\varepsilon}_{t+1}$, is regressed on linear combinations of the elements of the vector $x_t^{1:p}$.

In order to see the efficacy of the VAT strategy, rewrite the model in (2) as

$$y_{t+1} = x_t^{1:p'} \pi_0 \beta_m + w_t' \gamma + \varepsilon_{t+1} - x_t^{1:p'} (\alpha_m - \alpha) - x_t^{1:p'} (\pi_0 - \pi_m) \beta_m \quad (3)$$

by imposing the fixed null π_0 with the minimum MSFE coefficient β_m . The fitted residuals may be written as

$$\hat{\varepsilon}_{t+1} = \varepsilon_{t+1} - x_t^{1:p'} \pi_0 (\hat{\beta} - \beta_m) - w_t' (\hat{\gamma} - \gamma) - x_t^{1:p'} (\alpha_m - \alpha) - x_t^{1:p'} (\pi_0 - \pi_m) \beta_m \quad (4)$$

where $\hat{\beta}$ and $\hat{\gamma}$ denote least squares estimates. $(\alpha_m - \alpha)$ reflects the difference between the minimum MSFE parameter vector and the parameter vector given in the model, which is not zero when $\mathbf{E}[\varepsilon_{t+1} | \mathcal{F}_t]$ is not. $(\pi_0 - \pi_m) \beta_m$ reflects the difference between the null and the weights π_m implied by the minimum MSFE parameter vector α_m .

The variable addition test is implemented as a Wald test on an ancillary regression of the fitted residual $\hat{\varepsilon}_{t+1}$ onto $x_t^{1:p'} \pi_0$ and q additional variables (instruments) given by linearly independent combinations of $x_t^{1:p}$. The term $x_t^{1:p'} \pi_0$ is retained in the test regression to offset the correlation of $x_t^{1:p'} (\alpha_m - \alpha)$ with the instruments when $\mathbf{E}[\varepsilon_{t+1} | \mathcal{F}_t] \neq 0$ and more generally to control for any correlation or cointegration between the combinations constituting the instruments and the combination implied by aggregation under the null. Let Υ be the $p \times (q+1)$ matrix with columns given by π_0 and the q instruments. The ancillary regression for the test is

$$\hat{\varepsilon}_{t+1} = x_t^{1:p'} \Upsilon \varphi + e_{t+1} \quad (5)$$

where φ is a $(q+1)$ -vector of coefficients on data given by $x_t^{1:p'} \Upsilon$.

The instruments pick up correlations between $x_t^{1:p}$ and $x_t^{1:p'} (\pi_0 - \pi_m) \beta_m$ in $\hat{\varepsilon}_{t+1}$. The latter is obviously zero when $\pi_m = \pi_0$ under the null, so that the q coefficients on the instruments are all zero. The VAT statistic is natural here, because linear combinations of $x_t^{1:p}$ other than π_0 should be correlated with $x_t^{1:p'} (\pi_0 - \pi_m) \beta_m$, which may be viewed as an omitted variable under the alternative. The coefficients of these combinations are non-zero, providing a basis for discriminatory power under the alternative.

The VAT statistic is simply a standard regression-based test of the null that all coeffi-

cients but the first are zero. The least squares estimator of φ is given by

$$\hat{\varphi} = (\Upsilon' M_{xx}^T \Upsilon)^{-1} \Upsilon' \left(T^{-1} \sum x_t^{1:p} \hat{\varepsilon}_{t+1} \right), \quad (6)$$

where $M_{xx}^T = T^{-1} \sum x_t^{1:p} x_t^{1:p'}$, and the test may be implemented as a Wald statistic given by

$$V_T = T \hat{\varphi}'_q \hat{\Omega}_q^{-1} \hat{\varphi}_q, \quad (7)$$

where $\hat{\varphi}_q$ denotes the last q elements of $\hat{\varphi}$ and $\hat{\Omega}_q$ is a consistent estimator of the asymptotic variance Ω_q of $\hat{\varphi}_q$. In the absence of any serial correlation or heteroskedasticity in the error term (ε_{t+1}) in (2), V_T reduces to q multiplied by an F -test statistic, because these properties – or a lack of them – are passed through to (e_{t+1}) .

How should Υ be chosen? The combinations in Υ must be linearly independent and $q \leq p - 1$, so that Υ has full column rank of $q + 1$. Moreover, the limit of M_{xx}^T is shown below to have a rank of p (or $p - 1$ in the $I(0)$ direction), so $q \leq p - 1$ ensures that the limit of the inverse in (6) is well-defined.

The VAT statistic considered by Andreou *et al.* (2010) is quite similar to the test proposed here with $q = p - 1$. Specifically, they propose using the unrestricted regressor vector $x_t^{1:p}$ as instruments, which amounts to Υ being a $p \times p$ identity matrix. For reasons already discussed, the first column of Υ is π_0 in the tests considered here. As long as the $p - 1$ aggregation vectors used in the instruments are linearly independent of each other and of π_0 , Υ is $p \times p$ and full rank, so that this instrumentation scheme is just a full rank transformation of that used by those authors.

Of course, as Andreou *et al.* (2010) noted, the unrestricted DL approach is not feasible when p is large relative to T . This case motivates the use of linear combination with q much smaller than $p - 1$ proposed by Miller (2014) for a MIDAS null and in the present analysis for a fixed null.

The simplicity of the VAT statistic may be summarized by its construction in just two steps:

1. Estimate the null model in (2), and retain the fitted residuals ($\hat{\varepsilon}_{t+1}$).
2. Regress the fitted residuals on $x_t^{1:p'} \pi_0$ and $q \leq p - 1$ additional linearly independent combinations of $x_t^{1:p}$, as in (5). Test the null that all but the first regression coefficient are zero using the Wald statistic V_T in (7).

As shown below, the test statistic may be compared to χ_q^2 critical values.

2.3 Modified Variable Addition Test Statistic

Suppose that $x_t^{1:m} \sim I(1)$ and cointegrated either with $I(1)$ (y_{t+1}) or otherwise with $I(1)$ series in (w_t) , so that $\varepsilon_{t+1} \sim I(0)$. The presence of nuisance parameters in the limiting distributions of test statistics conducted on cointegrating regressions is very widely documented. Elliott (2006) notes the possibility over-rejection in hypothesis testing of forecasting models with $I(1)$ or near $I(1)$ data in particular.

A nuisance parameter problem arises when $\mathbf{E}[\varepsilon_{t+1}|\mathcal{F}_t] \neq 0$, which shows up in $x_t^{1:p'}(\alpha_m - \alpha)$ in the fitted residuals in (4). Since the fitted residuals are orthogonal to $x_t^{1:p'}\pi_0$, including the latter term in the test regression controls for this term in the $I(0)$ case. In the $I(1)$ case, $x_t^{1:p'}(\alpha_m - \alpha)$ is orthogonal to cointegrating but not $I(1)$ combinations of $x_t^{1:p}$, as shown in Lemma A2 of the appendix. Since π_0 does *not* cointegrate $x_t^{1:p}$, including $x_t^{1:p'}\pi_0$ in the test regression does not effectively control for $x_t^{1:p'}(\alpha_m - \alpha)$ so that the test regression may pick up correlations between the q instruments based on $x_t^{1:p}$ and $x_t^{1:p'}(\alpha_m - \alpha)$, causing over-rejection under the null.

If $(x_t^{1:m})$ were known to be $I(1)$ and the way in which the VAT test depended on nuisance parameters were known, it might be possible to design a test that eliminates the nuisance parameters, as unit root tests and cointegration tests generally do. These are not typically known with certainty. The nuisance parameter problem may be circumvented using a bootstrapping strategy, but such a strategy would likely require the user to know the order of integration so that bootstrap draws that preserve correlations between the error and predictors could be drawn from levels or differences of the predictors.

I propose a second, new, *modified* VAT statistic, in which variables are added to both sides of the ancillary regression in (5). In addition to the linear combinations of the HF predictors on the right-hand side of the ancillary regression, the modified test simply adds randomly generated error scaled by a nondecreasing function of the sample size to the regressand of the ancillary regression. The resulting test statistic has a χ_q^2 limiting distribution under the null and is consistent under the alternative, as long as the function doesn't increase too fast.

Specifically, replace (5) with

$$\hat{\varepsilon}_{t+1}^* = x_t^{1:p'}\Upsilon\varphi + e_{t+1}^* \quad (8)$$

where $\hat{\varepsilon}_{t+1}^* = \hat{\varepsilon}_{t+1} + T^\epsilon u_{t+1}$ and $e_{t+1}^* = e_{t+1} + T^\epsilon u_{t+1}$ with $u_{t+1} \sim \mathbf{iidN}(0, \sigma_u^2)$ and $0 < \epsilon < 1/2$. The Wald statistic in (7) may be replaced by V_T^* , which is simply q multiplied by an F -test statistic of the null that all but the first element of φ are zero in the ancillary regression, because the added iid error term dominates.

The two steps to implement the test are summarized as follows:

1. Estimate the null model in (2), and retain the fitted residuals ($\hat{\varepsilon}_{t+1}$).
2. Regress the *modified* fitted residuals on $x_t^{1:p'}\pi_0$ and $q \leq p - 1$ additional linear combinations of $x_t^{1:p}$, as in (8). Test the null that all but the first regression coefficient are zero using the *simplified* Wald statistic V_T^* , q multiplied by an F -test statistic.

Clearly the modification does not substantially complicate and may in fact simplify the implementation of the test.

Aside from simplicity, an advantage of the proposed test is the robustness of its asymptotic size to unit roots and the nuisance parameter problem noted above. This statistic is also robust to stationary serial correlation and heteroskedasticity of the error term in (2). The obvious disadvantage is power loss in finite samples and even in large samples against local alternatives, depending on the function g_T under the alternative $\theta_A = \theta_0 + cg_T^{-1}$. Following the large-sample analysis in the next section, the small-sample analysis sheds light on this size-power trade-off.

3 Large-Sample Analysis

Before a more detailed discussion of the I(0) and I(1) cases, it will be useful to introduce some assumptions and notation common to these cases. Recall that (w_t) may have I(0), I(1), and deterministic predictors, regardless of the order of integration of the HF predictor. Denote by (w_{0t}) all of the I(0) series and linear combinations of the I(1) series in the direction of their cointegrating relationships, if any, denote by (w_{1t}) linear combinations of the I(1) series in the direction of their stochastic trends, and denote the deterministic trends by (w_{2t}) with individual trends – intercept or linear trend, e.g. – given by (w_{2jt}) .

The linear combinations of the I(1) series in (w_{0t}) and (w_{1t}) are partitions of an orthogonal matrix along the lines of Phillips (1988) and discussed in more detail below for the case in which $(x_{t-i/m})$ is I(1). The coefficients on (w_{1t}) are zeros, unless cointegrated with (y_{t+1}) , but these zeros reflect linear combinations of γ rather than γ itself, unless the I(1) series in (w_t) do not share any stochastic trends. Elements of γ that must be zero to ensure that the error is I(0) are still estimated, since these are not generally known to be zero. The distinction between the predictors (w_t) and their linear combinations in (w_{0t}) and (w_{1t}) is purely for the theoretical analysis and does not require any consideration in the setup of the model or implementation of the tests.

Let (ξ_t) generically denote all of the I(0) series in the model. The series (ε_{t+1}) , (w_{0t}) , and (Δw_{1t}) will always belong to (ξ_t) , but the remaining components will vary depending on the specification of $(x_{t-i/m})$ as discussed in Sections 3.1 and 3.2 below.

A law of large numbers (LLN) and central limit theorem (CLT) for the sample variance of (ξ_t) , an invariance principle (IP) for (ξ_t) , and a formal assumption about the asymptotic order of the deterministic sequences (w_{2jt}) will be critical for the ensuing asymptotic analysis of the test. Recall that (ξ_t) is assumed to have a mean of zero since an intercept may be included in the model, and I also assume that it has a negligible initial value for the purpose of the IP.

Assumption 1.

- (a) $T^{-1} \sum \xi_t \xi_t' \rightarrow_p \Sigma$ (LLN),
- (b) $T^{-1/2} \sum (\xi_t \xi_t' - \Sigma) \rightarrow_d \mathbf{N}(0, \lim_{T \rightarrow \infty} \text{var}(T^{-1/2} \sum \xi_t \xi_t'))$ (CLT),
- (c) $T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_t \rightarrow_d B(r) = \mathbf{BM}(\Omega)$ for $r \in [0, 1]$ (IP), and
- (d) $T^{-\delta_j} w_{2j \lfloor Tr \rfloor} \rightarrow f_j(r)$ for square integrable f_j and $\delta_j \geq 0$.

Σ denotes $\mathbf{var}(\xi_t)$ and the long-run variance is given by $\Omega = \lim_{T \rightarrow \infty} \mathbf{var}(T^{-1/2} \sum \xi_t)$. Strictly speaking Σ is not a ‘‘contemporaneous’’ variance, since it includes autocovariances of the HF predictors. However, it is contemporaneous at the LF. Assumption 1 may be derived under more primitive assumptions, but at the cost of extraneous notation.

Because the components of (w_t) are treated as nuisance predictors, it will be useful in the subsequent discussion to define additional notation to condition these out of the model. First, let

$$\begin{bmatrix} M_{\varepsilon\varepsilon}^T & M_{\varepsilon x}^T & M_{\varepsilon w}^T \\ M_{x\varepsilon}^T & M_{xx}^T & M_{xw}^T \\ M_{w\varepsilon}^T & M_{wx}^T & M_{ww}^T \end{bmatrix} = T^{-1} \sum \begin{bmatrix} \varepsilon_{t+1}^2 & \varepsilon_{t+1} x_t^{1:p'} & \varepsilon_{t+1} w_t' \\ x_t^{1:p} \varepsilon_{t+1} & x_t^{1:p} x_t^{1:p'} & x_t^{1:p} w_t' \\ w_t \varepsilon_{t+1} & w_t x_t^{1:p'} & w_t w_t' \end{bmatrix},$$

and then let $M_{xx \perp w}^T = M_{xx}^T - M_{xw}^T (M_{ww}^T)^{-1} M_{wx}^T$ and $M_{x\varepsilon \perp w}^T = M_{x\varepsilon}^T - M_{xw}^T (M_{ww}^T)^{-1} M_{w\varepsilon}^T$. Naturally, these reflect sample moments of the residuals from regressing $(x_t^{1:p})$ and (ε_{t+1}) on (w_t) .

By regressing out (w_t) , the least squares estimator of β_m in (3) may be written as

$$\hat{\beta} - \beta_m = (\pi_0' M_{xx \perp w}^T \pi_0)^{-1} \pi_0' (P_T + R_T(\pi_m))$$

where $P_T = M_{x\varepsilon\perp w}^T - M_{xx\perp w}^T(\alpha_m - \alpha)$ and $R_T(\pi) = M_{xx\perp w}^T(\pi - \pi_0)\beta_m$. Substituting $\hat{\beta} - \beta_m$ into the residuals in (4), some algebra shows that the key moment in the least squares estimator in (6) of the ancillary regression in (5) may be written as

$$\Upsilon' \left(T^{-1} \sum x_t^{1:p} \hat{\varepsilon}_{t+1} \right) = \Upsilon'(I - A_T)(P_T + R_T(\pi_m)). \quad (9)$$

where $A_T = M_{xx\perp w}^T \pi_0 (\pi_0' M_{xx\perp w}^T \pi_0)^{-1} \pi_0'$. Note that $R_T(\pi_m) = R_T(\pi_0) = 0$ under the null, but $R_T(\pi_m) \neq 0$ when the null is falsely imposed.

Clearly, the rates of convergence and limiting distributions of the key moment in (9) used to determine that of $\hat{\varphi}$ in (6) and thus of the VAT statistic in (7) critically depend on the rates and limits of $M_{x\varepsilon\perp w}^T$ and $M_{xx\perp w}^T$, which in turn depend on the order of integration of $(x_{t-i/m})$.

3.1 I(0) High-Frequency Predictor

First, suppose that the HF series $(x_{t-i/m})$ is I(0). The target (y_{t+1}) may also be I(0), or it may be I(1) and cointegrated with I(1) series in (w_t) . One such predictor in (w_t) might be (y_t) or another lag of the target. If both $(x_{t-i/m})$ and (y_{t+1}) are I(0), then all I(1) series in (w_t) must be cointegrated. Since we treat (w_t) as nuisance predictors for the purpose of the test, such specification does not affect the VAT statistic's limiting distribution, as long as the error (ε_{t+1}) is I(0).

The I(0) case is relatively straightforward once some additional notation is defined. Let (ξ_t) , defined above as all I(0) series in the model, be $\xi_t = (\varepsilon_{t+1}, w'_{0t}, \Delta w'_{1t}, x_t^{1:p'})'$, so that its variance may be partitioned as

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon 0} & \sigma_{\varepsilon 1} & \sigma_{\varepsilon x} \\ \sigma_{0\varepsilon} & \Sigma_{00} & \Sigma_{01} & \Sigma_{0x} \\ \sigma_{1\varepsilon} & \Sigma_{10} & \Sigma_{11} & \Sigma_{1x} \\ \sigma_{x\varepsilon} & \Sigma_{x0} & \Sigma_{x1} & \Sigma_{xx} \end{bmatrix}.$$

It will be useful to define $\Sigma_{xx\perp 0} = \Sigma_{xx} - \Sigma_{x0}\Sigma_{00}^{-1}\Sigma_{0x}$ and $\sigma_{x\varepsilon\perp 0} = \sigma_{x\varepsilon} - \Sigma_{x0}\Sigma_{00}^{-1}\sigma_{0\varepsilon}$ to be the variance of $(x_t^{1:p})$ and the covariance of $(x_t^{1:p})$ and (ε_{t+1}) in the direction orthogonal to (w_{0t}) .

The following theorem gives the limiting distribution and consistency of the VAT statistic for an I(0) HF predictor.

Theorem 1. *Consider the test statistic V_T with $x_{t-i/m} \sim I(0)$. Assume that $\alpha_m \neq 0$ and that Assumption 1 holds.*

[a] Under the null hypothesis $\pi(\theta_0)$, $V_T \rightarrow_d \chi_q^2$, while

[b] Under an alternative hypothesis $\pi(\theta_0 + cg_T^{-1})$ with continuously differentiable π ,
 $V_T = O_p(Tg_T^{-2})$,

as $T \rightarrow \infty$.

The results support [a] the asymptotic validity of χ_q^2 critical values and [b] the consistency of the test statistic against an alternative with $g_T = o(T^{1/2})$. If $g_T = 1$ (a fixed alternative), then $V_T = O_p(T)$ and the rate at which power increases to one is maximized. At the other extreme, if $g_T = T^{1/2}$, then $V_T = O_p(1)$ suggests that – although the test may have some power, since the limit is not χ_q^2 – the test is inconsistent.

Even without considering the possibility of a unit root in $(x_{t-i/m})$, the result in part [a] of the theorem substantively extends that of Andreou, *et al.* (2010) for the flat null. First, although this result is for a fixed null, it may be a null other than flat sampling. Second, the instrumentation scheme allows $q < p - 1$, similarly to that of Miller (2014), so that the test is feasible when p is large relative to T . Third, the model is quite a bit more general, allowing the error to be correlated with the predictors and allowing deterministic trends, unit roots, etc. in the other predictors.

3.2 I(1) High-Frequency Predictor

Now suppose that $(x_{t-i/m})$ is I(1). It is well known (Phillips, 1991; Marcellino, 1999, e.g.) that aggregation preserves the long-run properties, so (x_t^p) and $(x_t^{1:p})$ are also I(1). The p -vector $(x_t^{1:p})$ is clearly mutually cointegrated by $p - 1$ linearly independent cointegrating relationships and has a single stochastic trend. Since π_0 is known under the null, $x_t^{1:p'} \pi_0 \sim I(1)$ – i.e., there is no possibility that estimation error in π lies in the cointegrating space of $(x_t^{1:p})$, in contrast to the MIDAS null studied by Miller (2014). Nevertheless, the asymptotic analysis is still complicated by the cointegration of the regressors in the test regression.

Two subcases of the I(1) case must be considered. If $(x_{t-i/m})$ is not cointegrated with any other I(1) predictors, then any aggregation of this predictor must be cointegrated with (y_{t+1}) under the maintained assumption that $\alpha_m \neq 0$. Call this subcase A. Let $H^x = (H_0^x, H_1^x)$ be an orthogonal matrix, such that $H_0^x x_t^{1:p} = x_{0t}$ with $x_{0t} \sim I(0)$, while $H_1^x x_t^{1:p} = x_{1t}$ with $x_{1t} \sim I(1)$, along the lines of Phillips (1988). Obviously, x_{1t} is a scalar, since $(x_t^{1:p})$ has only a single common stochastic trend. Moreover, $x_t^{1:p'} \pi = x_{1t} H_1^{x'} \pi$, because $H_0^{x'} \pi = 0$ for any non-negative aggregation weight vector. In words, aggregation of a HF I(1) series cannot cointegrate it.

In subcase B, $(x_t^{1:p})$ may be cointegrated with other predictors, such as (y_t) or another lag of (y_{t+1}) . Let \tilde{x}_t denote a vector consisting of $x_t^{1:p}$ and all predictors cointegrated with it – e.g., $\tilde{x}_t = (x_t^{1:p'}, y_t)'$. Redefine $H^x = (H_0^x, H_1^x)$, such that $H_0^x \tilde{x}_t = x_{0t}$ and $H_1^x \tilde{x}_t = x_{1t}$. The dimensions of x_{0t} and x_{1t} depend on the number of common stochastic trends in (\tilde{x}_t) . When $\tilde{x}_t = (x_t^{1:p'}, y_t)'$, x_{1t} is still a scalar, since the $p + 1$ series in (\tilde{x}_t) still share only a single stochastic trend.

To accommodate the I(1) case, redefine $\xi_t = (\varepsilon_{t+1}, w'_{0t}, \Delta w'_{1t}, x'_{0t}, \Delta x'_{1t})'$. In subcase A, x_{0t} and x_{1t} are defined as above, and w_{0t} and w_{1t} are defined as in the previous section. In subcase B, x_{0t} and x_{1t} are defined with \tilde{x}_t as in the preceding paragraph, and w_{0t} and w_{1t} are defined similarly to subcase A, but do not include series such as (y_t) that are included in (\tilde{x}_t) due to cointegration with $(x_t^{1:p})$. The distinction between these two cases is purely theoretical – any series in (w_t) that is cointegrated with $(x_t^{1:p})$ remains in (w_t) for the purpose of the test. However, for expositional simplicity, I assume hereafter that subcase A holds when $x_{t-i/m} \sim I(1)$.

The variance of ξ_t is now partitioned as

$$\Sigma = \begin{bmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon 0} & \sigma_{\varepsilon 1} & \sigma_{\varepsilon x} & \sigma_{\varepsilon z} \\ \sigma_{0\varepsilon} & \Sigma_{00} & \Sigma_{01} & \Sigma_{0x} & \Sigma_{0z} \\ \sigma_{1\varepsilon} & \Sigma_{10} & \Sigma_{11} & \Sigma_{1x} & \Sigma_{1z} \\ \sigma_{x\varepsilon} & \Sigma_{x0} & \Sigma_{x1} & \Sigma_{xx} & \Sigma_{xz} \\ \sigma_{z\varepsilon} & \Sigma_{z0} & \Sigma_{z1} & \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}.$$

The subscript x now corresponds to I(0) combinations of $(x_t^{1:p})$ or (\tilde{x}_t) instead of corresponding directly to $(x_t^{1:p})$ in the I(0) case, and the subscript z corresponds to the stochastic trend(s) of $(x_t^{1:p})$ or (\tilde{x}_t) . $\Sigma_{xx\perp 0}$ and $\sigma_{x\varepsilon\perp 0}$, are defined as in the I(0) case but with (x_{0t}) instead of $(x_t^{1:p})$.

The modified VAT statistic allows a robust asymptotic result, given in the following theorem.

Theorem 2. *Consider the test statistic V_T^* with $x_{t-i/m} \sim I(0)$ or $x_{t-i/m} \sim I(1)$. Assume that $\alpha_m \neq 0$ and that Assumption 1 holds.*

- [a] *Under the null hypothesis $\pi(\theta_0)$, $V_T^* \rightarrow_d \chi_q^2$, while*
- [b] *Under an alternative hypothesis $\pi(\theta_0 + cg_T^{-1})$ with continuously differentiable π , $V_T^* = O_p(T^{1-2\varepsilon} g_T^{-2})$ when $x_{t-i/m} \sim I(0)$ and $V_T^* = O_p(T^{2(1-\varepsilon)} g_T^{-2})$ when $x_{t-i/m} \sim I(1)$,*

as $T \rightarrow \infty$.

Part [a] precisely echoes that of Theorem 1, but holds when $x_{t-i/m} \sim I(1)$ and nuisance parameters would otherwise undermine the limit of V_T . Part [b] of Theorem 2 is the source of power loss in comparison with part [b] of Theorem 1. Adding $T^\epsilon u_t$ results in the correct asymptotic size, but at the expense of power.

Test consistency requires $T^{1-2\epsilon} g_T^{-2} \rightarrow \infty$ in the purely $I(0)$ case. If $g_T = \text{const}$ (a fixed alternative), then $O_p(T^{1-2\epsilon} g_T^{-2}) = O_p(T^{1-2\epsilon})$, so that the test is consistent against fixed alternatives as long as $\epsilon < 1/2$. For $\epsilon = 1/2$, the test *may* have power against a fixed alternative asymptotically, but the power is *not* one and the test is therefore inconsistent. The test is inconsistent against local alternatives such that $g_T = T^{1/2-\epsilon}$ or g_T increases faster than $T^{1/2-\epsilon}$.

When $(x_{t-i/m})$ has a unit root, a weaker condition that $T^{2(1-\epsilon)} g_T^{-2} \rightarrow \infty$ allows $\epsilon < 1$ for consistency. In case the order of integration is unknown, $\epsilon < 1/2$ should be set, but power against a given alternative is expected to be greater if a unit root is in fact present.

The appeal of the modified VAT statistic, as Theorem 2 elucidates, is that knowledge of the order of integration is not required. Unit root pretesting is not necessary to test for the optimal aggregation or mixed-frequency specification using V_T^* . However, this knowledge increases power: if $x_{t-i/m} \sim I(0)$ is known, V_T is correctly sized and more powerful than V_T^* .

Otherwise, choosing σ_u^2 and ϵ is not obvious. The following small-sample analysis sheds light on the size-power trade-off between the unmodified and modified VAT tests.

4 Small-Sample Analysis

The main issues on the implementation of the test that have yet to be addressed pertain to the selection of the instruments (variables added) in both VAT statistics and the parameters σ_u^2 and ϵ of the variable added in the modified VAT statistic. In particular, the latter implies a trade-off between size and power that needs to be addressed. I rely on Monte Carlo simulations for these purposes.

The HF data-generating process (DGP) is essentially that of Miller (2014). I consider $y_{t+1-i/m} = x_{t-i/m} \beta + \varepsilon_{t+1-i/m}$, where

$$\zeta_{t+1-i/m} = \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix} \zeta_{t+1-(i+1)/m} + \xi_{t+1-i/m}, \quad \text{with } \xi_{t+1-i/m} \sim \text{iidN} \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

and $\zeta_{t+1-i/m} = (\varepsilon_{t+1-i/m}, x_{t-i/m})'$. In this DGP, $d = \{0, 1\}$ denotes the order of integration

of (x_{t-im}) , $\rho = \{0, 1/2\}$ and $\beta = 10$. With a non-zero nuisance parameter ρ , $\beta_m \neq \beta$ is implied, so that the coefficient that minimizes MSFE is not the same as that in the DGP.

Mixed frequencies are introduced by aggregating $(y_{t+1-i/m})$ with $p = m$. The span is fixed at $T = 200$, but the frequency is varied by letting $m = \{4, 150, 365\}$. Two are chosen as representative of quarters (4) and days (365) per year. The intermediate choice (150) requires a large number of degrees of freedom, but is still small enough (compared to T) to be feasible under an alternative of an unrestricted DL with $\ell = m$ or when using $q = m - 1$ instruments.

Since the predictor is assumed to be observable only at the LF, it is aggregated such that $y_{t+1} = y_{t+1}^{1:m'} \pi(\theta)$ implying an error term of $\varepsilon_{t+1} = \varepsilon_{t+1}^{1:m'} \pi(\theta)$ with $y_{t+1}^{1:m} = (y_{t+1}, \dots, y_{t+1-(m-1)/m})'$ and $\varepsilon_{t+1}^{1:m} = (\varepsilon_{t+1}, \dots, \varepsilon_{t+1-(m-1)/m})'$ defined analogously to $x_t^{1:m}$. Consider 3 models given by

$$\begin{aligned} \text{Model 1: } & y_{t+1} = x_t^{1:p'} \pi \beta + \varepsilon_{t+1} & d = 0 \\ \text{Model 2: } & y_{t+1} = x_t^{1:p'} \pi \beta + \varepsilon_{t+1} & d = 1 \\ \text{Model 3: } & y_{t+1} = y_t \gamma_1 + x_t^{1:p'} \pi \beta + \varepsilon_{t+1} & d = 1 \end{aligned}$$

The first two models are the simplest forms of (2) with I(0) and cointegrated I(1) series. The third model adds a single lag of the target. The latter is characteristic of forecasting practice and illustrates a case with cointegrated predictors. Note that the asymptotic results are robust cointegrated predictors, so no major differences are expected between Models 2 and 3.

In order to generate aggregation weights under the null and alternatives, I use $\pi_j(\theta) = (2 - j/m)^{4\theta} / \sum_{u=1}^m (2 - u/m)^{4\theta}$ for $j = 1, \dots, m$ with scalar θ – i.e., $\ell = 1$ – and such that $\pi(\theta) = (\pi_1(\theta), \dots, \pi_m(\theta))'$. I consider the interval $\theta = [0, 2]$, where $\theta = 0$ corresponds to the null (flat sampling) with increasing weight given to π_1 , the weight on the most recent HF observation, up to $\theta = 2$. At $\theta = 2$, the weight given to π_1 is effectively one with all other weights effectively zero (end-of-period sampling). This specification is similar to the exponential Almon lag popular in the mixed-frequency literature, but is numerically more stable for the larger values of $m = 150$ and $m = 365$ in this analysis. From the perspective of local alternatives, $\theta = \theta_0 + c/T$ with $\theta_0 = 0$ and $c = [0, 400]$.

I consider two instrumentation schemes. To implement the unrestricted scheme with $q = m - 1$, which I refer to as the qU scheme (U for unrestricted), I set the last $m - 1$ columns of Υ to be the last $m - 1$ columns of an $m \times m$ identity matrix. For the second scheme, I keep the columns of Υ to 3 by choosing only $q = 2$ instruments regardless of m , which I refer to as $q2$. These instruments have aggregation weights given by a geometrically decaying sequence $0.9^{j-1} / \sum_{u=1}^m 0.9^{u-1}$ and a linearly decreasing sequence $2(m + 1 - j) / (m(m + 1))$ for $j = 1, \dots, m$. Using decreasing weights reflects a belief that, if the null of equal weights

is false, the minimum MSFE aggregation weights will give more weight to more recent observations.

Finally, σ_u^2 and ϵ , must be specified for V_T^* . The error term in (8) is $e_{t+1}^* = e_{t+1} + T^\epsilon u_{t+1}$ with $0 < \epsilon < 1/2$ and with $u_{t+1} \sim \text{iidN}(0, \sigma_u^2)$ and independent of $(\xi_{t+1-i/m})$. I suggest setting $\sigma_u^2 = T^{-1} \sum \hat{e}_{t+1}^2 \approx \mathbf{var}(e_{t+1})$ as a rule of thumb, where (\hat{e}_{t+1}) are the fitted residuals from the test regression in (5) for the unmodified statistic. $T^\epsilon u_{t+1}$ is thus large enough to dominate the finite sample distribution under the null when $\epsilon > 0$, but not so large that $T^\epsilon u_{t+1}$ reduces finite sample power under the alternative when $\epsilon = 0$ – i.e., when $T^\epsilon u_{t+1}$ is just random noise.

The choice of ϵ is more delicate. If chosen too small, correct size cannot be expected in the I(1) case. If chosen too large, too much power may be lost. In the simulations, I consider $\sigma_u^2 = 0$ (unmodified VAT statistic) and $\epsilon = \{0.25, 0.33, 0.45\}$ with σ_u^2 chosen as above (modified VAT statistic). Both modified and unmodified VAT statistics are implemented as q multiplied by an F -test statistic of the null that all but the first element of φ are zero, since $(\xi_{t+1-i/m})$ has neither serial correlation nor heteroskedasticity.

Table 1 shows the rejection rates – i.e., the empirical power function – for the VAT and modified VAT statistics with a 0.05 nominal size based on 1,000 simulations of Model 1 with I(0) data. Note that rejection rates yield empirical sizes under the null ($c/T = 0$) and empirical power under the alternatives.

Looking at the small frequency ratio of $m = 4$, all test statistics appear to be properly sized. The nuisance parameter ρ causes almost no difference in the empirical sizes using either the unmodified or modified VAT statistics, but $\rho \neq 0$ seems to increase power against local alternatives slightly. There is no major difference between qU and $q2$, because $q = 3$ in the former and $q = 2$ in the latter. V_T is very powerful against every alternative considered. As expected, V_T^* reduces power as ϵ increases. Even in the cases of $\epsilon = \{0.33, 0.45\}$, these statistics are quite powerful, but numbers shaded gray are empirical powers that do not exceed 0.90. To put this into perspective, $c/T = 0.2$ corresponds to a weight vector of $(0.32, 0.27, 0.22, 0.18)'$ – nearly equal in comparison to the null of equal weights. It seems difficult to go wrong with any of the test statistics when $m = 4$ and the data are known to be I(0).

Extending the frequency ratio to $m = 150$ causes substantial over-fitting using qU in the test regression, with $q = 149$ degrees of freedom and a LF sample size of $T = 200$. Rejection rates over 20% under the null reflect this over-fitting. When $m = 365$, qU is obviously not feasible, since the test regression would have more regressors than observations. The results for $q2$ are similar to those with $m = 150$.

The power functions for V_T^* and $m = \{150, 365\}$ exhibit a non-monotonicity in θ in the

Model 1	c/T=	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0			
$m = 4$	$\rho = 0$	$\sigma_u = 0$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
		$\varepsilon = 0.25$	0.06	0.81	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.33$	0.07	0.45	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.45$	0.06	0.16	0.58	0.97	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99
		$\sigma_u = 0$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.05	0.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	$\rho = 1/2$	$\varepsilon = 0.33$	0.08	0.49	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.45$	0.06	0.18	0.66	0.97	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	
		$\sigma_u = 0$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.06	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.33$	0.06	0.62	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.45$	0.07	0.23	0.77	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.00	
$m = 150$	$\rho = 0$	$\sigma_u = 0$	0.21	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
		$\varepsilon = 0.25$	0.24	0.33	0.76	0.99	1.00	1.00	1.00	1.00	1.00	0.99	0.84	0.64	0.52	0.45	0.43	0.42	0.41	0.38	0.41	0.39	0.39	0.39	
		$\varepsilon = 0.33$	0.21	0.27	0.48	0.76	0.90	0.92	0.92	0.99	1.00	1.00	0.83	0.51	0.42	0.34	0.34	0.29	0.30	0.29	0.30	0.30	0.29	0.29	
		$\varepsilon = 0.45$	0.21	0.25	0.29	0.39	0.49	0.46	0.50	0.66	0.98	0.85	0.42	0.33	0.28	0.28	0.27	0.26	0.24	0.26	0.26	0.24	0.24	0.25	
		$\sigma_u = 0$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.05	0.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.81	0.64	0.49	0.43	0.38	0.37	0.38	0.38	
	$\rho = 1/2$	$\varepsilon = 0.33$	0.06	0.48	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.66	0.47	0.33	0.26	0.22	0.22	0.20	0.19	0.19	
		$\varepsilon = 0.45$	0.06	0.19	0.69	0.97	1.00	0.99	1.00	1.00	1.00	1.00	0.97	0.64	0.36	0.21	0.15	0.13	0.12	0.10	0.08	0.08	0.08	0.08	
		$\sigma_u = 0$	0.24	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.24	0.40	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.85	0.66	0.55	0.47	0.44	0.42	0.39	0.39	0.41	0.40	0.40	
		$\varepsilon = 0.33$	0.22	0.28	0.62	0.89	0.95	0.95	0.94	1.00	1.00	1.00	0.86	0.52	0.42	0.34	0.35	0.29	0.31	0.32	0.29	0.29	0.29	0.29	
		$\varepsilon = 0.45$	0.24	0.26	0.32	0.47	0.51	0.48	0.52	0.69	1.00	0.91	0.43	0.32	0.27	0.29	0.27	0.26	0.25	0.23	0.26	0.26	0.25	0.25	
$m = 365$	$\rho = 0$	$\sigma_u = 0$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
		$\varepsilon = 0.25$	0.07	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.86	0.69	0.52	0.45	0.39	0.39	
		$\varepsilon = 0.33$	0.06	0.49	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.73	0.53	0.36	0.28	0.21	0.18	0.18	
		$\varepsilon = 0.45$	0.06	0.18	0.66	0.96	0.95	0.87	0.78	0.73	0.85	0.99	1.00	0.99	0.74	0.43	0.26	0.19	0.13	0.11	0.09	0.09	0.09	0.09	
		$\sigma_u = 0$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.06	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.87	0.71	0.52	0.46	0.39	0.39	
	$\rho = 1/2$	$\varepsilon = 0.33$	0.05	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.74	0.53	0.38	0.29	0.20	0.19	0.19	
		$\varepsilon = 0.45$	0.06	0.24	0.84	0.99	0.97	0.89	0.80	0.76	0.87	0.99	1.00	1.00	1.00	0.75	0.44	0.24	0.21	0.13	0.11	0.10	0.09	0.09	

Table 1: **Rejection Rates for Model 1: I(0) Case.** Rejection rates with a nominal size of 0.05 from 1,000 simulations with $d = 0$, $\gamma = 0$, $m = \{4, 150, 365\}$, $T = 200$, $\rho \in \{0, 1/2\}$, $\sigma_u^2 = 0$, and $\sigma_u^2 = T^{-1} \sum \hat{\varepsilon}_{t+1}^2$ with $\varepsilon \in \{0.25, 0.33, 0.45\}$. qU and $q2$ correspond to unrestricted ($q = m - 1$) and restricted ($q = 2$) instrumentation schemes, respectively. (In the font used for the table $\varepsilon = \varepsilon$.) Bold numbers correspond to rejection rates exceeding 0.10 under the null. Shaded numbers correspond to rejection rates not exceeding 0.90 under the alternative.

sense that power seems to *decrease* for alternatives further away from the null beyond a certain point. The theory suggests that this is a small-sample effect, and, in fact, increasing T in the simulations (not shown) increases the low rejection rates for higher values of θ . Nevertheless, V_T^* does not work well when the data are all $I(0)$ and $\epsilon > 1/4$.

Rejection rates using Model 2 with $I(1)$ data are displayed in Table 2. As before, there is little difference between qU and $q2$ when $m = 4$. Note also that when $\rho = 0$ and $m = 4$, the size and power results are strikingly similar to the $I(0)$ case of Model 1. The main difference appears when $\rho \neq 0$, where the unmodified tests V_T have 100% rejection rates under the null! Such over-rejection in general inference about forecasting models with possible unit roots due to nuisance parameters is well-established (Elliott, 2006). V_T^* pushes the size down below 0.10, but not until $\epsilon = 0.45$. Power against almost all alternatives is maintained at 1.

Moving to $m = 150$ for Model 2, the effects of over-fitting again using qU are clear. The apparent size distortion using V_T with $\rho \neq 0$, is similar to when $m = 4$. Even with $\epsilon = 0.45$, the empirical size is still (slightly) above 0.10, but there is little or no power loss from increasing ϵ for $q2$, in contrast to the $I(0)$ case of Model 1. This increase in power due to the presence of a unit root is suggested by Theorem 2. The results for $q2$ with $m = 365$ are qualitatively similar to those with $m = 150$.

Finally, the rejection rates using Model 3, with a lagged target as a (cointegrated) predictor, are shown in Table 3. The results for Model 3 are qualitatively very similar to those for Model 2, but with one notable exception: when $\rho \neq 0$, the modified statistic V_T^* successfully pushes the size below 0.10, once ϵ increases to 0.45. Cointegrating the regressors by including y_t – which is natural in a forecasting model – seems to improve the size of the test.

Synthesizing the results of the simulations provides the following insights. If unit roots *or* nuisance parameters can be ruled out safely, the unmodified VAT statistic V_T , which was proposed by Miller (2014) for a MIDAS null, can also be used for nulls of a known aggregation schemes. The statistic is also similar to that proposed by Andreou *et al.* (2010) for the null of a flat aggregation scheme, but with the critical difference of using a small number of instruments given by linearly independent combinations of the HF predictors rather than the HF predictors themselves.

Otherwise, if neither unit roots nor nuisance parameters such as ρ can be ruled out, the modified VAT statistic V_T^* should be used. A choice of $\epsilon = 0.45$ reduces size distortion effectively in the presence of such nuisance parameters, but at the expense of power when m is relatively large and the predictor happens to be $I(0)$. Power against both fixed and local alternatives is excellent when m is relatively small or when the predictor happens to be $I(1)$.

Model 2	c/T =	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0			
$m = 4$	$\rho = 0$	$\sigma_u = 0$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
		$\varepsilon = 0.25$	0.05	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.33$	0.04	0.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.45$	0.05	0.24	0.84	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\sigma_u = 0$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.25$	0.05	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\rho = 1/2$	$\varepsilon = 0.33$	0.04	0.76	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.45$	0.05	0.27	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\sigma_u = 0$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.25$	0.28	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.33$	0.14	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.45$	0.08	0.52	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$m = 150$	$\rho = 0$	$\sigma_u = 0$	0.21	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.26	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.33$	0.24	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.96	0.95	0.93	0.92	0.92	0.91	
		$\varepsilon = 0.45$	0.24	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.78	0.67	0.59	0.55	0.52	0.51	0.50	0.52	0.49		
		$\sigma_u = 0$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.25$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\rho = 1/2$	$\varepsilon = 0.33$	0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.45$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	
		$\sigma_u = 0$	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.25$	0.27	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.33$	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.96	0.93	0.92	0.91	0.91	0.90	
		$\varepsilon = 0.45$	0.26	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.79	0.64	0.58	0.53	0.48	0.50	0.49	0.49	0.48		
$m = 365$	$\rho = 0$	$\sigma_u = 0$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.37	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.33$	0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.45$	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	
		$\sigma_u = 0$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.25$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$\rho = 1/2$	$\rho = 0$	$\varepsilon = 0.33$	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.45$	0.04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\sigma_u = 0$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.37	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.33$	0.19	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.45$	0.10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2: **Rejection Rates for Model 2: I(1) Case.** Rejection rates with a nominal size of 0.05 from 1,000 simulations with $d = 0$, $\gamma = 0$, $m = \{4, 150, 365\}$, $T = 200$, $\rho \in \{0, 1/2\}$, $\sigma_u^2 = 0$, and $\sigma_u^2 = T^{-1} \sum \hat{e}_{t+1}^2$ with $\varepsilon \in \{0.25, 0.33, 0.45\}$. qU and $q2$ correspond to unrestricted ($q = m - 1$) and restricted ($q = 2$) instrumentation schemes, respectively. (In the font used for the table $\varepsilon = \varepsilon$.) Bold numbers correspond to rejection rates exceeding 0.10 under the null. Shaded numbers correspond to rejection rates not exceeding 0.90 under the alternative.

Model 3	c/T=	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0				
$m = 4$	$\rho = 0$	$\sigma_u = 0$	0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00			
		$\varepsilon = 0.25$	0.06	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
		$\varepsilon = 0.33$	0.04	0.53	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.45$	0.05	0.18	0.67	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\sigma_u = 0$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.06	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	$\rho = 1/2$	$\sigma_u = 0$	0.05	0.61	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.05	0.22	0.74	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.33$	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.45$	0.14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\sigma_u = 0$	0.08	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		$\varepsilon = 0.25$	0.07	0.35	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$m = 150$	$\rho = 0$	$\sigma_u = 0$	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
		$\varepsilon = 0.25$	0.22	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.99	0.99	
		$\varepsilon = 0.33$	0.23	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.97	0.91	0.88	0.87	0.84	0.83	0.82	0.81	0.81	
		$\varepsilon = 0.45$	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.84	0.65	0.55	0.49	0.50	0.45	0.43	0.42	0.42	0.42	0.43	0.43	
		$\sigma_u = 0$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
	$\rho = 1/2$	$\sigma_u = 0$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.98	0.98	0.98	0.98	0.98	0.97	
		$\varepsilon = 0.33$	0.66	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.45$	0.25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	
		$\sigma_u = 0$	0.23	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.96	0.92	0.90	0.85	0.82	0.83	0.84	0.80	0.80	
		$\varepsilon = 0.25$	0.22	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.84	0.65	0.54	0.52	0.50	0.43	0.43	0.41	0.43	0.41	0.41	0.41	
$m = 365$	$\rho = 0$	$\sigma_u = 0$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
		$\varepsilon = 0.25$	0.26	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.33$	0.13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.45$	0.07	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.98	0.97	0.97	0.98	0.97	
		$\sigma_u = 0$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
		$\varepsilon = 0.25$	0.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

Table 3: **Rejection Rates for Model 2: I(1) Case with Lagged Target.** Rejection rates with a nominal size of 0.05 from 1,000 simulations with $d = 0$, $\gamma = 0$, $m = \{4, 150, 365\}$, $T = 200$, $\rho \in \{0, 1/2\}$, $\sigma_u^2 = 0$, and $\sigma_u^2 = T^{-1} \sum \hat{\varepsilon}_{t+1}^2$ with $\varepsilon \in \{0.25, 0.33, 0.45\}$. qU and $q2$ correspond to unrestricted ($q = m - 1$) and restricted ($q = 2$) instrumentation schemes, respectively. (In the font used for the table $\varepsilon = \epsilon$.) Bold numbers correspond to rejection rates exceeding 0.10 under the null. Shaded numbers correspond to rejection rates not exceeding 0.90 under the alternative.

5 Application to State Retail Gasoline Prices

Suppose a researcher is interested in forecasting state-level retail gasoline prices. The U.S. has a federal gasoline tax rate of \$0.184/gallon, but state-levied gasoline tax rates vary widely. California, New York, and Pennsylvania have particularly high tax rates, while Missouri, Oklahoma, and Texas have particularly low tax rates, for example. A state policy maker evaluating a change in the gasoline tax rate would be particularly interested in a state-level price forecast. In fact, law makers in Missouri are considering an increase of \$0.015/gallon from \$0.17/gallon at the time of this writing.³

Data. Like the tax rates themselves, the ease in obtaining readily usable state-level data on retail gasoline prices also varies by state. The US Energy Information Administration (EIA) publishes weekly data for 9 states (California, Colorado, Florida, Massachusetts, Minnesota, New York, Ohio, Texas, and Washington), 10 cities (Boston, Chicago, Cleveland, Denver, Houston, Los Angeles, Miami, New York, San Francisco, and Seattle) and 5 regions (East Coast, Midwest, Gulf Coast, Rocky Mountain, and West Coast). Missouri is included in the Midwest region, but is not one of the 9 states for which data are published by the EIA. Missouri's largest city, St. Louis, is not one of the 10 cities, and the closest of these cities, Chicago, is hours away.

The Missouri Department of Economic Development publishes a state-level retail price in the Missouri Energy Bulletin⁴ roughly twice per month, but these bulletins are publicly available only since 2011. The University of Missouri's Economic and Policy Analysis Research Center maintains monthly historical data since 1999.⁵ I apply the tests developed in this paper to forecasting and nowcasting models of this monthly series from 1999.1-2014.12.

A natural predictor of the Missouri retail price is the aforementioned weekly series of Midwest retail gasoline prices,⁶ which includes Missouri, as well as Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Nebraska, North Dakota, South Dakota, Ohio, Oklahoma, Tennessee, and Wisconsin. Another publicly available series from the EIA that might predict Missouri retail gasoline prices is the Gulf Coast daily spot price series.⁷ Gulf

³http://www.stltoday.com/news/senate-panel-finds-support-for-gas-tax-hike/article_c6e95ff2-d91b-55c1-8ed4-be9548c6af77.html, St. Louis Post-Dispatch, accessed January 13, 2016.

⁴<https://energy.mo.gov/energy/stay-informed/missouri-energy-bulletins>, Regular Unleaded Gasoline - Retail Prices, Missouri, accessed January 11, 2016.

⁵<http://eparc.missouri.edu/publications/archive/files/f/f15.pdf>, Missouri Transportation Fuels Average Retail Price by Month, accessed January 11, 2016.

⁶http://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_w.htm, Weekly Midwest Regular Conventional Retail Gasoline Prices, accessed January 11, 2016.

⁷http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm, U.S. Gulf Coast Conventional Gasoline Regular Spot Price FOB, accessed January 11, 2016.

Coast spot prices may influence future retail prices over a region much wider than Missouri, but are available at a much higher frequency.

High-frequency and high-resolution data – i.e., daily state and local data – are available from other sources, such as the American Automobile Association’s Fuel Gauge Report,⁸ but they publish only limited historical data, as of this writing.

Models. I consider a forecasting model (Model F) and a nowcasting model (Model N) given by

$$\begin{aligned} \text{Model F: } y_{t+1} &= \gamma_0 + y_t \gamma_1 + x_{1t}^{1:m_t'} \alpha_{1t} + x_{2t}^{1:n_t'} \alpha_{2t} + \varepsilon_{t+1}^F \\ \text{Model N: } y_{t+1} &= \gamma_0 + y_t \gamma_1 + x_{1,t+1}^{1:m_t'} \alpha_{1t} + x_{2,t+1}^{1:n_t'} \alpha_{2t} + \varepsilon_{t+1}^N \end{aligned}$$

where $x_{1t}^{1:m_t'}$ denotes a monthly vector of $m_t = 4\text{-}5$ weeks of Midwest retail prices and $x_{2t}^{1:n_t'}$ denotes a monthly vector of $n_t = 18\text{-}23$ business days of Gulf Coast spot prices. Due to calendar irregularities, the coefficients α_{1t} and α_{2t} are also calendar dependent, but only through the aggregation weight vectors. That is, $\alpha_{1t} = \pi_{1t}(\theta_1)\beta_1$ and $\alpha_{2t} = \pi_{2t}(\theta_2)\beta_2$ where β_1 , β_2 , θ_1 and θ_2 are assumed to be invariant, so that the functions π_{1t} and π_{2t} vary solely due to the length of month t .

In Model F, the HF predictors contain all current information other than that in y_t relevant to future y_{t+1} , while nowcasting with Model N uses contemporaneous information. Substantive differences in the weights between these two models are expected, because the forecasting model should give more weight to more recent observations, while the nowcasting model should weigh daily and weekly observations proportionally to their influence on the contemporaneous month’s state price.

Because the target and predictors are contemporaneous in Model N, nuisance parameters from substantial correlation between the predictors and error are expected. Also, because prices are typically viewed to be I(1), one should proceed with caution when using V_T to test model N – simulations in the previous section suggest size distortion in this case.

Null and Alternative Hypotheses. I consider three options for handling the different frequencies of the data: aggregate the HF predictors similarly to (1) (null of the tests), estimate an unrestricted DL model similar to (2) with no restrictions on the weights (an alternative of the tests), or use a parsimonious MIDAS specification (another alternative).

Aggregation sets $x_{1t}^{m_t} = x_{1t}^{1:m_t'} \pi_{m_t}$ and $x_{2t}^{n_t} = x_{2t}^{1:n_t'} \pi_{n_t}$, where π_{m_t} and π_{n_t} are known. Specifically, I consider flat sampling (equal weights for each week or day), end-of-period (EOP) sampling (unit weight for the last week or day only), and beginning-of-period (BOP) sampling (unit weight for the first week or day only). Calendar irregularities cause no

⁸<http://www.fuelgaugereport.com>

problems for aggregation, other than that the information may not be aligned in the sense that the last week or day of each month is not the same from month to month.

Calendar irregularities cause rather obvious problems for the second option, an “unrestricted” DL model. A typical approach is to consider only $m = \min(m_t) = 4$ weeks per month and $n = \min(n_t) = 18$ days per month, losing $\sum_t m_t - mT$ and $\sum_t n_t - nT$ HF observations. For model F, I consider both a *beginning-weighted* scheme, in which

$$x_{1t}^{1:m_t'} \alpha_{1t} + x_{2t}^{1:n_t'} \alpha_{2t} + \varepsilon_{t+1} = x_{1t}^{1:m'} \alpha_1^{bw} + x_{2t}^{1:n'} \alpha_2^{bw} + \varepsilon_{t+1}^{bw}$$

uses only the *first* 4 weeks and *first* 18 days, and an *end-weighted* scheme, in which

$$x_{1t}^{1:m_t'} \alpha_{1t} + x_{2t}^{1:n_t'} \alpha_{2t} + \varepsilon_{t+1} = x_{1t}^{(m_t-m+1):m_t'} \alpha_1^{ew} + x_{2t}^{(n_t-n+1):n_t'} \alpha_2^{ew} + \varepsilon_{t+1}^{ew}$$

uses only the *last* 4 weeks and *last* 18 days. The respective error terms simply augment ε_{t+1}^F in Model F with the omitted HF variables. The same schemes are used for Model N with obvious notational differences. Note that although these schemes are less restrictive than full aggregation, they are not really unrestricted.

Parsimonious MIDAS specifications, the third option, impose restrictions by reducing the dimension of θ in $\pi(\theta)$, but these nonlinear specifications are well-equipped to deal with calendar irregularities. I consider a first-order exponential Almon lag given by

$$\begin{aligned} \pi_{i,Almon}(\theta_1) &= \exp(\theta_1 i) / \sum_{u=1}^{m_t} \exp(\theta_1 u) \\ \pi_{j,Almon}(\theta_2) &= \exp(\theta_2 j) / \sum_{u=1}^{n_t} \exp(\theta_2 u) \end{aligned}$$

for $i = 1, \dots, m_t$ and $j = 1, \dots, n_t$.

Although these functions use all of the HF observations, the issue of aligning those observations remains. The above MIDAS specifications are *end-aligned*, in the sense that the last weeks or days of each month have the same index – π_1 corresponds to the last rather than first HF period. *Beginning-aligned* MIDAS specifications are given by

$$\begin{aligned} \pi_{i,Almon}(\theta_1) &= \exp(\theta_1 i) / \sum_{u=m_*-m_t+1}^{m_*} \exp(\theta_1 u) \\ \pi_{j,Almon}(\theta_2) &= \exp(\theta_2 j) / \sum_{u=n_*-n_t+1}^{n_*} \exp(\theta_2 u) \end{aligned}$$

for $i = m_* - m_t + 1, \dots, m_*$ and $j = n_* - n_t + 1, \dots, n_*$, where $m_* = \max(m_t) = 5$ and $n_* = \max(n_t) = 23$.

Note that, although a MIDAS specification is nested by an unrestricted DL specification in the absence of calendar effects, these MIDAS specifications with calendar irregularities

employ all of the data and are therefore not nested by the distributed lag specifications that do not.

Extending the tests themselves to allow for calendar effects is not problematic, because the null may be framed in terms of θ for a given function π . A null of flat sampling corresponds to $\theta_{1,2} = 0$ in the above MIDAS functions, an EOP null corresponds to $\theta_{1,2} \rightarrow -\infty$, while a BOP null corresponds to $\theta_{1,2} \rightarrow \infty$. As long as $q < \min(m_t) - 1 = 3$ for a test on the weekly data and $q < \min(n_t) - 1 = 17$ for a test on the daily data, the χ_q^2 limiting distribution may be expected to hold. In the case of a *joint* test of θ_1 and θ_2 , which may be conducted simply by imposing both constraints to get fitted residuals under the null, $q < \min(m_t, n_t) - 1 = 3$, but a χ_{2q}^2 limiting distribution is expected, since there are $2q$ instruments added in the test regression.

Sequential Testing Procedure. I consider the following sequential testing procedure:

1. Use V_T or V_T^* to test aggregation nulls of interest, such as flat sampling, EOP sampling, or BOP sampling.
 - (a) Fail to Reject: Use that sampling scheme, stop
 - (b) Reject:
 - i. Consider a different sampling scheme, repeat step 1
 - ii. Or, consider a MIDAS specification, go to step 2
2. Use V_T to test the MIDAS specification against a more general alternative
 - (a) Fail to Reject: Use the MIDAS specification, stop
 - (b) Reject:
 - i. Consider a different MIDAS specification, repeat step 2
 - ii. Or, if feasible, use a distributed lag specification, stop

The two-step implementation of the test is the same for the MIDAS null in step 2 as for a fixed null, except nonlinear least squares is used to estimate the model in (2) with a MIDAS parameterization of $\alpha(\theta)$ in the first step of the test. Miller (2014) derived the limiting distribution of V_T under this null. No modification is necessary to preserve size under this null.

MODEL F	RMSE	V_T	V_T^* with $\epsilon =$		
			0.25	0.33	0.45
Flat sampling	0.128	13.23	18.01	12.28	7.93
EOP sampling	0.089	0.38	6.20	5.63	5.10
BOP sampling	0.165	31.30	27.04	17.14	9.97
End-aligned MIDAS	0.089	0.38	6.23	5.65	5.11
Beginning-aligned MIDAS	0.089	0.38	6.22	5.64	5.10
End-weighted distributed lag	0.073				
Beginning-weighted distributed lag	0.090				

Table 4: **RMSE's and Test Statistics for Model F.** Bold numbers correspond to rejections using a size-0.05 χ_4^2 critical value of 9.49.

Numerical Results. I set $q = 2$, using instruments generated according to those in the simulations of the previous section but using the appropriate number of days and weeks in each month. I conduct joint size-0.05 tests of the null $\theta_1 = \theta_2 = \theta_0$ with $2q = 4$ instruments, so that a χ_4^2 critical value of 9.49 is applicable.

Tables 4 and 5 show fitted root mean squared errors (RMSE's) and the test statistics for five nulls – three fixed nulls: flat sampling, EOP sampling, BOP sampling, and two MIDAS nulls: end-aligned and beginning-aligned – for Models F and N respectively. V_T^* is calculated as $2q$ times an F -test statistic of the joint null, but V_T is calculated using a robust covariance estimator.⁹

For the forecasting Model F, BOP sampling is firmly ruled out. Flat sampling is rejected except when $\epsilon = 0.45$. The decrease in V_T^* as ϵ increases could be due to decreasing size, if prices are I(1), or decreasing power, if prices are I(0). In any case, the evidence most strongly supports EOP sampling among the three fixed nulls.

According to the sequential testing procedure, one should choose EOP sampling. For the purpose of illustration, however, it is informative to consider tests of the MIDAS nulls. Neither of these are rejected against a less restrictive alternative. Moreover, the RMSE's using the MIDAS models are identical to that using EOP sampling, and slightly better than that using a distributed lag model that equalizes weeks and days by dropping the last ones.

Figure 1 sheds some light on the similarity between the results using EOP sampling and using MIDAS. The figure shows the estimated weekly and daily beginning-aligned MIDAS weights, where MIDAS(4) and MIDAS(5) show the weights estimated for months with 4 and 5 weeks, respectively, and MIDAS(18)-MIDAS(23) show the weights estimated for months

⁹I use a rectangular window with a lag truncation of 1, but the numerical results in this application are quite robust to this choice of lag truncation.

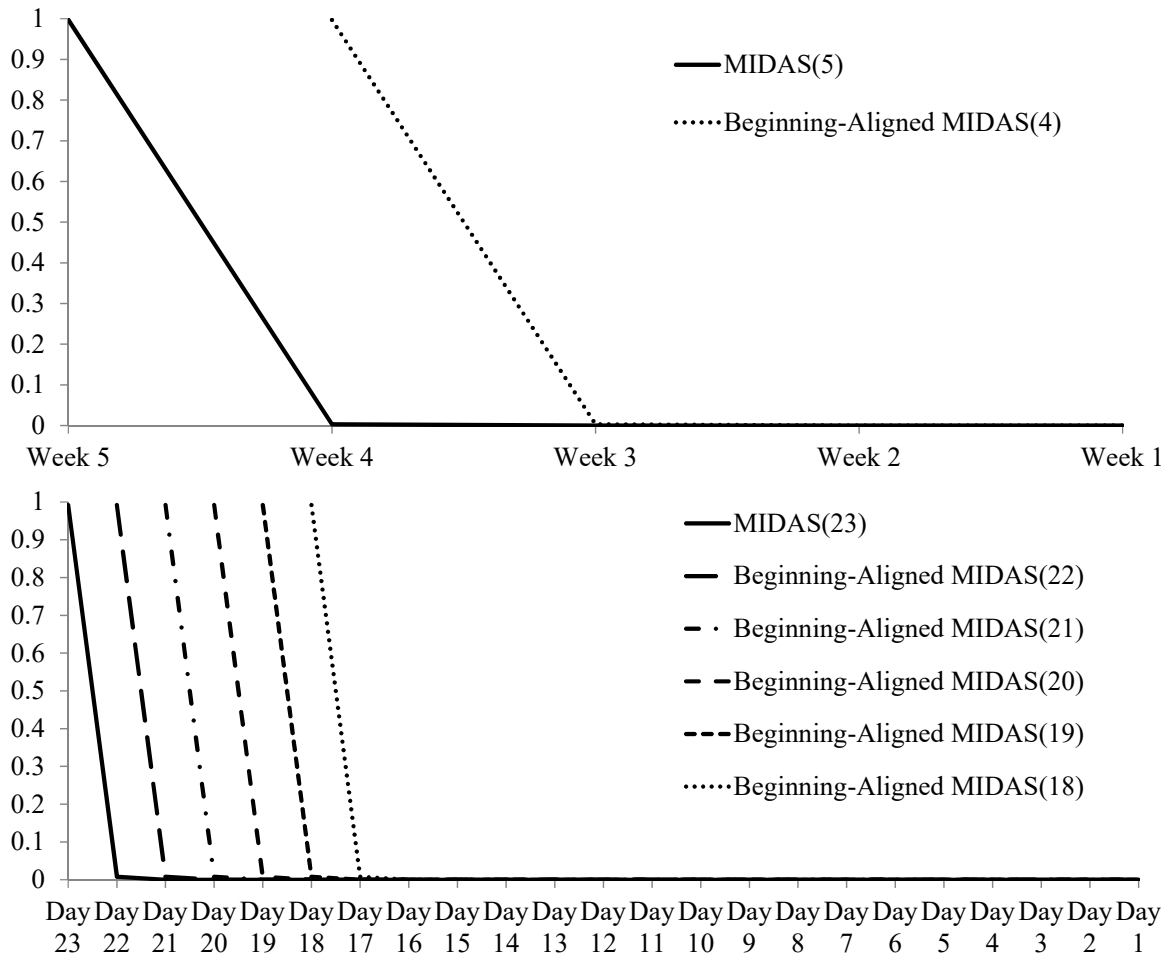


Figure 1: **Weights Estimated for Model F.** Weekly weights (top panel) and daily weights (bottom panel) for beginning-aligned MIDAS specifications.

MODEL N	RMSE	V_T	V_T^* with $\epsilon =$		
			0.25	0.33	0.45
Flat sampling	0.042	5.43	4.03	2.73	2.00
EOP sampling	0.077	34.41	9.68	5.13	2.51
BOP sampling	0.069	48.53	33.72	18.31	8.53
End-aligned MIDAS	0.077	37.04	10.35	5.46	2.64
Beginning-aligned MIDAS	0.037	0.64	1.39	1.39	1.41
End-weighted distributed lag	0.039				
Beginning-weighted distributed lag	0.039				

Table 5: **RMSE's and Test Statistics for Model N.** Bold numbers correspond to rejections using a size-0.05 χ_4^2 critical value of 9.49.

with the respective number of days. Both weekly and daily weights are estimated to be indistinguishable from EOP sampling. The same holds for end-aligned MIDAS, the weights of which are essentially the same as MIDAS(5) and MIDAS(23) in the figure.

Overall, the evidence supports using EOP sampling for Model F. Keeping in mind that this is a forecasting model, the result is sensible as it suggests that the most recent prices – measured at any frequency – have the most predictive power for next period's target.

Moving to the nowcasting Model N, V_T and V_T^* with $\epsilon = 0.25$ reject BOP sampling, EOP sampling, and end-aligned MIDAS. Flat sampling and beginning-aligned MIDAS are not rejected with any of the tests. Beginning-aligned MIDAS shows the lowest RMSE.

The sequential testing procedure suggests using flat sampling for Model N, but Figure 2, showing the weights estimated using beginning-aligned MIDAS, tells a different story. The weekly weights are indeed close to flat, but with more emphasis on the earlier weeks. The daily weights for months with 23 business days show a similar decreasing pattern, but they are closer to BOP sampling than flat sampling.

Since Model N is a nowcasting model, the interpretation of the non-zero weights are those at weeks or days when Midwest retail and Gulf Coast spot prices best explain the *contemporaneous* Missouri retail price. Recall that the Missouri data are collected roughly twice per month – typically during the first and third week or second and fourth week of each month, but never during the fifth week. In this light, it is sensible to give more weight to earlier weeks in each month, as beginning-aligned MIDAS does. In contrast, end-aligned MIDAS would misalign fifth weeks, which should all receive zero weight, with some fourth weeks that should receive a positive weight. Beginning-aligned MIDAS with a higher-order exponential Almon lag specification might capture the bimonthly nature of the data better than the first-order lag considered here.

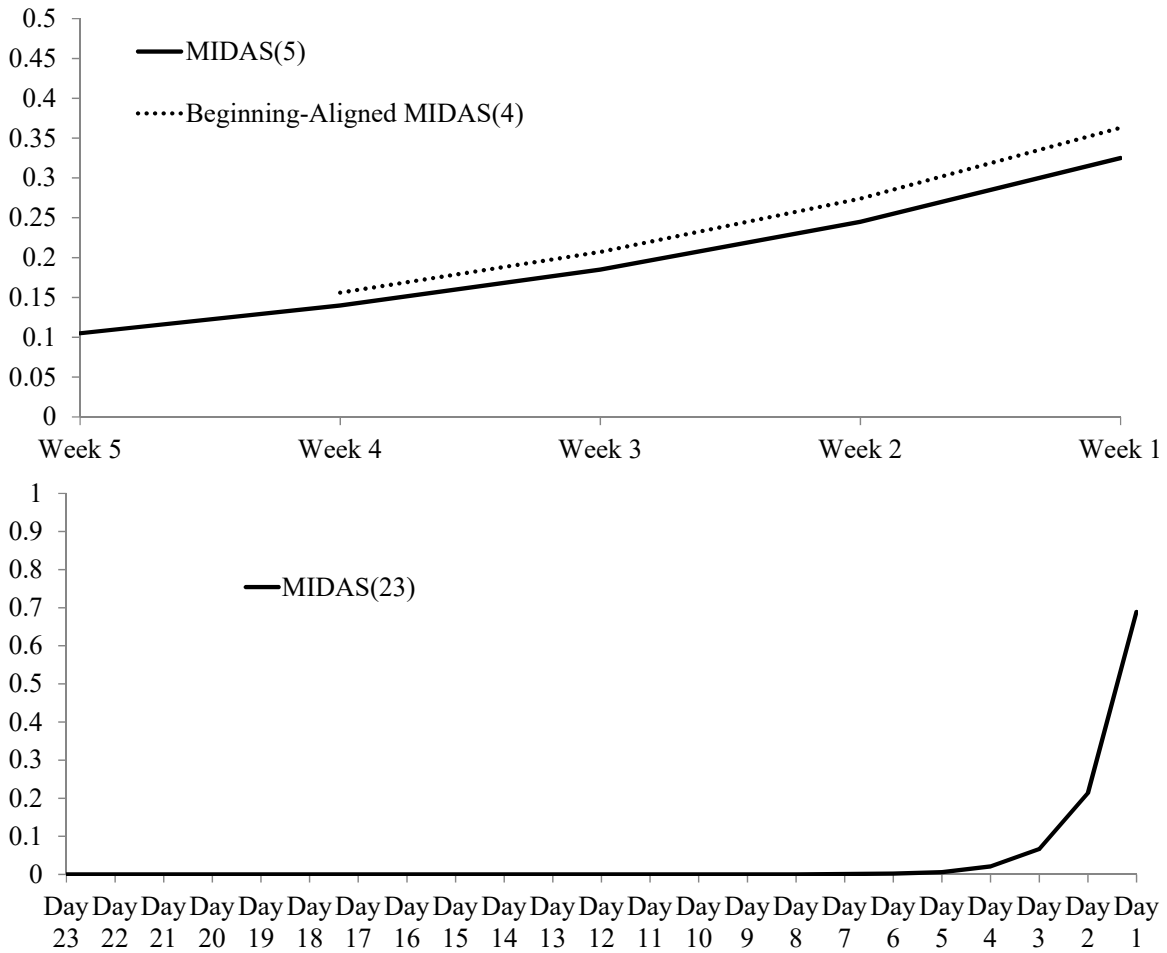


Figure 2: **Weights Estimated for Model N.** Weekly weights (top panel) and daily weights (bottom panel) for beginning-aligned MIDAS specifications.

6 Some Conclusions

The two VAT statistics, unmodified V_T and modified V_T^* , are very easy to compute, and they are robust to a wide range of conditions that have the potential to derail inference using other tests. In particular, I have shown their robustness to deterministically and stochastically trending covariates and to infeasibility of the alternative hypothesis when m is large relative to the LF sample size. I emphasize the desirability of choosing a small number of linear combinations of the HF predictor as instruments rather than using the HF predictor vector itself.

V_T^* is also robust to nuisance parameters that may plague inference when the HF predictor on which the tests are conducted is stochastically trending. Because of this robustness,

knowledge about the order of integration and cointegrating relationships is not needed for V_T^* .

The robustness of V_T^* naturally comes at a price of power loss. Simulations suggest that the power loss may be negligible or substantial. It is substantial when m is relatively large and the HF predictor is stationary, suggesting that V_T should be used if these are known to be the case. Otherwise, if m is not very big, the power loss is minor even against local alternatives, suggesting that V_T^* should be used to guard against over-rejection from nuisance parameters.

If m is sizeable, but the order of integration is unknown, the user faces some risk of over-rejection under the null with V_T or under-rejection under the alternative with V_T^* . In this case, it might be desirable to pretest for a unit root. Of course, there may be a compelling reason to do so for the purpose of forecasting – even without the MF issue considered here.

Appendix A: Proofs of the Theoretical Results

For the reader's convenience, seventeen preliminary results are presented in Lemma A1. Lemma A2 gives the minimum MSFE parameters vectors α_m for the I(0) and I(1) cases. Proofs of the main results follow these lemmas.

First, I introduce some notation used throughout the proofs. Let $=_a$ denote an asymptotic approximation when the right-hand side contains a mixture of a limit and expressions of asymptotic order. **MN** denotes a mixed normal distribution – i.e., normal with stochastic variance.

Define the diagonal normalization matrix ϖ_T with the first set of diagonal elements of T corresponding to the dimension of w_{0t} , the second set of diagonal elements of T^2 corresponding to the dimension of w_{1t} , and the last set of diagonal elements $T^{1+2\delta_j}$ corresponding to each of the deterministic trends w_{2jt} in w_{2t} . Let v_T be a $(q+1) \times (q+1)$ diagonal normalization matrix with T in the first q diagonals and T^2 in the last one.

I use the typical notation of $\Omega = \sum_{k=-\infty}^{\infty} \mathbf{E}[\xi_t \xi_{t-k}']$ for the long-run variance of (ξ_t) , the same $\Omega = \lim_{T \rightarrow \infty} \mathbf{var}(T^{-1/2} \sum \xi_t)$ as in Assumption 1, and $\Delta = \sum_{k=0}^{\infty} \mathbf{E}[\xi_t \xi_{t-k}']$ for the one-sided long-run variance of (ξ_t) , and I employ subscripts to denote partitions of these identical to those of Σ in Section 3.1 for the I(0) case and in Section 3.2 for the I(1) case. The Brownian motion B_z is the subvector of Brownian motion B corresponding to Δx_{1t} in ξ_t in the limit in Assumption 1(c).

Finally, in Lemma A1, let π and $\tilde{\pi}$ denote generic aggregation vectors – non-negative elements with a unit sum. These may simply be replaced by H_1^x in the appropriate parts of the lemma.

Lemma A1.

- [a.i] $T\varpi_T^{-1/2}M_{ww}^T\varpi_T^{-1/2} =_a \begin{bmatrix} \Sigma_{00} & O_p(T^{-1/2}) \\ O_p(T^{-1/2}) & O_p(1) \end{bmatrix}$
- [b.i] $T^{1/2}\varpi_T^{-1/2}M_{w\varepsilon}^T =_a (\sigma'_{0\varepsilon}, O_p(T^{-1/2}))'$
- [b.ii] $T\varpi_T^{-1/2}M_{w\varepsilon}^T = (T^{-1/2}\sum \varepsilon_{t+1}w_{0t}, O_p(1))'$
- [c.i] $H'_0M_{x\varepsilon}^T \rightarrow_p \sigma_{x\varepsilon}$ and $T^{1/2}H'_0M_{xu}^T \rightarrow_d \mathbf{N}(0, \sigma_u^2\Sigma_{xx})$
- [c.ii] $\pi'M_{x\varepsilon}^T = O_p(1)$ and $\pi'M_{xu}^T \rightarrow_d \mathbf{MN}(0, \sigma_u^2\pi' \int B_z B'_z \pi)$
- [d.i] $H'_0M_{xw}^T\varpi_T^{-1/2}T^{1/2} =_a (\Sigma_{x0}, O_p(T^{-1/2}))'$
- [d.ii] $\pi'M_{xw}^T\varpi_T^{-1/2} = (O_p(T^{-1/2}), O_p(1))'$
- [e.i] $H'_0M_{xx}^T H_0 \rightarrow_p \Sigma_{xx}$
- [e.ii] $T^{-1}\pi'M_{xx}^T \tilde{\pi} \rightarrow_d \pi' \int B_z B'_z \tilde{\pi}$
- [e.iii] $\pi'M_{xx}^T H_0 = O_p(1)$
- [f.i] $H'_0M_{x\varepsilon\perp w}^T \rightarrow_p \sigma_{x\varepsilon\perp 0}$
- [f.ii] $\pi'M_{x\varepsilon\perp w}^T = O_p(1)$
- [g.i] $H'_0M_{xx\perp w}^T H_0 \rightarrow_p \Sigma_{xx\perp 0}$
- [g.ii] $T^{-1}\pi'M_{xx\perp w}^T \tilde{\pi} = O_p(1)$
- [g.iii] $\pi'M_{xx\perp w}^T H_0 = O_p(1)$
- [h.i] $T^{1/2}H'_0P_T \rightarrow_d \Omega_P^{1/2}\mathbf{N}(0, I_p)$ with $\Omega_P = (\sigma_\varepsilon^2 - \sigma_{\varepsilon x\perp 0}\Sigma_{xx\perp 0}^{-1}\sigma_{x\varepsilon\perp 0})\Sigma_{xx\perp 0}^{-1}$
- [h.ii] $\pi'P_T = O_p(1)$

Proof of Lemma A1. I omit proofs of parts [a.i]-[e.iii], which are straightforward from Assumption 1 and presented for easy reference in proofs of the subsequent parts. Parts [f.i] and [g.i] follow directly from parts [a.i], [b.i], and [c.i] and parts [a.i], [d.i], and [e.i] respectively.

The moment in part [f.ii] is $\pi' M_{x\varepsilon}^T - \pi' M_{xw}^T (M_{ww}^T)^{-1} M_{w\varepsilon}^T$, where the first term is $O_p(1)$ from part [c.ii]. The second term may be written as

$$\begin{aligned} \pi' M_{xw}^T (M_{ww}^T)^{-1} M_{w\varepsilon}^T &= T^{1/2} \left[\pi' M_{xw} \varpi_T^{-1/2} \right] \left[T \varpi_T^{-1/2} M_{ww}^T \varpi_T^{-1/2} \right]^{-1} \left[T^{1/2} \varpi_T^{-1/2} M_{w\varepsilon} \right] \\ &= T^{1/2} \begin{bmatrix} O_p(T^{-1/2}) \\ O_p(1) \end{bmatrix}' \begin{bmatrix} O_p(1) & O_p(T^{-1/2}) \\ O_p(T^{-1/2}) & O_p(1) \end{bmatrix} \begin{bmatrix} O_p(1) \\ O_p(T^{-1/2}) \end{bmatrix} \end{aligned}$$

from parts [a.i], [b.i], and [d.ii]. The result follows from matrix multiplication.

The proofs of parts [g.ii] and [g.iii] are similar to that of [f.ii], using parts [a.i], [d.ii], and [e.ii] and parts [a.i], [d.i], [d.ii], and [e.iii] respectively.

For parts [h.i] and [h.ii], note that $P_T = M_{x\varepsilon\perp w}^T - M_{xx\perp w}^T H_0^x \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0} + O_p(T^{-1})$ from the minimum MSFE parameter vector in Lemma A2[b]. Up to the negligible term, $T^{1/2} H_0' P_T$ in part [h.i] may be written as

$$T^{1/2} (H_0^{x'} M_{x\varepsilon}^T - H_0^{x'} M_{xx}^T H_0 \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0}) - T^{1/2} H_0^{x'} M_{xw}^T (M_{ww}^T)^{-1} (M_{w\varepsilon}^T - M_{wx}^T H_0^x \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0})$$

by expanding $M_{x\varepsilon\perp w}^T$ and $M_{xx\perp w}^T$ and rearranging terms. This expression may be rewritten as

$$\begin{aligned} &T^{-1/2} \sum x_{0t} (\varepsilon_{t+1} - x'_{0t} \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0}) - (H_0^{x'} M_{xw}^T \varpi_T^{-1/2} T^{1/2}) \\ &\times (T \varpi_T^{-1/2} M_{ww}^T \varpi_T^{-1/2})^{-1} \varpi_T^{-1/2} \sum w_t (\varepsilon_{t+1} - x'_{0t} \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0}) \end{aligned}$$

which is

$$T^{-1/2} \sum (x_{0t} - \Sigma_{x0} \Sigma_{00}^{-1} w_{0t}) (\varepsilon_{t+1} - x'_{0t} \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0}) + O_p(T^{-1/2})$$

from parts [a.i], [b.ii], and [d.i].

The limiting normal in part [h.i] follows from the CLT for covariances in Assumption 1, because $(x_{0t} - \Sigma_{x0} \Sigma_{00}^{-1} w_{0t})$ is uncorrelated with $(\varepsilon_{t+1} - x'_{0t} \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0})$. The limiting variance is given by

$$\begin{aligned} &\lim_{T \rightarrow \infty} \text{var} (T^{-1/2} \sum (x_{0t} - \Sigma_{x0} \Sigma_{00}^{-1} w_{0t}) (\varepsilon_{t+1} - x'_{0t} \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0})) \\ &= \lim_{T \rightarrow \infty} T^{-1} \mathbf{E} (\varepsilon_{t+1} - x'_{0t} \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0})^2 \mathbf{E} (x_{0t} - \Sigma_{x0} \Sigma_{00}^{-1} w_{0t}) (x_{0t} - \Sigma_{x0} \Sigma_{00}^{-1} w_{0t})', \end{aligned}$$

which is Ω_P .

Finally, premultiplying P_T by π' and using parts [f.ii] and [g.iii] yields the result for part [h.ii]. \square

Lemma A2. *The MSFE of the MF model in (2) is minimized by*

$$[a] \quad \alpha_m = \alpha + \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0} + O(T^{-1}) \text{ when } x_{t-i/m} \sim I(0), \text{ and}$$

$$[b] \quad \alpha_m = \alpha + H_0^x \Sigma_{xx\perp 0}^{-1} \sigma_{x\varepsilon\perp 0} + O(T^{-1}) \text{ when } x_{t-i/m} \sim I(1),$$

using the variance notations of Section 3.1 for part [a] and those of Section 3.2 for part [b].

Proof of Lemma A2. The MSFE from a forecast using model (2) is given by

$$\mathbf{E}(\mathbf{E}(\varepsilon_{T+1} - x_T^{1:p'}(\alpha_m - \alpha) - w_T'(\gamma_m - \gamma)) | \mathcal{F}_T)^2.$$

Expanding the square, taking the derivative with respect to the two parameter vectors α and γ , setting the derivative equal to zero, and solving for α_m yields $\alpha_m = \alpha + B^{-1}C$ where

$$B = \mathbf{E}[x_T^{1:p} x_T^{1:p'}] - \mathbf{E}[x_T^{1:p} w_T'] \mathbf{E}[w_T w_T']^{-1} \mathbf{E}[w_T x_T^{1:p'}], \quad \text{and}$$

$$C = \mathbf{E}[x_T^{1:p} \varepsilon_{T+1}] - \mathbf{E}[x_T^{1:p} w_T'] \mathbf{E}[w_T w_T']^{-1} \mathbf{E}[w_T \varepsilon_{T+1}].$$

for both parts [a] and [b].

For part [a], $\mathbf{E}[x_T^{1:p} x_T^{1:p'}] = \Sigma_{xx}$ and $\mathbf{E}[x_T^{1:p} \varepsilon_{T+1}] = \sigma_{x\varepsilon}$ since both series are $I(0)$. Rewrite the second term of B as

$$T^{1/2} \mathbf{E}[x_T^{1:p} w_T'] \varpi_T^{-1/2} \left(T \varpi_T^{-1/2} \mathbf{E}[w_T w_T'] \varpi_T^{-1/2} \right)^{-1} T^{1/2} \varpi_T^{-1/2} \mathbf{E}[w_T x_T^{1:p'}]$$

using ϖ_T defined at the beginning of this appendix. Now, $T^{1/2} \varpi_T^{-1/2} \mathbf{E}[w_T x_T^{1:p'}] = (\Sigma_{x0}, T^{-1/2} \Delta_{x1}, 0)'$, where $\Delta_{x1} = \sum_{k=0}^{\infty} \mathbf{E}[x_T^{1:p} \Delta w'_{1,T-k}]$. Along the same lines, $T^{1/2} \varpi_T^{-1/2} \mathbf{E}[w_T \varepsilon_{T+1}] = (\sigma_{\varepsilon 0}, T^{-1/2} \delta_{\varepsilon 1}, 0)'$ in C , where $\delta_{\varepsilon 1} = \sum_{k=0}^{\infty} \mathbf{E}[\varepsilon_{T+1} \Delta w'_{1,T-k}]$.

The more complicated factor $T \varpi_T^{-1/2} \mathbf{E}[w_T w_T'] \varpi_T^{-1/2}$ in B and C is equal to

$$\begin{bmatrix} \Sigma_{00} & T^{-1/2} \Delta_{01} & 0 \\ T^{-1/2} \Delta'_{01} & \Omega_{11} & 0 \\ 0 & 0 & T^{-2\delta} w_{2t} w'_{2t} \end{bmatrix},$$

where $\Delta_{01} = \sum_{k=0}^{\infty} \mathbf{E}[w_{0T} \Delta w'_{1,T-k}]$ and $\Omega_{11} = \sum_{k=0}^{\infty} \mathbf{E}[\Delta w_{1T} \Delta w'_{1,T-k}]$. Some algebra shows that its inverse is equal to

$$\begin{bmatrix} \Sigma_{00}^{-1} + O(T^{-1}) & -T^{-1/2} \Sigma_{00}^{-1} \Delta_{01} \Omega_{11}^{-1} + O(T^{-3/2}) & 0 \\ -T^{-1/2} \Omega_{11}^{-1} \Delta'_{01} \Sigma_{00}^{-1} + O(T^{-3/2}) & \Omega_{11}^{-1} + O(T^{-1}) & 0 \\ 0 & 0 & (T^{-2\delta} w_{2t} w'_{2t})^{-1} \end{bmatrix}, \quad (.1)$$

and the stated result for part [a] follows from multiplying out the expressions for B and C .

For part [b], it is straightforward to see that

$$\mathbf{E}[x_T^{1:p} x_T^{1:p'}] = H^x \begin{bmatrix} \Sigma_{xx} & \Delta_{xz} \\ \Delta'_{xz} & T\Omega_{zz} \end{bmatrix} H^{x'}$$

in the first term of B , with H^x defined at the beginning of Section 3.2. Now,

$$T^{1/2} \mathbf{E}[x_T^{1:p} w_T'] \varpi_T^{-1/2} = H^x \begin{bmatrix} \Sigma_{x0} & T^{-1/2} \Delta_{x1} & 0 \\ \Delta'_{0z} & T^{1/2} \Omega_{z1} & 0 \end{bmatrix}$$

and together with the expression for the inverse of $T\varpi_T^{-1/2} \mathbf{E}[w_T w_T'] \varpi_T^{-1/2}$ in (.1), some algebra reveals that

$$B = H^x \begin{bmatrix} \Sigma_{xx\perp 0} + O(T^{-1}) & O(1) \\ O(1) & O(T) \end{bmatrix} H^{x'} \quad \text{and} \quad B^{-1} = H^x \begin{bmatrix} \Sigma_{xx\perp 0}^{-1} & 0 \\ 0 & 0 \end{bmatrix} H^{x'} + O(T^{-1}).$$

Algebra along similar lines shows that

$$C = H^x \begin{bmatrix} \sigma_{x\varepsilon\perp 0} \\ O(1) \end{bmatrix} + O(T^{-1}).$$

Postmultiplying B^{-1} by C and noting that $H^{x'} H^x = I$ yields the stated result for part [b].

□

Proof of Theorem 1. Under the null, $\pi_m = \pi_0$ so that $R_T(\pi_0)$ is identically zero and $\Upsilon'(I - A_T)P_T$ remains in (9). Lemma A1 may be employed by letting $x_t^{1:p} \sim \text{I}(0)$, defining $H^x = H_0^x = I$ so that $x_{0t} = x_t^{1:p}$, and ignoring results pertaining to H_1^x . Consequently,

$$A_T \rightarrow_p \Sigma_{xx\perp 0} \pi_0 (\pi_0' \Sigma_{xx\perp 0} \pi_0)^{-1} \pi_0' = A,$$

follows from part [g.i] of the lemma, and $T^{1/2} P_T \rightarrow_d \Omega_P^{1/2} \mathbf{N}(0, I_p)$ follows from part [h.i]. The limiting variance Ω_P has full rank p unless the HF predictors are perfectly collinear and reflects the variance of the sample covariance of (a) the part of $(x_t^{1:p})$ orthogonal to the other $\text{I}(0)$ predictors (w_{0t}) with (b) the part of (ε_{t+1}) orthogonal to all of the $\text{I}(0)$ regressors.

The limit of the expression in (9) inflated by $T^{1/2}$ is therefore given by

$$T^{-1/2} \sum \Upsilon' x_t^{1:p} \hat{\varepsilon}_{t+1} \rightarrow_d \Upsilon'(I - A) \Omega_P^{1/2} \mathbf{N}(0, I_p),$$

the variance of which has a rank of q . The rank follows because $I - A$ is an orthogonal projection onto a $(p - 1)$ -dimensional space. Υ has $q + 1$ columns, but since one of them is π_0 and $\pi_0'(I - A) = 0$, the rank is reduced to q as long as $q \leq p - 1$.

Looking now at the inverse in the expression for φ in (6), the limit of $\Upsilon' M_{xx}^T \Upsilon$ is clearly $\Upsilon' \Sigma_{xx} \Upsilon$, which has a full rank of $q + 1$ and is therefore invertible. The whole expression therefore has a limiting distribution given by

$$T^{1/2} \hat{\varphi} \rightarrow_d (\Upsilon' \Sigma_{xx} \Upsilon)^{-1} \Upsilon' (I - A) \Omega_P^{1/2} \mathbf{N}(0, I_p),$$

the variance of which has rank q .

The q degrees of freedom of the limiting distribution of the test statistic under the null follows, because the test omits the first regressor, which corresponds to the direction in which $\Upsilon' (I - A)$ is already rank deficient.

To obtain the χ_q^2 limiting distribution under the null,

$$\hat{\Omega}_q \rightarrow_p \Omega_q = (\Upsilon' \Sigma_{xx} \Upsilon)^{-1} \Upsilon' (I - A) \Omega_P (I - A) \Upsilon (\Upsilon' \Sigma_{xx} \Upsilon)^{-1}$$

must hold. This limit follows by using a standard heteroskedasticity autocorrelation robust variance estimator $\hat{\Omega}_q$ with stationary $(\varepsilon_{t+1}, x_t^{1:p'})'$.

Now consider an alternative given by $\pi_A = \pi(\theta_0 + c g_T^{-1})$. Denote the $m \times \ell$ matrix of first derivatives of $\pi(\theta)$ evaluated at θ_0 by $\Pi(\theta_0)$. Under such alternative,

$$R_T(\pi_A) = M_{xx \perp w}^T (\pi_A - \pi_0) \beta_m = g_T^{-1} \Sigma_{xx \perp 0}^{-1} \Pi(\theta_0) c \beta_m + o_p(g_T^{-1})$$

in (9). In this case,

$$g_T \hat{\varphi} = g_T (\Upsilon' M_{xx}^T \Upsilon)^{-1} \Upsilon' \left(T^{-1} \sum x_t^{1:p} \hat{\varepsilon}_{t+1} \right) \rightarrow_p (\Upsilon' \Sigma_{xx} \Upsilon)^{-1} \Upsilon' (I - A) \Sigma_{xx \perp 0}^{-1} \Pi(\theta_0) c \beta_m$$

so that $T^{1/2} \hat{\varphi} = O_p(T^{1/2} g_T^{-1})$. Since the test statistic is quadratic in $T^{1/2} \hat{\varphi}$, it is $O_p(T g_T^{-2})$ under the alternative. \square

Proof of Theorem 2. Consider the I(1) case first. Recall that the columns of H_0^x characterize the cointegrating relationships of $x_t^{1:p}$ and H_1^x characterizes the stochastic trend, where $H^x = (H_0^x, H_1^x)$ is an orthogonal matrix. The regressor vector in the test regression $\Upsilon' x_t^{1:p}$ is itself a $(q + 1)$ -vector of cointegrated regressors with a single common stochastic trend(s). Similarly to H^x , define the $(q + 1) \times (q + 1)$ orthogonal matrix $H^q = (H_0^q, H_1^q)$, such that $H_0^q \Upsilon' x_t^{1:p} \sim I(0)$ and $H_1^q \Upsilon' x_t^{1:p} \sim I(1)$, where the former is a q -vector series while the latter is a scalar series.

The modified VAT statistic $V_T^* = T\hat{\varphi}'\Xi\hat{\sigma}_u^{-2}\Upsilon'M_{xx}^T\Upsilon\Xi\hat{\varphi}$ may be written as

$$V_T^* = \frac{1}{T^{-2\epsilon}\hat{\sigma}_u^2} \left[T^{-\epsilon}v_T^{1/2}H^{q'}\hat{\varphi} \right]' \Xi \left\{ Tv_T^{-1/2}H^{q'}\Upsilon'M_{xx}^T\Upsilon H^q v_T^{-1/2} \right\} \Xi \left[T^{-\epsilon}v_T^{1/2}H^{q'}\hat{\varphi} \right], \quad (.2)$$

where $\hat{\sigma}_u^2 = T^{-1} \sum (\hat{\varepsilon}_{t+1}^* - x_t^{1:p'}\Upsilon\hat{\varphi})^2$ and $\Xi = E(E'E)^{-1}E'$ such that E' chooses all but the first row of the matrix that follows.

Because the series $(x_t^{1:p})$ has only a single stochastic trend and unit cointegrating coefficients, and because the columns of Υ sum to one, $\Upsilon'H_1^x = \iota_{q+1}c_p$ where ι_{q+1} is a $(q+1)$ -vector of ones and c_p is a scalar that depends on p . Further, because $(\Upsilon'x_t^{1:p})$ has only a single stochastic trend and unit cointegrating coefficients, $H_0^{q'}\Upsilon'H_1^x = H_0^{q'}\iota_{q+1}c_p = 0$, and thus, $H_0^{q'}\Upsilon'x_t^{1:p} = (H_0^{q'}\Upsilon'H_0^x)x_{0t}$.

Looking first at the factor in curly brackets in (.2), some manipulation using these facts allows

$$\begin{aligned} & Tv_T^{-1/2}H^{q'}\Upsilon'M_{xx}^T\Upsilon H^q v_T^{-1/2} \\ &= \begin{bmatrix} H_0^{q'}\Upsilon'H_0^x T^{-1} \sum x_{0t}x'_{0t}H_0^{x'}\Upsilon H_0^q & O_p(T^{-1/2}) \\ O_p(T^{-1/2}) & H_1^{q'}\Upsilon'H_1^x T^{-2} \sum x_{1t}x'_{1t}H_1^{x'}\Upsilon H_1^q + O_p(T^{-1}) \end{bmatrix} \\ &\rightarrow_d \begin{bmatrix} H_0^{q'}\Upsilon'H_0\Sigma_{xx}H_0^q & 0 \\ 0 & H_1^{q'}\Upsilon'H_1^x \int B_z B'_z H_1^{x'}\Upsilon H_1^q \end{bmatrix} \end{aligned} \quad (.3)$$

from parts [e.i]-[e.iii] of Lemma A1. The limiting matrix is a $(q+1) \times (q+1)$ matrix with full rank.

Using the least squares estimator of φ in (6) augmented with $\hat{\varepsilon}_{t+1}^*$, the factors in square brackets in (.2) may be written as

$$\begin{aligned} T^{-\epsilon}v_T^{1/2}H^{q'}\hat{\varphi} &= \left\{ Tv_T^{-1/2}H^{q'}\Upsilon'M_{xx}^T\Upsilon H^q v_T^{-1/2} \right\}^{-1} \\ &\times \begin{bmatrix} H_0^{q'}\Upsilon'H_0^x T^{-1/2-\epsilon} \sum x_{0t}\hat{\varepsilon}_{t+1}^* \\ H_1^{q'}\Upsilon'H_0^x T^{-1-\epsilon} \sum x_{0t}\hat{\varepsilon}_{t+1}^* + H_1^{q'}\Upsilon'H_1^x T^{-1-\epsilon} \sum x_{1t}\hat{\varepsilon}_{t+1}^* \end{bmatrix}. \end{aligned} \quad (.4)$$

as in (.2). Suppose the null holds, so that $R_T(\pi_m) = 0$. The factor in square brackets in (.4) may be written as

$$T^{-\epsilon} \begin{bmatrix} T^{1/2}H_0^{q'}\Upsilon'H_0^{x'}(I - A_T)P_T \\ H_1^{q'}\Upsilon'(I - A_T)P_T \end{bmatrix} + \begin{bmatrix} H_0^{q'}\Upsilon'T^{-1/2} \sum x_{0t}u_{t+1} \\ H_1^{q'}\Upsilon'H_1^x T^{-1} \sum x_{1t}u_{t+1} + O_p(T^{-1/2}) \end{bmatrix} \quad (.5)$$

using (9).

Looking at the top part of the first term of (.5), $T^{1/2}H_0^{x'}P_T = O_p(1)$ from Lemma

A1[h.i]. Moreover,

$$T^{1/2}H'_0A_T P_T = T^{-1/2}H'_0M_{xx\perp w}^T\pi_0 (T^{-1}\pi'_0M_{xx\perp w}^T\pi_0)^{-1}\pi'_0P_T,$$

which is $O_p(T^{-1/2})$ from parts [g.ii], [g.iii], and [h.ii] of Lemma A1. Looking at the bottom part of the first term of (.5), $H_1^{q'}\Upsilon'P_T = O_p(1)$ by Lemma A1[h.ii]. Write

$$H_1^{q'}\Upsilon'A_T P_T = T^{-1}H_1^{q'}\Upsilon'M_{xx\perp w}^T\pi_0 (T^{-1}\pi'_0M_{xx\perp w}^T\pi_0)^{-1}\pi'_0P_T$$

and then use parts [g.ii] (twice) and [g.iii] to see that this expression is also $O_p(1)$. Hence, the first term of (.5) is $T^{-\epsilon}O_p(1) = o_p(1)$.

Since $u_{t+1} \sim \mathbf{iidN}(0, \sigma_u^2)$ and independent of (x_{0t}) and (Δx_{1t}) , the limit of the second term of (.5) and thus of the factor in square brackets in (.4) is

$$\left[\begin{array}{c} T^{-1/2-\epsilon}H_0^{x'} \sum x_t^{1:p} \hat{\varepsilon}_{t+1}^* \\ T^{-1-\epsilon}H_1^{x'} \sum x_t^{1:p} \hat{\varepsilon}_{t+1}^* \end{array} \right] \rightarrow_d \mathbf{MN} \left(0, \sigma_u^2 \left[\begin{array}{cc} H_0^{q'}\Upsilon'H_0\Sigma_{xx}H_0'\Upsilon H_0^q & 0 \\ 0 & H_1^{q'}\Upsilon'H_1^x \int B_z B_z' H_1^{x'}\Upsilon H_1^q \end{array} \right] \right)$$

from Lemma A1[c.i]-[c.ii].

The factor in curly brackets in (.4) is the same as that in (.2). Taken together,

$$T^{-\epsilon}v_T^{1/2}H^{q'}\hat{\varphi} \rightarrow_d \mathbf{MN} \left(0, \sigma_u^2 \left[\begin{array}{cc} H_0^{q'}\Upsilon'H_0\Sigma_{xx}H_0'\Upsilon H_0^q & 0 \\ 0 & H_1^{q'}\Upsilon'H_1^x \int B_z B_z' H_1^{x'}\Upsilon H_1^q \end{array} \right]^{-1} \right) \quad (.6)$$

for the factors in square brackets in (.2).

Now, $T^{-2\epsilon}\hat{\sigma}_u^2$ in (.2) may be written as

$$T^{-1-2\epsilon} \sum (T^\epsilon u_{t+1} + (\hat{\varepsilon}_{t+1} - x_t^{1:p'}\Upsilon\hat{\varphi}))^2 = T^{-1} \sum u_{t+1}^2 + T^{-2\epsilon}Q_{1T} + T^{-\epsilon}2Q_{2T}$$

by using the definition of $\hat{\varepsilon}_{t+1}^*$ and defining

$$\begin{aligned} Q_{1T} &= T^{-1} \sum (\hat{\varepsilon}_{t+1} - x_t^{1:p'}\Upsilon\hat{\varphi})^2 \\ Q_{2T} &= T^{-1} \sum (\hat{\varepsilon}_{t+1} - x_t^{1:p'}\Upsilon\hat{\varphi})u_{t+1}. \end{aligned}$$

Clearly, $T^{-2\epsilon}\hat{\sigma}_u^2$ is a consistent estimator of σ_u^2 if Q_{1T} and Q_{2T} are $O_p(1)$. An expansion of Q_{1T} has (a) $T^{-1} \sum x_t^{1:p} \hat{\varepsilon}_{t+1}$, which has already shown to be $O_p(1)$, (b) $\hat{\varphi}'\Upsilon'M_{xx}^T\Upsilon\hat{\varphi} = (T^{-1/2}v_T^{1/2}H^{q'}\hat{\varphi})'(T v_T^{-1/2}H^{q'}\Upsilon'M_{xx}^T\Upsilon H^q v_T^{-1/2})(T^{-1/2}v_T^{1/2}H^{q'}\hat{\varphi})$, which is $O_p(1)$ for $\epsilon < 1/2$

along the lines discussed above, and (c)

$$T^{-1} \sum \hat{\varepsilon}_{t+1}^2 = T^{-1} \sum \varepsilon_{t+1} \hat{\varepsilon}_{t+1}$$

which is $O_p(1)$ similarly to $T^{-1} \sum x_t^{1:p} \hat{\varepsilon}_{t+1}$. Q_{2T} may likewise be shown to be $O_p(1)$.

Under the null, the χ^2 limit of V_T^* follows from the mixed normal distribution in (.6) of the factors in square brackets in (.2), the distribution in (.4) of the factor in curly brackets in (.2), and consistency of $T^{-2\epsilon} \hat{\sigma}_u^2$. The degrees of freedom q comes from selecting only q elements of $\hat{\varphi}$ in $\Xi \hat{\varphi}$ for the test.

Now, suppose that the alternative holds. The square-bracketed term in (.4) has an additional term that is identical to that with P_T but with P_T replaced by $R_T(\pi_A)$, which has a leading term of $g_T^{-1} M_{xx \perp w}^T \Pi(\theta_0) c \beta_m$ as in the proof of Theorem 1. Since $\beta_m \neq 0$,

$$g_T^{-1} T^{-\epsilon} \begin{bmatrix} T^{1/2} H_0^{q'} \Upsilon' H_0^x (I - A_T) M_{xx \perp w}^T \Pi(\theta_0) \\ H_1^{q'} \Upsilon' (I - A_T) M_{xx \perp w}^T \Pi(\theta_0) \end{bmatrix} c \beta_m = g_T^{-1} T^{-\epsilon} \begin{bmatrix} O_p(T^{1/2}) \\ O_p(T) \end{bmatrix}$$

where the $O_p(T^{1/2})$ and $O_p(T)$ follow by rewriting as above and repeatedly using parts [g.ii] and [g.iii] of Lemma A1. It follows that

$$T^{-\epsilon} v_T^{1/2} H^{q'} \hat{\varphi} = g_T^{-1} T^{-\epsilon} \begin{bmatrix} O_p(T^{1/2}) \\ O_p(T) \end{bmatrix},$$

too, so that the test statistic, which is quadratic in this expression, is $O_p(T^{2(1-\epsilon)} g_T^{-2})$.

Finally, consider the I(0) case. In this case, $H_0^q = I_{q+1}$ and $H_0^x = I_p$ and H_1^q and H_1^x are not defined. The proof carries through as a special case of that above until $T^{1/2-\epsilon} v_T^{1/2} \hat{\varphi} = O_p(T^{1/2-\epsilon} g_T^{-1})$ under the alternative, so that the test statistic is only $O_p(T^{1-2\epsilon} g_T^{-2})$ in that case. \square

References

- Andreou, E., E. Ghysels, and A. Kourtellos (2010). "Regression Models with Mixed Sampling Frequencies," *Journal of Econometrics* 158, 246-261.
- Andreou, E., E. Ghysels, and A. Kourtellos (2011). "Forecasting with Mixed-Frequency Data," in *Oxford Handbook of Economic Forecasting*, M.P. Clements and D.F. Hendry, eds. Oxford: Oxford University Press, 225-246.
- Andreou, E., E. Ghysels, and A. Kourtellos (2013). "Should Macroeconomic Forecasters

- Use Daily Financial Data and How?" *Journal of Business and Economic Statistics* 21, 240-251.
- Armesto, M.T., K.M. Engemann, and M.T. Owyang (2010). "Forecasting with Mixed Frequencies," *Federal Reserve Bank of St. Louis Review* 92, 521-536.
- Breitung, J. and C. Roling (2015). "Forecasting Inflation Rates Using Daily Data: A Nonparametric MIDAS Approach," *Journal of Forecasting* 34, 588-603.
- Chambers, M.J. (2015). "Testing for a Unit Root in a Near-Integrated Model with Skip-Sampled Data," *Journal of Time Series Analysis* 36, 630-649.
- Clark, T.E. and M.W. McCracken (2001). "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105, 85-110.
- Clements, M.P. and A. Galvão (2008). "Macroeconomic Forecasting with Mixed-Frequency Data: Forecasting Output Growth in the United States," *Journal of Business and Economic Statistics* 26, 546-554.
- Clements, M.P. and A. Galvão (2009). "Forecasting US Output Growth Using Leading Indicators: An Appraisal Using MIDAS Models," *Journal of Applied Econometrics* 24, 1187-1206.
- Diebold, F.X. (2015). "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," *Journal of Business and Economic Statistics* 33, 1-24.
- Elliott, G. (2006). "Forecasting with Trending Data," in *Handbook of Economic Forecasting, Volume 1*, G. Elliott, C.W.J. Granger, and A. Timmermann, eds. Amsterdam: North Holland, 555-604.
- Ghysels, E. and J.I. Miller (2015). "Testing for Cointegration with Temporally Aggregated and Mixed-Frequency Time Series," *Journal of Time Series Analysis* 36, 797-816.
- Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). "The MIDAS Touch: Mixed Data Sampling Regression Models," CIRANO Working Paper 2004s-20.
- Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). "Predicting Volatility: Getting the Most Out of Return Data Sampled at Different Frequencies," *Journal of Econometrics* 131, 59-95.

- 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.
- Hausman, J.A. (1978). “Specification Tests in Econometrics,” *Econometrica* 46, 1251-1271.
- Hogrefe, J. (2008). “Forecasting Data Revisions of GDP: A Mixed Frequency Approach,” *Advances in Statistical Analysis* 92, 271-296.
- Kuzin, V., M. Marcellino, and C. Schumacher (2011). “MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area,” *International Journal of Forecasting* 27, 529-542.
- Kvedaras, V. and V. Zemlys (2012). “Testing the Functional Constraints on Parameters in Regressions with Variables of Different Frequency,” *Economics Letters* 116, 250-254.
- Lütkepohl, H. (1987). *Forecasting Aggregated Vector ARMA Processes*. Berlin: Springer-Verlag.
- Marcellino, M. (1999). “Some Consequences of Temporal Aggregation in Empirical Analysis,” *Journal of Business and Economic Statistics* 17, 129-136.
- Marcellino, M. and C. Schumacher (2010). “Factor-MIDAS for Now- and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP,” *Oxford Bulletin of Economics and Statistics* 72, 518-550.
- 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.
- Park, J.Y. (1990). “Testing for Unit Roots and Cointegration by Variable Addition.” In: G.F. Rhodes & T.B. Fomby (eds.), *Advances in Econometrics*. Greenwich: JAI Press, 107-133.
- Phillips, P.C.B. (1988). “Multiple Regression with Integrated Time Series,” *Contemporary Mathematics* 80, 79-105.
- Phillips, P.C.B. (1991). “Error Correction and Long-Run Equilibrium in Continuous Time,” *Econometrica* 59, 967-980.

- Schumacher, C. and J. Breitung (2008). "Real-Time Forecasting of German GDP Based on a Large Factor Model with Monthly and Quarterly Data," *International Journal of Forecasting* 24, 386-398.
- Tay, A.S. (2007). "Financial Variables as Predictors of Real Output Growth," School of Economics and Social Sciences, Singapore Management University, mimeo.
- West, K.D. (1996). "Asymptotic Inference about Predictive Ability," *Econometrica* 64, 1067-1084.
- Working, H. (1960) "Note on the Correlation of First Differences of Averages in Random Chain," *Econometrica* 28, 916-918.
- Wu, D.-M. (1973). "Alternative Tests of Independence between Stochastic Regressors and Disturbances," *Econometrica* 41, 733-750.