Cointegrating MiDaS Regressions and a MiDaS Test

J. Isaac Miller[1,2]

Department of Economics, University of Missouri

118 Professional Building, Columbia, Missouri 65211, USA

Abstract

This paper introduces cointegrating mixed data sampling (CoMiDaS) regressions, generalizing nonlinear MiDaS regressions in the extant literature. Under a linear mixed-frequency data-generating process, MiDaS regressions provide a parsimoniously parameterized nonlinear alternative when the linear forecasting model is over-parameterized and may be infeasible. In spite of potential correlation of the error term both serially and with the regressors, I find that nonlinear least squares consistently estimates the minimum mean-squared forecast error parameter vector. The exact asymptotic distribution of the difference may be non-standard. I propose a novel testing strategy for nonlinear MiDaS and CoMiDaS regressions against a general but possibly infeasible linear alternative. An empirical application to nowcasting global real economic activity using monthly covariates illustrates the utility of the approach.

[This paper was substantially revised and updated as "Mixed-frequency Cointegrating Regressions with Parsimonious Distributed Lag Structures." The interest reader is referred to that version, available as WP 12-11 in the working paper series of the Economics Department at Missouri and posted in August 2012.]

First draft: April 15, 2011
Current draft: February 8, 2012

JEL Classification: C12, C13, C22

Key words and phrases: cointegration, mixed-frequency series, mixed data sampling

[1]Correspondence: millerjisaac@missouri.edu.

[2]This draft was prepared for submission to the 2012 North American Summer Meeting of the Econometric Society and the 2012 SMU-ESSEC Symposium on Empirical Finance and Financial Econometrics. The author is grateful to Eric Ghysels, Peter Schmidt, Jeff Wooldridge, and other participants of the 2011 NBER-NSF Time Series Conference, the 2011 meeting of the Midwest Econometrics Group, seminars at UNC-Chapel Hill and Michigan State, and a colloquium at Missouri for constructive feedback. The author acknowledges financial support from the Economic & Policy Analysis Research Center at the University of Missouri and excellent reseach assistance from April Li. The usual caveat applies.

1. Introduction

Since the introduction of nonlinearly specified mixed data sampling (MiDaS) regressions by Ghysels *et al.* (2004) and Ghysels *et al.* (2007), the MiDaS methodology has been adopted for a variety of applied and empirical topics in economics and finance, particularly for the purposes of forecasting – e.g., Ghysels and Wright (2009), Armesto *et al.* (2010), and Andreou *et al.* (2010b). The proliferation of empirical applications of MiDaS forecasting models include primarily output – e.g., Tay (2007), Clements and Galvão (2008, 2009), Hogrefe (2008), Benjanuvatra (2009), Frale and Montefort (2010), Marcellino and Schumacher (2010), and Kuzin *et al.* (2011). Another common application of MiDaS is to modeling volatility – e.g., Ghysels *et al.* (2006), Alper *et al.* (2008), Ghysels *et al.* (2009), Chen and Ghysels (2010), and Ghysels and Valkanov (2010). Also, MiDaS forecasting models have been applied to study the effects of high-frequency monetary policy shocks (Armesto *et al.*, 2009, Francis *et al.*, 2010), inflation (Montefort and Moretti, 2010), and the risk-return trade-off (Ghysels *et al.* 2005).

The principal advantage of the parsimonious nonlinear autoregressive distributed lag (ADL) specification provided by the MiDaS approach lies in allowing regressors measured at a much higher frequency than the regressand. In particular, if the high-low frequency ratio exceeds the number of low frequency observations – e.g., 50 years of weekly data – a linear ADL regression is infeasible. Adding regressors exacerbates this problem. For a lower ratio, such a model may be feasible but simply not parsimonious. The MiDaS specification offers a trade-off. The gain from a reduction in the dimension of the parameter space may be quite large. In the extreme, an infeasible model becomes feasible. More generally, an efficiency gain is possible. The price paid may be inconsistency from neglected nonlinearity, if the MiDaS specification does not nest or at least approximate the data-generating process (DGP).

This paper aims for three important contributions to the existing MiDaS literature. I introduce cointegrating MiDaS (CoMiDaS) regressions, which nest stationary MiDaS regressions as a special case. The generalization allows for the possibility of stochastic trends common to the time series in the model. Cointegration in mixed-frequency and closely related temporally aggregated models has been studied extensively. Such studies include Granger (1990), Gomez and Maravall (1994), Granger and Siklos (1995), Marcellino (1999), Haug (2002), Chambers (2003, 2009, 2010), Pons and Sansó (2005), Chambers and McCrorie (2007), Seong *et al.* (2007), and Miller (2010, 2011). The mixed-frequency framework of Ghysels *et al.* (2004) does not allow for the possibility of stochastic trends. On the other hand, some authors who have worked with macroeconomic series believed to contain unit roots but with unique stochastic trends, such as Clements and Galvão (2008), have taken first differences. The CoMiDaS regressions I introduce nests both of these cases. The statistical analysis allows for the possibility of any number of trends, with these two cases as the extremes.

Second, I allow for the realistic possibility that the error term is correlated both serially and with the regressors. Although this allowance is somewhat standard for models with I(1) series, since estimators may still be consistent, I allow correlation with both I(1) and I(0) regressors. Such correlation causes inconsistency in estimating the parameters of the

DGP. However, since MiDaS regressions are often used in forecasting, a more appropriate outcome is to minimize mean-squared forecast error (MSFE) or a similar loss function.[3] I show that even with such correlation, the MiDaS specification may consistently estimate the minimum MSFE parameter vector within the class of models nested by the MiDaS specification and assuming a MiDaS DGP, as is assumed by Ghysels *et al.* (2004), Andreou *et al.* (2010a), *inter alia.*

The extension of the MiDaS framework in these two directions provides a justification for the MiDaS approach taken by previous authors, but develops a broader framework on which to base further analysis.

Third, I propose a novel test of the MiDaS null against a more general alternative. Like those of Andreou *et al.* (2010a), the proposed test is for in-sample rather than out-of-sample fit. Their tests posit a specific weighting scheme nested by the MiDaS specification as the null, with a maintained hypothesis that the DGP is nested by the MiDaS specification under both the null and the alternative. In contrast, I make this assumption only under the null. If estimating the linear ADL model is feasible (m is small), a likelihood ratio test similar to that proposed for ADL models by Godfrey and Poskitt (1975) could be used to test the MiDaS null.[4] If estimating the linear ADL model is infeasible, which is precisely the case in which one would expect the most gain from the parsimonious specification, the Godfrey-Poskitt test is infeasible.

The test I propose is feasible in either case. Specifically, I base the test on a traditional variable addition test (Wu, 1973), where the variables added are arbitrary linear combinations of the high-frequency regressors. Although one of the tests proposed by Andreou *et al.* (2010a) is also a variable addition test, it has a different null and requires m to be small enough for the linear ADL estimator to be feasible. Whereas both their tests with MiDaS alternatives are useful for testing one MiDaS specification against another, my proposed test, like that of Godfrey and Poskitt (1975), is useful for distinguishing a general MiDaS specification from a more general DGP with linear ADL structure.

The remainder of the paper is structured along the following lines. I outline the basic cointegrating MiDaS regressions in Section 2 and present a somewhat more general cointegrating regression with nonlinearity in the regression coefficients. I derive the difference between the NLS estimator and the minimum MSFE parameter vector. In section 3, I show the consistency of the regression coefficients and of the deep parameters underlying these coefficients to the analogous minimum MSFE parameters. I derive the asymptotic distributions of the respective differences, which may be Gaussian or nonstandard under alternative assumptions. I then introduce the proposed test, evaluating both its asymptotic distribution and small sample performance in Section 4. In Section 5, I present an illustrative application to forecasting global real economic activity, and I conclude with Section 6. An appendix contains proofs of the theoretical results.

---

[3] Clements and Hendry (1995) and Christofferson and Diebold (1998) suggested alternative measures for forecasts from a cointegrated system, but I consider forecasts from only a single cointegrating regression.

[4] The Godfrey-Poskitt results are derived using an Almon lag – rather than *exponential* Almon lag – specification. Similar results may hold for the exponential specification.

2. CoMiDaS Regressions

Consider the task of forecasting the change in a macroeconomic series observed at a low frequency, using lags of the same series, several series observed at a higher frequency during the previous low-frequency period, and perhaps some additional regressors observed at the previous low-frequency period. In levels, such a forecasting regression may be written as

$$y_{t+1} = \sum\nolimits_{k=1}^{p} \rho_k y_{t+1-k} + \varphi' w_t + \beta' \sum\nolimits_{k=0}^{m-1} \Pi_{k+1} x_{t-k/m}^{(m)} + \varepsilon_{t+1}, \tag{1}$$

where the superscript (m) denotes the higher frequency – specifically, m times the lower frequency, and Π_s is a diagonal matrix of unknown weights. The weight structure may differ across regressors series, but for expositional simplicity I assume that all regressor series are observed at the same frequency.

As an example, the best forecast of an annual series may use an annual average (*average* or *flat* sampling) of each monthly regressor, in which case the optimal weights would be $1/m$ for each of these high-frequency regressors. End-of-period sampling (a special case of *selective* or *skip* sampling), provides another example. That case assigns a unit to the first weight and zeros to the remaining weights.

Subtracting y_t from both sides and manipulating $x_{t-k/m}^{(m)}$ under the assumption that $\sum_{k=0}^{m-1} \Pi_{k+1} = I$ allows an error correction representation

$$\triangle y_{t+1} = \left(\rho^* y_t + \varphi' w_t + \beta' x_t\right) - \rho^*(L) \triangle y_t - \beta' \Pi(L) \triangle^{(1/m)} x_t^{(m)} + \varepsilon_{t+1}, \tag{2}$$

where $\rho^* \equiv \sum_{k=1}^{p} \rho_k - 1$, $\rho^*(z) \equiv \sum_{k=1}^{p-1} \sum_{s=k+1}^{p} \rho_s z^{k-1}$ for $p \geq 2$ and $\rho^*(z) = 0$ for $p = 1$, and where $\Pi(z) \equiv \sum_{k=0}^{m-2} \sum_{s=k+1}^{m-1} \Pi_{s+1} z^{k/m}$. I use the notation $\triangle^{(1/m)}$ to denote a high-frequency difference, and I drop the superscript (m) from $\beta' x_t^{(m)}$ to signify that this term includes only low-frequency observations of the high-frequency regressors.

The above models are not parsimonious. For m sufficiently large, they may not even be feasible. Ghysels *et al.* (2004) proposed MiDaS regressions, featuring an ADL with parsimonious nonlinear specification to overcome the infeasibility. Under such a specification, the weight matrices Π_k are parameterized by γ, so that $\Pi(z;\gamma) \equiv \sum_{k=0}^{m-2} \sum_{s=k+1}^{m-1} \Pi_{s+1}(\gamma) z^{k/m}$.

The MiDaS forecasting regression becomes

$$y_{t+1} = \sum\nolimits_{k=1}^{p} \rho_k y_{t+1-k} + \varphi' w_t + \beta' \sum\nolimits_{k=0}^{m-1} \Pi_{k+1}(\gamma) x_{t-k/m}^{(m)} + \eta_{t+1} \tag{3}$$

in levels, or

$$\triangle y_{t+1} = \left(\rho^* y_t + \varphi' w_t + \beta' x_t\right) - \rho^*(L) \triangle y_t - \beta' \Pi(L;\gamma) \triangle^{(1/m)} x_t^{(m)} + \eta_{t+1} \tag{4}$$

in differences, under this parameterization. The former is more similar to the I(0) MiDaS regression of Ghysels *et al.* (2004), while the latter is more similar to the I(1) MiDaS regression of Clements and Galvão (2008). When both low-frequency and high-frequency regressors contain a mix of I(0) and I(1) series, the CoMiDaS regression

$$\begin{aligned} \triangle y_{t+1} = {} & \left(\rho^* y_t + \varphi_1' w_{1t} + \beta_1' x_{1t}\right) + \varphi_0' w_{0t} + \beta_0' x_{0t} \\ & - \rho^*(L) \triangle y_t - \beta' \Pi(L;\gamma) \triangle^{(1/m)} x_t^{(m)} + \eta_{t+1}, \end{aligned} \tag{5}$$

where the subscripts 0 and 1 denote the order of integration, generalizes that in (4). As Andreou *et al.* (2010a) noted, all of the terms with I(1) regressors are linear in parameters.

Since the weight matrices (Π_s) are parameterized directly, there is no practical difference between defining the lag polynomial in (3), (4), and (5) except that the latter two impose that the weights sum to unity, resulting in $m-1$ rather than m weight matrices. Typical nonlinear weight specifications impose this restriction anyway.

The lag structure most commonly employed in the MiDaS literature is the exponential Almon lag, modified from Almon (1965). For the first diagonal of the weight matrix, the exponential Almon polynomial may be written as

$$\pi_{1,s}\left(\gamma\right) z^{s/m} = \frac{\exp(\gamma_1 s + \gamma_2 s^2)}{\sum_{j=1}^{m} \exp(\gamma_1 j + \gamma_2 j^2)} z^{s/m},$$

and the remaining diagonals are similar. The remaining diagonals may depend on γ_1 and γ_2, if the weight structure is assumed to be the same for all regressor series, or they may depend on additional pairs of parameter.

The literature posits alternative lag specifications, but the exponential Almon lag is employed for its flexibility in mimicking reasonable economic assumptions about the relationship between the low- and high-frequency data. For example, if the regressand is observed at the lower frequency due to average sampling and is cointegrated with the regressors, then Chambers (2003) and Miller (2011) showed that average sampling the regressors is most efficient. The exponential Almon lag achieves this for $\gamma_1 = \gamma_2 = 0$. If instead the optimal scheme is thought to be end-of-period sampling, then the exponential Almon lag provides an adequate approximation by setting, $\gamma = (-5, -5)$ for $m = 12$, say. In the cointegrating case, Miller (2011) showed that matching the aggregation scheme of the regressors to that of the regressand is efficient in the absence of correlation of the error with the regressors. Aggregating the regressors directly is more parsimonious than using a MiDaS approach, but would be inappropriate when the regressand aggregation scheme is unknown. The exponential Almon lag can take other shapes, assigning more weights to middle observations if, perhaps, seasonality is a concern.

2.1. Generalized Specification

As with a typical I(0) error-correction model, in order for the CoMiDaS regression to be well-specified, there are three possibilities. Either all series are I(0), some series are I(1) but no cointegrating relationships exist, $(\rho^*, \varphi_1', \beta_1')' = 0$, or some series are I(1) but at least one cointegrating relationship exists, $\rho^* y_t + \varphi' w_{1t} + \beta' x_{1t} \sim I(0)$.

To simplify notation, let (p_{1t}) denote the n_1 unique common stochastic trends of the I(1) vector $(y_t, w_{1t}', x_{1t}')'$ in (5), and let (p_{0t}) denote an n_0 series containing both the I(0) regressors and any cointegrating combinations of the I(1) regressors. The regression has $n = n_1 + n_0$ regressors: the number of stochastic trends of the I(1) series, plus the number of cointegrating combinations of the I(1) series, plus the number of stationary covariates. A univariate high-frequency (x_t) spawns m regressors, but no more than one stochastic trend. Such a stochastic trend may be shared with (y_t) and (w_t). Note that $n_1 = 0$ if all series are stationary, but that $n_0 = 0$ *cannot* hold, except in the case of selective sampling, or in the trivial case of no mixed frequencies. In other words, even when all high-frequency regressors are I(1), there will generally be I(0) terms in (2), (4), and (5) due to the error correction representation.

The regression may be rewritten as

$$\triangle y_{t+1} = \theta_1' p_{1t} + g_0' (\theta) p_{0t} + \eta_{t+1} \tag{6}$$

where $\theta_{g\times 1} = (\rho^*, \varphi_1', \beta_1', \rho_1^*, ..., \rho_p^*, \varphi_0', \beta_0', \gamma')'$ is an element of the parameter space $\Theta \subseteq \mathbb{R}^g$ and $g_0(\theta)$ is a vector of linear and nonlinear functions. I refer to the models in (3)-(6) as *CoMiDaS regressions* or *CoMiDaS models*.

If the model contains any unit roots, θ_1 must be zero, because the cointegrating combinations of $(y_t, w_{1t}', x_{1t}')'$ are in (p_{0t}) rather than (p_{1t}). However, the dimension n_1 of θ_1 varies depending on the number of unit roots. To allow for their presence, but since the stochastic trends are latent, I do not assume θ_1 to be zero in estimation, and I estimate all n coefficients. In other words, I do not impose a unit root. Nor do I impose a certain number of trends. Although determining the number of trends may be desirable in certain applications, it is not necessary in the present analysis.[5]

Imposing a unit root when the DGP does not contain one is clearly not desirable. Clements and Hendry (1995) suggested not imposing a unit root, but Christofferson and Diebold (1998) found that this leads to weaker forecasts when the DGP contains one. Hansen (2010) recently suggested averaging forecasts from models in which a unit root is imposed and not imposed when the smallest root of the DGP is uncertain but thought to be near unity.

Aside from potential unit roots, there is a major difficulty to overcome in analyzing the statistical properties of an estimator of θ using (6): potential correlation of (ε_t) both serially and with the regressors. Although it is well-known that such correlations pose no problem in the consistent estimation of the cointegrating vectors in a linear model, the effects on forecasting and on coefficients of I(0) regressors under a nonlinear specification are not obvious. I show that such correlation is not problematic for consistently estimating the minimum MSFE parameter vector in either I(0) or I(1) cases.

In order to analyze the statistical properties of an NLS estimator of the CoMiDaS regression in (6), I consider a more general model, which allows for nonlinearity in the coefficients of all terms. The linear model in (2) may be written as very simply as

$$\triangle y_{t+1} = p_t' \alpha + \varepsilon_{t+1}, \tag{7}$$

where $p_t = (p_{0t}', p_{1t}')'$. I assume that the DGP is linear, so that the minimum MSFE forecast of $\triangle y_{t+1}$ is given by $\mathbf{E}[\triangle y_{t+1}|\mathcal{F}_t] = p_t'\alpha + \mathbf{E}[\varepsilon_{t+1}|\mathcal{F}_t]$. I make a realistic allowance for the error sequence to be correlated serially and with both the I(0) and I(1) regressors, so that $\mathbf{E}[\varepsilon_{t+1}|\mathcal{F}_t] \neq 0$ in general.

As described above, forecasts are made using a general nonlinear model

$$\triangle y_{t+1} = p_t' g(\theta) + \eta_{t+1}. \tag{8}$$

[5]See Marcellino (1999) and Haug (2002) for discussions of such tests when the low-frequency series is temporally aggregated.

I define the space $\mathcal{G} \subseteq \mathbb{R}^n$ such that $g : \Theta \mapsto \mathcal{G}$, where n is the number of regressors, as above.

I assume that

[N1] there exists $\theta \in \Theta$ such that $g(\theta) = \alpha$,

so that the nonlinear model nests the linear DGP. Otherwise, there is a latent term in the residual stemming from the nonlinear approximation. Correlation of this term with the regressors may generate bias in the forecasts, and it is not obvious that such bias is sufficiently small to be offset by the efficiency gain from the parsimonious specification. Although my proposed test can provide evidence for or against the nonlinear specification by allowing a general linear alternative, like Andreou *et al.* (2010a), I do not address this difficulty in the asymptotic analysis of the estimators.

I consider (8) to be the model estimated henceforth. The CoMiDaS regression of interest in (6) is a special – but important – case, with $g(\theta) = (g_0'(\theta), \theta_1')'$.

2.2. Minimum MSFE Parameter Vector and NLS Estimation

The MSFE using the model in (8) may be written as

$$Q(\theta) = \mathbf{E}(\mathbf{E}[\varepsilon_{t+1} - p_t'(g(\theta) - \alpha)|\mathcal{F}_t])^2$$

by substituting (7). The first-order condition is

$$0 = \frac{\partial g'(\theta)}{\partial \theta}(M(g(\theta) - \alpha) - N),$$

where $M \equiv \mathbf{E}p_t p_t'$ and $N \equiv \mathbf{E}p_t \varepsilon_{t+1}$. M and N may be partitioned as

$$M = \begin{bmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{bmatrix} = \begin{bmatrix} \Sigma_{pp} & \Delta_{p1} + o(1) \\ \Delta_{p1}' + o(1) & T\Omega_{11} + o(T) \end{bmatrix}, \qquad \text{and}$$

$$N = \begin{bmatrix} N_0 \\ N_1 \end{bmatrix} = \begin{bmatrix} \sigma_{p\varepsilon} \\ \delta_{\varepsilon 1}' + o(1) \end{bmatrix},$$

where $\Sigma_{pp} \equiv \mathbf{E}p_{0t}p_{0t}'$, $\Delta_{p1} \equiv \sum_{k=0}^{\infty} \mathbf{E}p_{0t}\triangle p_{1,t-k}'$, $\delta_{\varepsilon 1} \equiv \sum_{k=0}^{\infty} \mathbf{E}\varepsilon_{t+1}\triangle p_{1,t-k}'$, and $\Omega_{11} = \sum_{k=-\infty}^{\infty} \mathbf{E}\triangle p_{1t}\triangle p_{1,t-k}'$.[6] The remainder terms come from $\sum_{k=T+1}^{\infty} \mathbf{E}p_{0t}\triangle p_{1,t-k}'$, etc.

I assume that

[A1] $M_{11} - M_{10}M_{00}^{-1}M_{10}$ is nonsingular for $n_1 > 0$ and for any T, and that

[A2] Σ_{pp} is nonsingular,

[6] With I(1) series, the moment matrix M is a function of T. However, I suppress the argument to reserve the notation M_T for sample moment matrices below.

which are sufficient to ensure the invertibility of M. As $T \to \infty$, Assumption [A1] requires only that Ω_{11} is nonsingular, since $T^{-1}(M_{11} - M_{10}M_{00}^{-1}M_{10}) = \Omega_{11} + o_p(1)$. I maintain [A1] to ensure that M is invertible for any sample size. Assumption [A2] does not place any unreasonable restrictions on the high-frequency regressors. If the increments of these regressors are perfectly correlated, as in the case with interpolated series, then there is no information gain from including the high-frequency increments. In that case, the increments may be dropped, retaining only the levels (x_t) observed at the lower frequency.

Under the additional assumptions that

[N2] $\partial g_i(\theta)/\partial\theta$ exists and is bounded in a neighborhood of $\theta_{\min}$ for $i = 1, \ldots, n$ and that

[N3] $g(\theta_{\min}) \in \text{int}(\mathcal{G})$,

a solution to the minimization problem is given by $g(\theta_{\min}) - \alpha = M^{-1}N$. The vector $\theta_{\min}$ is the minimum MSFE parameter vector of the nonlinear model in (8). If the error is a martingale difference sequence (an mds) with respect to $(\mathcal{F}_t)$, then $N = 0$ and $p_t'\alpha$ gives the optimal forecast. Deviations from $N = 0$ may occur because of correlation between the error and regressors.

NLS is a natural estimator for forecasting with a nonlinear model. The NLS objective function may be written as

$$Q_T(\theta) = \frac{1}{2}\sum\nolimits_t (\varepsilon_{t+1} - p_t'(g(\theta) - \alpha))^2,$$

which has a first-order condition of

$$0 = T\frac{\partial g'(\theta)}{\partial\theta}(M_T(g(\theta) - \alpha) - N_T),$$

where $M_T \equiv T^{-1}\sum_t p_t p_t'$ and $N_T \equiv T^{-1}\sum_t p_t \varepsilon_{t+1}$. A solution is given by $g(\hat{\theta}_{NLS}) - \alpha = M_T^{-1}N_T$, similarly to $g(\theta_{\min})$.

The difference $g(\hat{\theta}_{NLS}) - g(\theta_{\min})$ may be written as $M_T^{-1}P_T$, where $P_T \equiv (N_T - N) - (M_T - M)M^{-1}N$. This difference reveals the extent to which the NLS estimator minimizes MSFE, which is the aim of the next section.

3. Asymptotic Properties of the NLS Estimator

In order to carefully analyze the NLS estimator of θ in (8), I introduce additional notation and rely on additional assumptions. Let $b_t \equiv (\varepsilon_{t+1}, p_{0t}', \triangle p_{1t}')'$ denote the $n+1$ stationary components of the DGP in (7), where (p_{0t}) and (p_{1t}) are the I(0) and I(1) regressors respectively. Defining $b_{0t} \equiv (\varepsilon_{t+1}, p_{0t}')'$ and $b_{1t} \equiv \triangle p_{1t}$ allows a useful alternative partition $b_t = (b_{0t}', b_{1t}')'$. In the ensuing discussion, the partition $v_t = (v_{0t}', v_{1t}')'$ is conformable with $(b_{0t}', b_{1t}')'$.

I assume that

[A3] $b_t = \sum_{s=0}^{\infty} \Psi_s v_{t-s}$ where $\sum_{k=0}^{\infty} s \left\| \Psi_s \right\| < \infty$, and that

[A4] (v_t) is iid with finite second and fourth moment matrices, and with zero third moment matrix.

These assumptions allow multivariate versions of the LLN, CLT, and IP of Phillips and Solo (1992). Assumption [A3] is quite general, allowing correlation serially and across series, which applies to both the error, the I(0) regressors, and first-differences of the I(1) regressors. The cases of iid and I(0) autoregressive regressors considered by Andreou *et al.* (2010a) is a special case, up to the assumption on the third moment. The symmetry condition is not essential, but makes the distributions below more tractable. As it turns out, when the error is iid, as Andreou *et al.* (2010a) assume, asymmetry may be allowed.[7]

The main asymptotic results below should hold under more general assumptions, such as those for weakly dependent heterogeneous processes considered by Davidson (1994), *inter alia*. Such a generalization would allow for ARCH in both the regressors, as the third case considered by Andreou *et al.* (2010a), and in the error term, for example.

The analogy between the MiDaS regressions in (4)-(6) and that of the more general model in (8) requires some additional explanation in light of Assumption [A3]. The matrices Ψ_s need not be square, in order to allow for the possibility of common innovations. In particular, $(\triangle x_t)$ and $(\triangle^{(1/m)} x_t^{(m)})$ have common high-frequency innovations. For example, letting $v_t \equiv (v_t^{(2)}, v_{t-1/2}^{(2)})$, $(\triangle x_t)$ has innovations of (v_t), while $(\triangle^{(1/2)} x_t^{(2)})$ has innovations of $(v_t^{(2)})$.

An invariance principle holds under Assumptions [A3] and [A4], such that $T^{-1/2} \sum_{t=1}^{[Tr]} b_t \equiv B_n(r) \to B(r)$ with $B \equiv BM(\Omega)$. $B = (B_0', B_1')'$ may be partitioned as above. The variance $\Omega = \Delta' + \Sigma + \Delta$ may be partitioned as

$$\Omega = \begin{bmatrix} \omega_\varepsilon^2 & \omega_{\varepsilon p} & \omega_{\varepsilon 1} \\ \omega_{p\varepsilon} & \Omega_{pp} & \Omega_{p1} \\ \omega_{1\varepsilon} & \Omega_{1p} & \Omega_{11} \end{bmatrix} \quad \text{or} \quad \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{01} \\ \Omega_{10} & \Omega_{11} \end{bmatrix}$$

similarly to (b_t). In other words, $\Omega_{10} = (\omega_{1\varepsilon}, \Omega_{1p})$, etc. Subscripts on the components Σ and Δ denote the same partition, and the variances and covariances Σ_{pp}, Δ_{p1}, Ω_{11}, $\sigma_{p\varepsilon}$, and $\delta_{\varepsilon 1}$ are the same as those introduced in the previous section. As noted above, $\Omega_{11} > 0$ due to Assumption [A1], but this assumption does not rule out cointegration of $(y_t, w_{1t}', x_{1t}')'$ in (5), because the cointegrating combinations are contained in (p_{0t}).

All of the asymptotics in this paper are low-frequency asymptotics, but these may be derived from a high-frequency DGP. Chambers (2003) derived low-frequency asymptotics from a DGP defined by stocks and flows in continuous time using particular aggregations schemes. Miller (2011) derived low-frequency asymptotics from a high-frequency DGP using a general aggregation scheme that may include the MiDaS lag polynomials considered here.

3.1. Consistency

Of primary concern is the consistency of the coefficient vector $g(\hat{\theta}_{NLS})$ to $g(\theta_{\min})$, rather than that of $\hat{\theta}_{NLS}$ to the deep parameter vector $\theta_{\min}$, since the coefficient vector may be compared directly with that of the linear DGP.

[7]See the last part of the proof of Lemma A1.

The following lemma establishes consistency of the coefficient estimates.

Theorem 1. *Under Assumptions [A1]-[A4] and [N1]-[N3], $g(\hat{\theta}_{NLS}) \to_p g(\theta_{\min})$ as $T \to \infty$. Moreover, defining the partition $g(\theta) = (g_0'(\theta), g_1'(\theta))'$ such that $g_0(\theta)$ and $g_1(\theta)$ correspond to the I(0) and I(1) regressors respectively, $g_1(\hat{\theta}_{NLS}) = g_1(\theta_{\min}) + O_p(T^{-1})$.*

The result is more robust than the consistency result of Andreou *et al.* (2010a) in the sense that both I(0) and I(1) regressors are allowed and in the sense that consistency to the minimum MSFE parameter vector does *not* require $(\varepsilon_{t+1}, \mathcal{F}_t)$ to be an mds. The error may be correlated both serially and with the regressors. The second part of the result establishes superconsistency for the coefficients corresponding to the I(1) regressors. Of course, $g_1(\theta_{\min}) = \theta_{1,\min}$ in the CoMiDaS regression in (6).

The additional assumption that

[N3'] $g^{-1}(\theta)$ is continuous at $\theta_{\min}$ and $\theta_{\min} \in \mathrm{int}(\Theta)$

allows consistency of the NLS estimator of the deep parameter vector θ.

Corollary 2. *Under Assumptions [A1]-[A4], [N1]-[N2], and [N3'], $\hat{\theta}_{NLS} \to_p \theta_{\min}$ as $T \to \infty$.*

Note that since Θ is a subset of the real line that need not be compact or closed, a limiting solution of $\pm\infty$ is still an interior solution as long as the derivative of $g(\theta)$ is zero at that limit. In other words, $g(\theta)$ may have a horizontal asymptote. This allowance is important, since the exponential Almon lag may best approximate particular weighting schemes as γ_1 or γ_2 approach $\pm\infty$. For example, end-of-period sampling is approximated as $\gamma_2 \to -\infty$.

3.2. Asymptotic Distributions

For the purpose of forecasting, the limiting distribution of $g(\hat{\theta}_{NLS}) - g(\theta_{\min})$ is sufficient to establish the asymptotic difference between an actual forecast and the minimum MSFE forecast. To the extent that inference about the deep parameters is desirable, the limiting distribution of $\hat{\theta}_{NLS} - \theta_{\min}$ is also useful.

When first and second derivatives are taken with respect to g rather than θ, the gradient and Hessian reduce to $J_T(g) \equiv T\left(M_T(g(\theta) - \alpha) - N_T\right)$ and $H_T \equiv TM_T$. Note that when evaluated at $g(\theta_{\min})$, $J_T = -TP_T$. Define $n \times n$ diagonal the normalization matrix $\nu_T \equiv diag(T^{1/2}I_{n_0}, TI_{n_1})$. Further, define

$$H \equiv \begin{bmatrix} \Sigma_{pp} & 0 \\ 0 & \int B_1 B_1' \end{bmatrix} \quad \text{and} \quad J \equiv \begin{bmatrix} (E' \otimes \kappa')\mathbf{N}(0, \Xi) \\ \int B_1 dB_0' \kappa - (\int B_1 B_1' \Omega_{11}^{-1} - I)\zeta \end{bmatrix},$$

where $\Xi \equiv \mathbf{E}[b_{0t}b_{0t}' \otimes b_{0t}b_{0t}'] - \mathbf{E}[b_{0t} \otimes b_{0t}]\mathbf{E}[b_{0t}' \otimes b_{0t}']$, $\kappa \equiv (1, -N_0'\Sigma_{pp}^{-1})'$ is an (n_0+1)-vector, $\zeta \equiv (N_1' - N_0'\Sigma_{pp}^{-1}\Delta_{p1})'$ is an n_1-vector, and E' is the unitary matrix that selects all but the first row of the following matrix of $n_0 + 1$ rows.

Using these definitions, the following theorem holds.

Theorem 3. *Under Assumptions [A1]-[A4] and [N1]-[N3],*

$$\nu_T(g(\hat{\theta}_{NLS}) - g(\theta_{\min})) \to_d H^{-1}J$$

as $T \to \infty$.

The asymptotic rates are as expected, but this $n \times 1$ distribution is quite nonstandard when (ε_{t+1}) is correlated serially and with the regressors. Normality holds for the coefficients of the I(0) regressors, however the variance is exacerbated by deviations from $\mathbf{E}p_{0t}\varepsilon_{t+1} = 0$.

In the special case in which (ε_{t+1}) is uncorrelated with the regressors, $N = 0$ and the $n_0 \times (n_0+1)^2$ matrix $(E' \otimes \kappa')$ becomes a matrix that selects rows $(n_0+1)j+1$, for $j = 1, \ldots, n_0$, of the subsequent matrix of $(n_0+1)^2$ rows. Since Ξ is an $(n_0+1)^2 \times (n_0+1)^2$ matrix of Kronecker products, some algebra shows that the rows and columns of Ξ selected are $\mathbf{E}[\varepsilon_{t+1}^2 p_{0t}p_{0t}'] - \mathbf{E}[\varepsilon_{t+1}p_{0t}]\mathbf{E}[\varepsilon_{t+1}p_{0t}']$, which simplifies to $\sigma_\varepsilon^2\Sigma_{pp}$ under the uncorrelatedness assumption.

If, moreover, the model contains no I(1) regressors, the limiting distribution $H^{-1}J$ reduces to $\sigma_\varepsilon\Sigma_{pp}^{-1/2}\mathbf{N}(0, I)$ when $N = 0$. In other words, when all of the regressors are stationary, the error term is ideal, and the MiDaS specification nests the underlying DGP, the NLS estimator of the coefficients – not of the deep parameters, which may enter nonlinearly – has the same asymptotic distribution as the LS estimator of the linear model.

If, on the other hand, the model contains no I(0) regressors, the limiting distribution $H^{-1}J$ reduces to $\sigma_\varepsilon(\int B_1B_1')^{-1/2}\mathbf{N}(0, I)$ when $N = 0$. This mixed normal distribution validates hypothesis testing using standard critical values.

Since CoMiDaS models *may* contain I(1) regressors but almost always contain I(0) regressors, the limiting distribution is generally $\sigma_\varepsilon H^{-1/2}\mathbf{N}(0, I)$, a vector containing both normal and mixed normal subvectors. These results are sensible in this special case, but Theorem 3 is much more broadly applicable.

Obtaining the exact asymptotic distribution of the NLS estimator $\hat{\theta}_{NLS}$ of the deep parameters is more involved, because the derivative structure is much more complicated.

I assume that

[N2'] the third derivative matrix of $g_i(\theta)$ exists for $i = 1, \ldots, n$, and $\partial g_i(\theta)/\partial\theta\partial\theta'$ is bounded in a neighborhood of $\theta_{\min}$,

and define

$$J_T^*(\theta) \equiv T(\partial g'(\theta)/\partial\theta)(M_T(g(\theta) - \alpha) - N_T), \quad \text{and}$$
$$H_T^*(\theta) \equiv T(\partial g'(\theta)/\partial\theta)M_T(\partial g(\theta)/\partial\theta') + T((M_T(g(\theta) - \alpha) - N_T)' \otimes I)(\partial vec\left(\partial g'(\theta)/\partial\theta\right)/\partial\theta')$$

to be the gradient vector and Hessian matrix with respect to θ. Again, these simplify at $\theta_{\min}$, since $M_T(g(\theta_{\min}) - \alpha) - N_T = P_T$.

I do not assume that the parameters involved in the coefficients on the I(0) regressors are distinct from those involved in the coefficients on the I(1) regressors. In particular, the coefficient β_1 in (5) is common to both. Recalling that $\theta \subseteq \mathbb{R}^g$ and $\mathcal{G} \subseteq \mathbb{R}^n$, $\partial g'(\theta)/\partial\theta$ is a $g \times n$ matrix. The $n \times n$ matrix M_T may be partitioned into four blocks: a block with products of only I(0) regressors, a block with products of only I(1) regressors, and two off-diagonal blocks with only cross-products of I(0) and I(1) regressors.

Sort and partition $\theta = (\theta_0', \theta_{10}', \theta_{11}')'$, so that θ_0 are g_0 parameters only in terms with I(0) regressors – e.g., γ in the weight functions above, θ_{10} are parameters in both terms with I(0) regressors and terms with I(1) regressors – e.g., β_1 in (5), and θ_{11} are parameters only in terms with I(1) regressors – e.g., ρ^* in (5). Let $g_1 \equiv g - g_0$, so that g_1 denotes the number of parameters in all terms with I(1) regressors: the rows of $(\theta_{10}', \theta_{11}')'$.

The $g \times g$ matrix $(\partial g'(\theta)/\partial\theta)M_T(\partial g(\theta)/\partial\theta')$ may be partitioned into four blocks: a $g_0 \times g_0$ block with terms consisting of products of only I(0) regressors, two off-diagonal blocks with terms consisting either of products of only I(0) regressors or cross-products of I(0) and I(1) regressors, but no products of only I(1) regressors, and a $g_1 \times g_1$ block with terms consisting of products of only I(1) regressors, products of only I(0) regressors, and cross-products.

Conformably with these blocks, define

$$G(\theta) \equiv \begin{bmatrix} \partial g_0(\theta)/\partial\theta_0' & 0 \\ 0 & \partial g_1(\theta)/\partial\theta_1' \end{bmatrix},$$

and define the $g \times g$ diagonal normalization matrix $\nu_T^* \equiv diag(T^{1/2}I_{g_0}, TI_{g_1})$. Finally, define $J^*(\theta) \equiv G'(\theta)J$ and $H^*(\theta) \equiv G'(\theta)HG(\theta)$. The following corollary holds using these definitions.

Corollary 4. *Under Assumptions [A1]-[A4], [N1], and [N2']-[N3'],*

$$\nu_T^*(\hat\theta_{NLS} - \theta_{\min}) \to_d H^*(\theta_{\min})^{-1}J^*(\theta_{\min})$$

as $T \to \infty$.

This distribution is g-dimensional, in contrast to the much less parsimonious n-dimensional distribution of the preceding theorem.

Again, it is instructive to consider the special case of $N = 0$, in which (ε_{t+1}) is uncorrelated with the regressors. The distribution simplifies to $\sigma_\varepsilon H^*(\theta_{\min})^{-1/2}\mathbf{N}(0, I)$ in the same way as above. And again, this distribution is a vector of normal and mixed normal subvectors. In the case of only I(0) regressors, this may be written as

$$\sigma_\varepsilon((\partial g_0'(\theta)/\partial\theta_0|_{\theta=\theta_{\min}})\Sigma_{pp}(\partial g(\theta_0)/\partial\theta_0'|_{\theta=\theta_{\min}}))^{-1/2}\mathbf{N}(0, I)$$

which is analogous to results in Section 3.2 of Andreou *et al.* (2010a).[8,9] Corollary 4 is more widely applicable than their results, however.

[8] The variance of $\hat\beta_{NLS}$ alone may be further isolated by conditioning on the remaining estimators in $\hat\theta_{NLS}$, as in line (3.4) of their paper.

[9] Even this special case generalizes the results Andreou *et al.* (2010a) in the sense that serial correlation of the error is still allowed, as long as the regressors are strictly exogeneous.

4. Testing for MiDaS and CoMiDaS

Mixed-frequency cointegrating regressions do not require a nonlinear MiDaS specification when the regressand has been average sampled. Chambers (2003) and Miller (2011) showed that the regressors should also be average sampled for maximum efficiency in that case, unless any of the series exhibit substantial seasonality, which may introduce inefficiency in estimating the long-run relationship. This recommendation holds regardless of whether or not the series are stocks or flows. When the regressand has been sampled differently – e.g., selectively sampled – the most efficient sampling method for the regressors when $N = 0$ is to *match* that of the regressand (Miller, 2011). However, Miller (2011) showed that efficiency gains are possible by exploiting serial correlation in the error term when $N \neq 0$ and the regressand aggregation scheme is known. In this case, or with unknown regressand aggregation scheme, or with I(0) series, and especially when the linear model is over-parameterized or infeasible, a (Co)MiDaS specification is sensible.

Andreou *et al.* (2010a) propose several tests of particular weighting schemes such as those mentioned above against alternatives of general nonlinear MiDaS specifications under the maintained hypotheses of no error correlation and that the MiDaS specification nests both the null and alternative. Using my notation, $N = 0$ under both their null and alternative. Because of the (mixed) normality in the CoMiDaS framework when $N = 0$, their testing strategy should also be valid for CoMiDaS models with strictly exogenous I(1) regressors.

The CoMiDaS alternative is useful to reject specific nested nulls, such as average sampling. However, a CoMiDaS null is more useful if the validity of the nonlinear specification is questionable. The difficulty in testing for CoMiDaS against a general linear alternative lies in infeasibility of estimating the alternative or of calculating the score when m is large. Unfortunately, this difficulty rules out traditional Wald, likelihood ratio, and Lagrange multiplier tests, such as the LR test proposed by Godfrey and Poskitt (1975) for a non-exponential Almon lag structure.

4.1. Restricted Minimum MSFE Parameter Vector and NLS Estimation

For the subsequent test, a restricted minimum MSFE parameter vector and a restricted nonlinear estimator are useful. In this case, Assumption [N1] is violated, except under the null that the restricted and unrestricted estimators are equal. Let $\theta_* \equiv \min_{\theta \in \Theta} \|g(\theta) - \alpha\|$ and $\tau_* \equiv \alpha - g(\theta_*)$. The norm of the vector τ_* reflects a lower bound on the difference $g(\theta) - \alpha$, based on both the functional form g used and the underlying coefficients, but not on the choice parameter vector θ. I assume that

[N1'] τ_* exists and is unique

for identification. Note that in any CoMiDaS regression, $\tau_{*,i} = 0$ for all i corresponding to the I(1) regressors in (5), since CoMiDaS regressions are linear in these parameters.

The null that a particular MiDaS specification nests the DGP is equivalent to $\|\tau_*\| = 0$, in which case Assumption [N1] holds and all of the above asymptotics are valid. The alternative that the MiDaS specification does not nest the DGP is equivalent to $\|\tau_*\| > 0$.

Of course, a rejection of the null may be a rejection of a particular lag structure, rather than a rejection of nonlinearity in general.

A Lagrangian objective function for the minimum MSFE vector,

$$Q(\theta) = \mathbf{E}(\mathbf{E}[\varepsilon_{t+1} - p_t'(g(\theta) - \alpha)|\mathcal{F}_t])^2 + 2\lambda'(g(\theta) - \alpha - \tau_*),$$

assigns a vector of Lagrange multipliers λ to the vector of constraints that $|g(\theta) - \alpha| > |\tau_*|$.[10]

The first-order conditions are

$$\begin{aligned} 0 &= (\partial g'(\theta)/\partial\theta)(M(g(\theta) - \alpha) - N + \lambda) \qquad \text{and} \\ 0 &= g(\theta) - \alpha - \tau_*, \end{aligned}$$

so that

$$\tau_* = g(\theta_{\min}) - \alpha = M^{-1}(N - \lambda)$$

at the minimum. Under the null that $\|\tau_*\| = 0$, $\lambda = N$, which is zero only if $N = 0$.

The NLS estimator under the same restriction has a Lagrangian given by

$$Q_T(\theta) = \frac{1}{2}\sum\nolimits_t (\varepsilon_{t+1} - p_t'(g(\theta) - \alpha))^2 + T\lambda_T'(g(\theta) - \alpha - \tau_*).$$

which has similar first-order conditions, so that

$$\tau_* = g(\hat{\theta}_{NLS}) - \alpha = M_T^{-1}(N_T - \lambda_T) \tag{9}$$

at the minimum.

4.2. A MiDaS Variable Addition Test

I propose a novel strategy to test the MiDaS null, based on possible neglected nonlinearity in the estimated series of residuals ($\hat{\eta}_{t+1}$). Consider a simplified (Co)MiDaS regression,

$$y_{t+1} = p_t' w(\gamma)\beta + \eta_{t+1} \tag{10}$$

with DGP given by

$$y_{t+1} = p_t' w\beta + \varepsilon_{t+1} \tag{11}$$

where $w(\gamma)$ is a weight vector and β is univariate – i.e., a single high-frequency series drives m regressors. Additional regressors in (3) add unnecessary expositional complexity for testing a single polynomial. The testing procedure would be exactly the same in that case.

The vector p_t may be viewed either as a vector of high-frequency regressors, as in (3), or as a single high-frequency regressor followed by high-frequency differences, as in (4). Note that $w(\gamma)\beta$ corresponds to $g(\theta)$ in (8) and elsewhere above, while $w\beta$ corresponds to α in (7) and elsewhere above. The nonlinear model nests the linear model *only* under the null.

The NLS residual series is

$$\hat{\eta}_{t+1} = \varepsilon_{t+1} - p_t'(w(\hat{\gamma})\hat{\beta} - w\beta), \tag{12}$$

[10] To form the Lagrangian, let the sign of each element λ_i of λ be positive if $g_i(\theta) - \alpha_i > \tau_{*,i}$ and negative if $g_i(\theta) - \alpha_i < \tau_{*,i}$, so that the second term of the objective function is non-negative for minimization.

where the NLS subscript is omitted from the estimators for brevity. The second term of (12) differs from zero asymptotically for two possible reasons. Either $N \neq 0$, so that NLS does not consistently estimate $w\beta$, which is not $w_{\min}\beta_{\min}$ in that case, or else the null is not true, in which case $w(\gamma)$ does not nest w and inconsistency results from neglected nonlinearity. If only the first reason is true – the null still holds – then the deviation is exactly offset by the correlation of (ε_{t+1}) with (p_t), so that $\sum_t \hat{\eta}_{t+1} p_t$ equals zero up to numerical approximation error.

If the second is true – the null does not hold – then the deviation of $(w(\hat{\gamma})\hat{\beta} - w\beta)$ from zero may be picked up by an ancillary regression. Regressing $(\hat{\eta}_{t+1})$ onto (p_t) is not parsimonious and may be infeasible. Regressing $(\hat{\eta}_{t+1})$ onto $(p_t' w(\hat{\gamma}))$ is feasible, but suffers from an endogeneity bias, since $(p_t' w(\hat{\gamma}))$ is correlated with (ε_{t+1}). A solution is offered by a Wu-type variable addition test (Wu, 1973). Regressing $(\hat{\eta}_{t+1})$ onto $(p_t' w(\hat{\gamma}))$ should capture all of the correlation between (ε_{t+1}) and (p_t), as well as estimation error in $(w(\hat{\gamma})\hat{\beta} - w\beta)$ from that correlation. (ε_{t+1}) should not be correlated with any other linear combination of (p_t). The remaining error in this ancillary regression stems from neglected nonlinearity under the alternative. So, controlling for $(p_t' w(\hat{\gamma}))$, $(\hat{\eta}_{t+1})$ should not be correlated with any other linear combination of (p_t) under the null, but should be correlated under the alternative.

Let $\hat{p}_t = p_t' w(\hat{\gamma})$ denote the estimated linear combination of high-frequency regressors using NLS and let W denote an $m \times q$ matrix of q other arbitrary linear combinations. The proposed test is based on the regression

$$\hat{\eta}_{t+1} = \hat{p}_t \varsigma_0 + p_t' W \varsigma_1 + e_t. \tag{13}$$

A null of $\varsigma_1 = 0$ coincides with a MiDaS null – i.e., that (10) nests (11). Letting F denote an F-test with this null, qF or an alternative test of $\varsigma_1 = 0$ may be used to test for MiDaS.

The fitted series $(\hat{p}_t)$ is included to ensure that estimates of ς_1 pick up only variations in $(p_t' W)$ orthogonal to $(\hat{p}_t)$. Such variations increase the power of the test. To ensure sufficient power, one may either choose multiple combinations in W, or run a preliminary regression of $(\hat{p}_t)$ onto $(p_t' W)$. A *lower* R^2 in this preliminary regression is desirable to increase power. However, keeping in mind that both the regressand and regressors of such a preliminary regression are linear combinations of the same high-frequency series, an R^2 near unity should be expected.

In the sense of latent omitted variables, my proposed test shares some similarities with the Wu-type test proposed by Andreou *et al.* (2010a). However, it differs substantially. Most importantly, their test is designed with a null of a specific aggregation scheme nested by MiDaS against an alternative of a more general MiDaS structure. They do not consider a DGP more general than a MiDaS regression. Second, they use m high-frequency instruments rather than linear combinations of the high-frequency regressors. Their IV test is infeasible when m is large relative to the sample size – precisely when the parsimonious MiDaS specification is most useful.

The following proposition cements the easily conjectured asymptotic χ^2 distribution of the proposed test statistic in most cases.

Proposition 5. *Let $p_t = (x_t, \triangle^{(1/m)} x_t^{(m)}, ..., \triangle^{(1/m)} x_{t-(m-1)/m}^{(m)})'$ in (10) and (11) and $q < m$. Under Assumptions [A1]-[A4], [N1'], and [N2]-[N3], the limiting distribution of the (Co)MiDaS variable addition test statistic qF coincides with a χ_q^2 distribution when either*

[a] *(x_t) is I(0), or*

[b] *(x_t) is I(1) and strictly exogenous and $(\varepsilon_{t+1}, \mathcal{F}_t)$ is an mds,*

as $T \to \infty$.

Note that the asymptotic distribution of the test is robust to correlation of the regressors and error term in the I(0) case, but not the I(1) case.

In the I(1) case, the error must be an mds sequence uncorrelated with the regressor at all lags. If the mds assumption is violated, it may be possible to construct a Wald test with a long-run variance estimator instead of the contemporaneous variance estimator of the traditional F-test, along the lines discussed by Park and Phillips (1988). If the exogeneity assumption is violated, it may be possible to correct for the resulting non-normality along the lines of Phillips and Hansen (1990) or Park (1992). However, variance estimation in these cases would be inconsistent due to correlation of the error with the I(0) regressors, further complicating the test and its limit.

Fortunately, the violation occurs in only one of the $(q + 1)$ elements of the regressor vector. When the number of variables added in the variable addition test is large, the non-normality of the single element may be diluted by the normality of the remaining q elements, so that the chi-squared distribution is a rough approximation to the actual distribution, which would involve stochastic integrals and nuisance parameters similar to the limit in Theorem 3.

The model and DGP in (10) and (11) are simple, but the principles of the test extend easily. For more complicated (Co)MiDaS models such as that in (5), the MiDaS specifications for each high-frequency regressor may be tested jointly or separately, depending on the desired null. The exact testing strategy depends on the parametric modeling assumptions underlying the nonlinear weight specification. The limiting distribution may be generalized from Proposition 5.

4.3. Small-Sample Performance of the Test

I perform simulations of the DGP in (11) to evaluate size and power properties of the proposed test using several nulls and alternatives. I set $m = 12$, as would be the case with an annual regressand and monthly regressor, and I consider $T = 25, 50, 100$. If T represents years in the sample, then mT represents months in the sample. All of the asymptotics above are as $T \to \infty$, so the sample sizes here are quite small but realistic for macroeconomic series.

I consider weights generated by 12 DGPs consisting of 6 nulls and 6 alternatives. The null models nest the MiDaS specification in (10), using an exponential Almon lag with γ equal to $(0, 0)$, $(-5, -5)$, $(1, 1)$, $(-0.5, 0.04)$, $(0.5, -0.04)$, and $(0.005, 0.02)$, labeled $H_0(j)$ with $j = 1, ..., 6$. The first of these is simply a flat aggregation scheme characteristic of

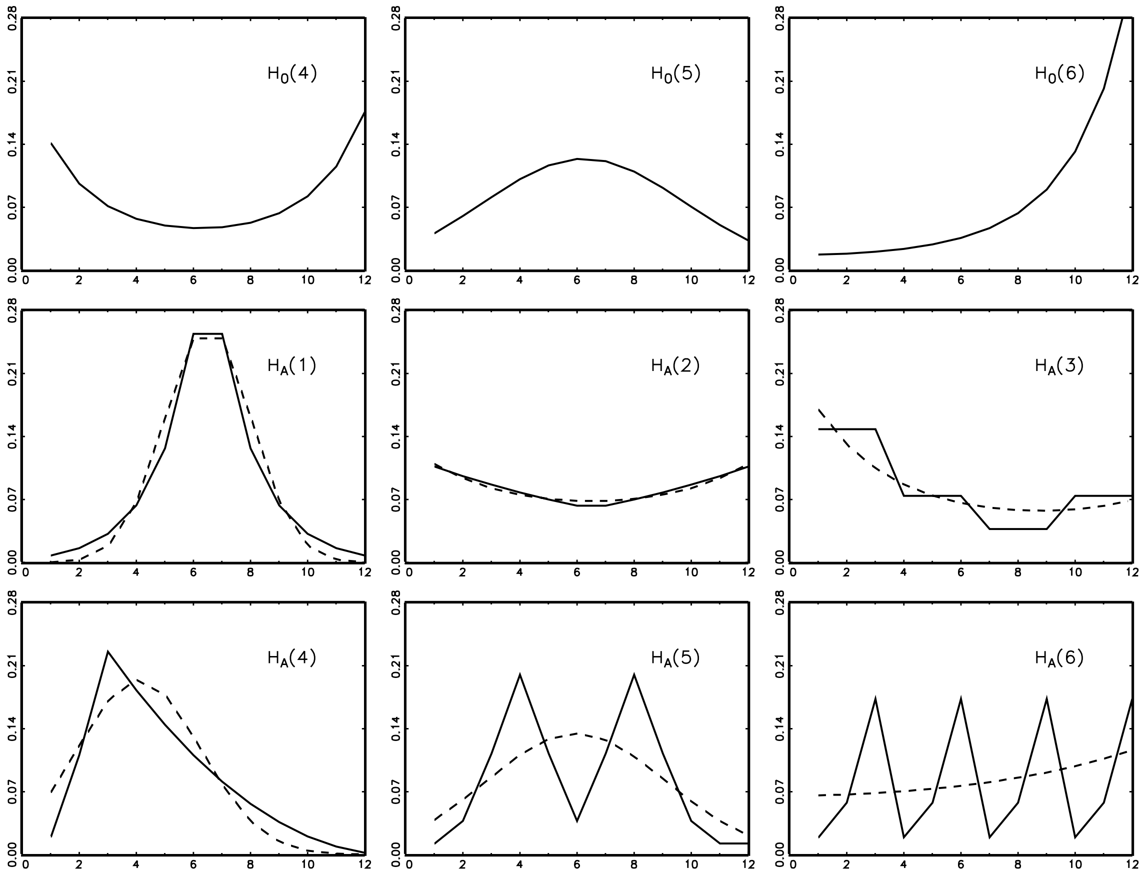

Figure 1: Three null weighting schemes and six alternative weighting schemes. Solid lines represent the DGP, while dashed lines represent the closest approximation to the alternative DGPs using an exponential Almon lag.

average sampling. The second two assign unit weights to the first high-frequency regressor and the last high-frequency regressor, respectively, and zero to the remaining high-frequency regressors, characteristic of selective sampling.

The last three are illustrated in the top panels of Figure 1. Specifically, $H_0(4)$ assigns more weight to the high-frequency regressors near the beginning and end of the low-frequency interval, $H_0(5)$ assigns more to those in the middle, and $H_0(6)$ assigns a gently increasing weight structure moving from the end to the beginning.

The bottom six panels of Figure 1 show six alternatives not nested by the exponential Almon lag. These panels also show the best fit curve, in the sense of $\min_{\theta\in\Theta}\|g(\theta)-\alpha\|$, using the exponential Almon function. A higher-order exponential Almon lag could better approximate some of these alternatives. Because the point of this exercise is to demonstrate the power of the test, only the second-order exponential Almon lag described in Section 2 is employed. Table 1 shows the weighting schemes for the six alternatives, labeled $H_A(j)$ with

$j = 1, ..., 6$. The table shows raw weights w^*, with actual weights set to $w_s = w_s^* / \sum_{i=1}^{12} w_i^*$.

Table 1: Alternative Weighting Schemes

	w_1^*	w_2^*	w_3^*	w_4^*	w_5^*	w_6^*	w_7^*	w_8^*	w_9^*	w_{10}^*	w_{11}^*	w_{12}^*
$H_A(1)$	1	2	4	8	16	32	32	16	8	4	2	1
$H_A(2)$	0.9^0	0.9^1	0.9^2	0.9^3	0.9^4	0.9^5	0.9^5	0.9^4	0.9^3	0.9^2	0.9^1	0.9^0
$H_A(3)$	1	1	1	$1/2$	$1/2$	$1/2$	$1/4$	$1/4$	$1/4$	$1/2$	$1/2$	$1/2$
$H_A(4)$	3^2	7^2	10^2	9^2	8^2	7^2	6^2	5^2	4^2	3^2	2^2	1^2
$H_A(5)$	3^0	3^1	3^2	4^2	3^2	3^1	3^2	4^2	3^2	3^1	3^0	3^0
$H_A(6)$	3^0	3^1	3^2	3^0	3^1	3^2	3^0	3^1	3^2	3^0	3^1	3^2

$H_A(1)$ is intuitively similar to $H_0(5)$, but with a more precipitous and less smooth peak in the middle. $H_A(2)$ is intuitively similar to $H_0(4)$ but flatter. Note that the exponential Almon function almost nests $H_A(2)$. $H_A(3)$ shows a quarterly pattern, if $m = 12$ is interpreted as months, with more weight given to the last quarter (smallest index). This weighting scheme might be reasonable for seasonal data in which the relationship modeled holds better during, say, a high-volume transaction season. $H_A(4)$ shows a sharp seasonal spike occurring at a time other than the ends or middle of the low-frequency period, and $H_A(5)$ shows a bimodal weighting scheme. Finally, for a shorter cycle within each quarter, say, $H_A(6)$ shows a repeating quarterly pattern.

These various null and alternative weighting schemes serve two purposes. I use each to illustrate size and power of the MiDaS IV test. I use the rest as instruments. That is, the q columns of W are the 11 linear combinations other than the one used in the DGP. Omitting the one used in the DGP rules out collinearity with $(\hat{z}_t)$ – especially for the null models.

Letting $u_t^{(m)} = (\varepsilon_{t+1}^{(m)}, x_t^{(m)})$ and

$$A = \begin{bmatrix} 0 & 0 \\ 0 & \varrho \end{bmatrix} \quad \text{and} \quad var(\varpi_t) = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix},$$

I consider a high-frequency DGP given by $u_t^{(m)} = Au_{t-1}^{(m)} + \varpi_t$ for $\varrho = 0, 1$ and $a = 0, 1/2$. As such, the regressor is $I(\varrho)$ and may or may not be correlated with the error term. I set $\beta = 10$, and perform each simulation $10,000$ times.

Table 2: Size and Power of the MiDaS Variable Addition Test

	I(0) Regressors						I(1) Regressors					
$a =$	0			1/2			0			1/2		
$T =$	25	50	100	25	50	100	25	50	100	25	50	100
$H_0(1)$	0.10	0.04	0.03	0.10	0.05	0.03	0.09	0.04	0.03	0.12	0.09	0.12
$H_0(2)$	0.16	0.09	0.06	0.15	0.09	0.07	0.16	0.09	0.06	0.46	0.75	0.98
$H_0(3)$	0.16	0.09	0.07	0.15	0.09	0.06	0.16	0.09	0.07	0.15	0.08	0.06
$H_0(4)$	0.10	0.05	0.03	0.10	0.04	0.03	0.09	0.05	0.03	0.15	0.14	0.26
$H_0(5)$	0.10	0.04	0.03	0.10	0.05	0.03	0.10	0.04	0.03	0.10	0.05	0.05
$H_0(6)$	0.10	0.04	0.03	0.09	0.04	0.03	0.10	0.05	0.03	0.10	0.06	0.07
$H_A(1)$	0.90	1.00	1.00	0.97	1.00	1.00	0.94	1.00	1.00	0.98	1.00	1.00
$H_A(2)$	0.22	0.31	0.64	0.28	0.46	0.84	0.22	0.33	0.67	0.31	0.54	0.90
$H_A(3)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$H_A(4)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$H_A(5)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$H_A(6)$	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 shows percentages of rejections using a test with a nominal size of 0.05. These are size under the six nulls and power against the six alternatives. The test appears to be extremely powerful against all of the alternatives except $H_A(2)$ – even at the small sample size of 25. Recall from Figure 1 that the exponential Almon lag approximates $H_A(2)$ quite well, so no significant power *should* be expected against that alternative. Imposing the Almon structure should not change the fit substantially, so a rejection might not be desirable.

Compared to a nominal size of 0.05, the test appears to be undersized for many of the nulls with a sample size as small as 50. For a sample size of 25, size distortion is more noticeable. None of the sizes appear larger than 0.10 for $T \geq 50$, except in cases with an I(1) regressor correlated with the error term, which are *not* covered by Proposition 5.

Nevertheless, even in cases with an I(1) regressor correlated with the error term, size distortion is not always discernible. The proof of Proposition 5 suggests that non-normality is only problematic for the single I(1) regressor in (p_t). The remaining $m-1$ regressors are I(0), and estimators of the corresponding coefficients are asymptotically normal, even with such correlation. The numerical results suggest that the normality of the I(0) regressors may in some sense dilute the non-normality of the single I(1) regressor. The case with the most size distortion occurs under $H_0(2)$, a null that assigns a unit weight to the first regressor, which is I(1), but zero weights to the remaining I(0) regressors. As the estimated model approximates the null model, more weight is given to the asymptotically non-normal coefficient estimates than to the asymptotically normal coefficient estimates. Similar intuition holds for $H_0(4)$, where relatively more weight is given to the first regressor, though the distortion is less dramatic. Asymptotic normality is also violated under the remaining four nulls when $a \neq 0$, but the weight given to the asymptotically non-normal coefficient estimate is small enough and the limiting distribution is close enough to a chi-squared distribution that the size distortion does not appear to be substantial.

5. An Application to Nowcasting Economic Activity

To illustrate the utility of CoMiDaS regressions, I consider a simple exercise of nowcasting annual log global real economic activity (RGDP) using past RGDP and high-frequency covariates measured over the same period.

A major factor of production in any modern economy is energy, and such economies are highly dependent on hydrocarbons and their substitutes. Oil use is particularly prevalent and its price naturally affects that of other energy sources. Although demand for and supply of oil in the United States were major movers of oil prices before the 1970's, the OPEC era changed the landscape of oil supply in the 1970's and into the 1980's. By 1985-86, the market power of OPEC collapsed. Moreover, in the last few decades the rise of demand in emerging economies, such as China, has fueled large demand increases. With a relatively inelastic supply of oil in both short- and long-run senses, demand has become an important driver of price. Hotelling (1931) provided theoretical models for the price of an exhaustible resource, such as oil. His perfect competition model provides a basis for a linear trend in the log of prices. Under the more realistic assumption of imperfect competition, Hotelling noted the importance of demand in determining price. Moreover, a number of recent papers have emphasized the role of demand in determining oil prices – e.g., Barsky and Kilian (2004), Kilian (2008, 2009), and Hamilton (2009).

Hamilton (2009) noted the stability of the relationship between US real GDP and US oil consumption over time. Extrapolating such a relationship to other developed and developing economies, and since oil is a factor production in global output and is traded globally, global real economic activity is a reasonable proxy for oil demand. Miller and Ni (2011) also made this argument, noting empirical evidence supporting cointegration of log real oil prices (ROIL) with RGDP since 1986, but with a common stochastic trend fluctuating around different linear trends.

Another type of series that may be tied to global real economic activity is an index for international maritime shipping rates, such as the one constructed by Kilian (2009). His purpose for creating this index was quite different – he used a shipping index as a proxy for oil demand, later found to be cointegrated with oil prices by He *et al.* (2010). As noted by Klovland (2002) and others, maritime shipping is linked with economic activity.[11] In this light, such an index may be viewed as a leading indicator of economic activity. The Baltic Dry Index (BDI or RBDI for log real BDI) is one such index.

It is not surprising that once the logs of these series are linearly detrended[12] the series are

[11] Klovland's (2002) analysis focused on pre-WW1 data. Using more recent data from the World Bank <http://data.worldbank.org> and UNCTAD <http://unctadstat.unctad.org>, international trade accounted for roughly 21% of world GDP in 2009.

[12] Since even the nonlinear-in-parameters MiDaS model is still linear in data, detrending is a logical first step for dealing with a model that contains such a trend. I do not specifically allow for deterministic trends in the econometric specification above, since the asymptotic distributions would contain additional nuisance parameters from these trends. In spite of the nuisance parameters, all of the qualitative results should hold with the addition of general trends of polynomial asymptotic order.

highly correlated. However, they are observed at different frequencies. Although RGDP may be a demand proxy for ROIL and RBDI, the latter two are measured much more frequently and accurately. Therefore, it is reasonable to use the informational content from ROIL and RBDI to try to predict RGDP.

No assumptions about the direction of causality need to made – nor should they be. Endogeneity may be expected in forecasting any of them using any of the others. Since the specification above allows for the error to be correlated with the regressors, endogeneity is not problematic from an econometric point of view. To highlight even more clearly the robustness of the model for forecasting, I adopt a *suboptimal* modeling strategy of not explicitly modeling lags of the indices in previous years, nor lags of the regressand beyond the previous year. The error structure may therefore be correlated both serially and with the regressors.

The model I estimate is adapted from (3) as

$$y_t = \delta' c_t + \rho y_{t-1} + \beta' \sum_{k=0}^{m-1} \Pi_{k+1}(\gamma) x_{t-k/m}^{(m)} + \eta_t,$$

where $c_t = (1, t)'$ and ρ is not restricted to unity. Rolling back the indices on the error term and regressand for nowcasting rather than forecasting poses no econometric problem, since endogeneity is allowed. The diagonal weight matrix Π_s is 2×2 with diagonal elements $\pi_{1,s}(\gamma_1, \gamma_2)$ and $\pi_{2,s}(\gamma_3, \gamma_4)$.

I use annual RGDP from 1985-2008 ($T = 24$) and monthly ROIL and RBDI over the same period ($m = 12$).[13] As a nowcasting exercise, I conduct one-step ahead forecasts of RGDP using lagged RGDP and contemporaneous RBDI and ROIL for 2001 through 2008. The smallest sample therefore has $T_1 = 16$, while the largest has $T_1 = 23$. Note that with $T = 24$, $m = 12$, and two high-frequency regressors, a linear ADL model, with 29 parameters, or 27 after the identifying restriction, would be infeasible. Average sampling reduces the parameters to 5, while using the two exponential Almon lag polynomials described above leaves 9.

Since the coefficients of the lag polynomial are exponential in their parameters, I restrict the parameter space of γ to the hypercube bounded by $(-5, 1)$ using a piecewise logistic function with each piece having equal derivatives at zero. These bounds were chosen so as *not* to restrict the weighting schemes for $m = 12$. There is no numerical difference – up to rounding error – between, say, $(\gamma_1, \gamma_2) = (-5, -5)$ and $(\gamma_1, \gamma_2) = (-6, -5)$. The monotonic transformation of the parameter space mutes the numerical sensitivity of the algorithm, providing more stable parameter estimates.

With such a small sample in mind, the number of instruments for the test should be kept to a minimal number, while still providing disparate weighting schemes. I choose three instruments – using weighting schemes described by $H_0(5)$, $H_A(5)$, and $H_A(6)$ above – based on diversity. I conduct tests on each of the two lag polynomials separately.

[13] The global real GDP series used in this study is from Angus Maddison's historical statistics (Maddison, 2010). Real oil price and real shipping rate series are created by dividing the West Texas Intermediate spot oil price series and the Baltic Dry Index by the US producer price index (all commodities). I take logs of the resulting series.

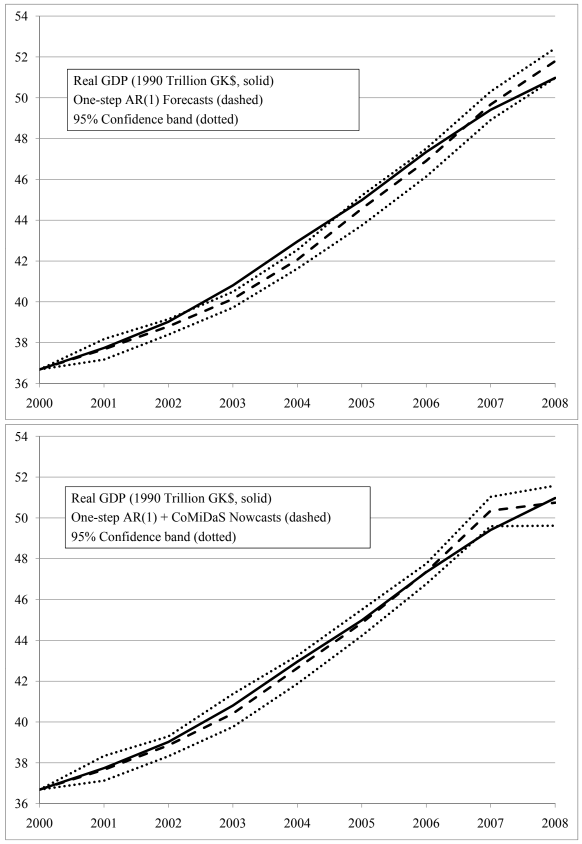

Figure 2: (a) RGDP target and one-step AR(1) forecasts, and (b) RGDP target and one-step AR(1) plus CoMiDaS nowcasts.

Table 3: Estimation of γ and VAT Results

	RBDI			ROIL		
Subsample	$\hat{\gamma}_1$	$\hat{\gamma}_2$	VAT	$\hat{\gamma}_3$	$\hat{\gamma}_4$	VAT
1985-00	-0.00000154	-0.00001770	2.06	-0.00000007	-0.00000246	21.13
1985-01	-0.00000025	-0.00000658	2.10	-0.00000043	-0.00000280	18.27
1985-02	-0.00000082	-0.00001051	1.71	-0.00000011	-0.00000061	16.25
1985-03	-0.00007397	-0.00090516	1.44	0.00000046	0.00005264	7.30
1985-04	-0.02267300	-0.28054330	0.29	-0.00067434	0.00130722	4.18
1985-05	-0.00010143	-0.00129093	0.85	-0.00000433	-0.00003896	6.22
1985-06	-0.00000212	-0.00002778	0.61	-0.00000006	-0.00000060	6.58
1985-07	-0.00000092	-0.00001092	1.25	0.00000017	0.00000200	4.77

Table 3 summarizes the results of estimation and testing. The deep parameter vector γ is estimated to be very close to zero and slightly negative for most of subsamples and lag

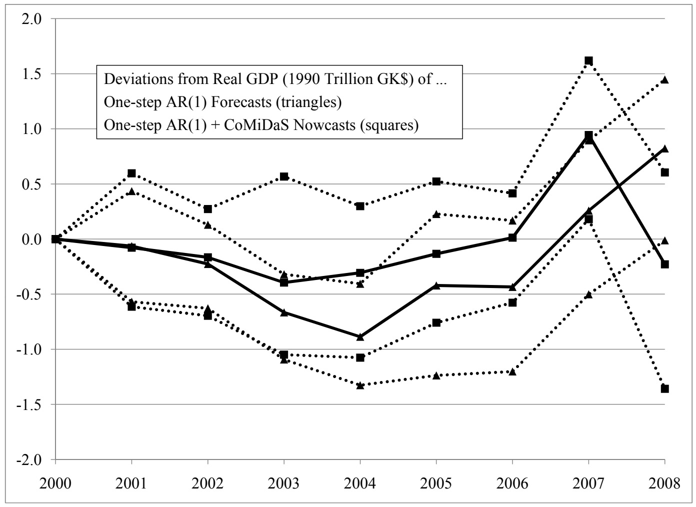

Figure 3: Deviations of one-step AR(1) forecasts and AR(1) plus CoMiDaS nowcasts from RGDP.

polynomials. This result suggests a nearly flat weighting scheme, with slightly more emphasis on recent RBDI and ROIL. In other words, shipping rates and oil prices in December are slightly more influential than those in January for nowcasting RGDP. This result is sensible in light of the results of Chambers (2003) and Miller (2011), which suggest that average sampling of the regressors is most efficient when the regressand is aggregated in the same way and the regression is cointegrating.

Compared to a χ^2_3 critical value of 7.81 at 5% significance, I cannot reject the MiDaS null for any of the subsamples for RBDI and most for ROIL. Non-rejection provides evidence supporting the adequacy of CoMiDaS specification.[14] In light of the simulation results in Table 2, some size distortion is to be expected, since the series are most likely I(1) and cointegrated, but the regressor is *not* strictly exogenous. Since the weights are estimated to be close to flat – i.e., $H_0(1)$, the size distortion is not likely to be severe.

Finally, Figures 2 and 3 show the CoMiDaS nowcasts and simple AR(1) forecasts for 2001-08, compared to the target expressed in trillions of 1990 Geary-Khamis dollars. 95% confidence bands were calculated using a simple bootstrap, which should be conservatively wide if serial correlation is present. The CoMiDaS nowcasts evidently track the target more closely than the AR(1) forecasts until 2007, when both forecasts overshoot the target. Overshooting in 2007 could result from revision of the 2007 RGDP datum following the global recession that began in late 2007. The difference in 2008 becomes clear, as the AR(1) forecast continues to overshoot, while that of the CoMiDaS nowcast corrects for the recession. The correction comes from a precipitous drop in ROIL and RBDI that did not occur until 2008.

[14]I cannot reject a null of flat aggregation against MiDaS, either, using an LR test similar to the LM test of Andreou *et al.* (2010a). Results not shown.

6. Concluding Remarks

The MiDaS specification introduced by Ghysels *et al.* (2004) provides a very useful parsimonious specification for regression and forecasting using high-frequency regressors. Asymptotic analyses by Ghysels and his coauthors support a range of mixed-frequency models for which the nonlinear ADL specification may be useful. This paper broadens this range substantially by introducing CoMiDaS regressions, allowing for the possibility of cointegration of series that contain unit roots. Moreover, the asymptotic analysis allows for the possibility of correlation of the error term serially and with both I(1) and I(0) regressors. In this light, my results promote the MiDaS approach to mixed-frequency time series by extending the validity of the approach.

Further, I present a simple variable addition test of the MiDaS null against a more general ADL specification. The test is feasible even when the number of high-frequency regressors is large relative to the low-frequency sample size, which is precisely when the MiDaS specification is most useful. Both asymptotic and simulated results suggest that the test suffers very little size distortion in most cases, while enjoying substantial power in samples as small as $T = 25$.

An application to nowcasting global real economic activity using a lag and using contemporaneous real oil prices and an index of real maritime shipping prices illustrates the utility of CoMiDaS regressions and of the proposed MiDaS variable addition test.

Appendix A: Proofs of the Theoretical Results

The following ancillary lemma collects results to make the ensuing proofs more tractable.

Lemma A1. *Under Assumptions [A3]-[A4] and for an arbitrary (n_0+1)-vector a with finite norm,*

[a] $T^{-1}\sum_t(b_{0t}b_{0t}' - \Sigma_{00})a \to_p 0$

[b] $T^{-1/2}\sum_t(b_{0t}b_{0t}' - \Sigma_{00})a \to_d (I_{n_0+1}\otimes a')\mathbf{N}(0,\Xi)$,

[c] $T^{-1}\sum_t p_{1t}b_{0t}' \to_d \int B_1 dB_0' + \Delta_{01}'$, *and*

[d] $T^{-2}\sum_t p_{1t}p_{1t}' \to_d \int B_1 B_1'$.

The convergences in parts [b]-[d] are joint, but the limiting distribution in [b] is independent of those in parts [c] and [d].

Proof of Lemma A1. Proofs of parts [a] and [b] for the univariate case are given by Theorems 3.7 and 3.8 of Phillips and Solo (1992). By Assumption [A3], the vector $b_{0t}b_{0t}'a$ may be written as

$$\begin{aligned}&\sum\nolimits_{i=0}^{\infty}\Psi_i v_{0,t-i}v_{0,t-i}'\Psi_i' a + \sum\nolimits_{r=1}^{\infty}\sum\nolimits_{i=0}^{\infty}\Psi_i v_{0,t-i}v_{0,t-i-r}'\Psi_{i+r}' a\\ &+\sum\nolimits_{r=1}^{\infty}\sum\nolimits_{i=0}^{\infty}\Psi_{i+r} v_{0,t-i-r}v_{0,t-i}'\Psi_i' a,\end{aligned}$$

where $(v_{0,t})$ is an mds with respect to $(\mathcal{F}_{t-1})$. This expression is

$$b_{0t}b_{0t}'a = (I \otimes a')\Xi_{00}(L)w_{00,t} + (I \otimes a')\sum\nolimits_{r=1}^{\infty}(\Xi_{0r}(L)w_{0r,t} + \Xi_{r0}(L)w_{r0,t}),$$

using the matrix polynomial $\Xi_{rs}(z) \equiv \sum_{k=0}^{\infty}(\Psi_{k+r} \otimes \Psi_{k+s})z^k$ and $w_{rs,t} \equiv (v_{0,t-r} \otimes v_{0,t-s})$.

Similarly to Phillips and Solo (1992), a Beveridge-Nelson decomposition (Beveridge and Nelson, 1981) may be applied to the matrix polynomial $\Xi_{rs}(z)$, so that $\Xi_{rs}(z) = \Xi_{rs}(1) - (1-z)\tilde{\Xi}_{rs}(z)$, where $\tilde{\Xi}_{rs}(z) = \sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}(\Psi_{j+r} \otimes \Psi_{j+s})z^i$. Subtracting $\mathbf{E}b_{0t}b_{0t}'a = (I \otimes a')\Xi_{00}(1)\mathbf{E}w_{00,t}$ and adding across t shows that the summation in parts [a] and [b] is

$$\begin{aligned}&(I \otimes a')\sum\nolimits_t\left[\Xi_{00}(1)(w_{00,t} - \mathbf{E}w_{00,t}) + \sum\nolimits_{r=1}^{\infty}(\Xi_{0r}(1)w_{0r,t} + \Xi_{r0}(1)w_{r0,t})\right] \\ &- (I \otimes a')\sum\nolimits_t\left[\tilde{\Xi}_{00}(L)\triangle w_{00,t} + \sum\nolimits_{r=1}^{\infty}(\tilde{\Xi}_{0r}(L)\triangle w_{0r,t} + \tilde{\Xi}_{r0}(L)\triangle w_{r0,t})\right],\end{aligned} \tag{14}$$

and the leading term in square brackets drives the asymptotic results.

Under Assumptions [A3]-[A4] the LLN and CLT for variances of Phillips and Solo (1992) apply directly for elements of $w_{rs,t}$ that are squares of single elements of v_t. The generalization of their asymptotics to multivariate series requires generalizing the LLN and CLT to covariances. I present only a sketch of the proof, since the details are tedious but follow in a straightforward manner from Phillips and Solo (1992).

The matrix Ξ in the limiting variance of the CLT is the same as that defined in Section 3. Specifically,

$$\begin{aligned}\Xi &\equiv \mathbf{E}[b_{0t}b_{0t}' \otimes b_{0t}b_{0t}'] - \mathbf{E}[b_{0t} \otimes b_{0t}]\mathbf{E}[b_{0t}' \otimes b_{0t}'] \\ &= \Xi_{00}(1)K_{00}\Xi_{00}'(1) + \sum\nolimits_{r=1}^{\infty}\{\Xi_{0r}(1)K_{0\bullet}\Xi_{0r}'(1) + \Xi_{r0}(1)K_{0\bullet}\Xi_{r0}'(1) \\ &+ \Xi_{0r}(1)K_{\bullet 0}\Xi_{r0}'(1) + \Xi_{r0}(1)K_{\bullet 0}\Xi_{0r}'(1)\}\end{aligned}$$

where

$$\begin{aligned}K_{00} &\equiv \mathbf{E}[v_{0t}v_{0t}' \otimes v_{0t}v_{0t}'] - \mathbf{E}[v_{0t} \otimes v_{0t}]\mathbf{E}[v_{0t}' \otimes v_{0t}'] \\ K_{0\bullet} &\equiv \mathbf{E}[v_{0t}v_{0t}' \otimes v_{0t-\bullet}v_{0,t-\bullet}'] \\ K_{\bullet 0} &\equiv \mathbf{E}[v_{0,t-\bullet}v_{0t}' \otimes v_{0t}v_{0,t-\bullet}']\end{aligned}$$

are time-invariant by assumption, so that $\bullet$ is just a placeholder.

A sufficient condition that Phillips and Solo (1992) employ is that $\sum_{s=1}^{\infty} s\psi_s^2 < \infty$. For the multivariate case, a representation element of the matrix $(\Psi_{i+r} \otimes \Psi_{i+s})$ may be written as $\psi_{jk,i+r}\psi_{uv,i+s}$, such that $\Psi_{i+r} = [\psi_{jk,i+r}]$ and $\Psi_{i+s} = [\psi_{uv,i+s}]$. The multivariate generalization is straightforward under the analogous sufficient condition that $\sum_{s=0}^{\infty} s\,|\psi_{jk,s}\psi_{uv,s}| < \infty$. Note that

$$\sum\nolimits_{s=0}^{\infty} s\,|\psi_{jk,s}\psi_{uv,s}| \leq \sum\nolimits_{s=0}^{\infty} s \max_{j,k}|\psi_{jk,s}|^2 \leq \sum\nolimits_{s=0}^{\infty} s \max_{j,k}|\psi_{jk,s}| \leq \sum\nolimits_{s=0}^{\infty} s\,\|\Psi_s\|$$

which is finite under Assumption [A3]. The last equality follows from one of the relationships between matrix norms presented by Lütkepohl (1996, pg. 111), for example. Similarly to

Phillips and Solo (1992), the second square-bracketed term of (14) may be shown to be $o_p\left(T^{1/2}\right)$, so that it may be ignored in both parts [a] and [b]. The limiting variance in part [b] follows directly from the variance of the square-bracketed expression in the first term of (14).

The proofs of parts [c] and [d] follow from standard asymptotic arguments using the invariance principle of Phillips and Solo (1992).

Finally, the joint convergence and independence of the limiting distribution in part [b] from those in [c] and [d] follows along logic similar to that of Chang *et al.* (2001). The joint convergence follows from the convergence of $T^{-1/2}\sum_{t=1}^{[Tr]} b_t$ under the stated assumptions. Moreover, the leading term of (14) is a martingale under Assumption [A4], so that $\mathbf{E}[((v_{0,t}\otimes v_{0,t}) - \mathbf{E}[v_{0,t}\otimes v_{0,t}])v_s'] = 0$ for $s \neq t$. For $s = t$, the equality holds if the third moment matrix of (v_t) is zero. Since this expected value is zero, the limiting distribution in part [b] is uncorrelated with the limiting Brownian motion $B(r)$ in parts [c] and [d]. Normality then implies independence. □

Proof of Theorem 1. As noted above, Assumptions [A1]-[A2] and [N1]-[N3] are sufficient to write the difference $g(\hat{\theta}_{NLS}) - g(\theta_{\min})$ as $M_T^{-1}P_T$, with $P_T = (N_T - N) - (M_T - M)M^{-1}N$ as defined above. Using parts [a] and [c] of Lemma A1, $N_T - N$ is $o_p(1)$ for I(0) regressors and $O_p(1)$ for I(1) regressors. Using part [a], [c], and [d], $M_T - M$ is $o_p(1)$ for products of I(0) regressors, $O_p(T)$ for products of I(1) regressors, and $O_p(1)$ for cross-products of I(0) and I(1) regressors. M is $O(1)$ for products of I(0) regressors and cross-products, but $O(T)$ for products of I(1) regressors. As a result, M^{-1} is $O(1)$ in the upper left block, but $O\left(T^{-1}\right)$ in all other blocks. Since N is $O(1)$, the entire second term of P_T is $o_p(1)$ for rows corresponding to the the I(0) regressors and $O_p(1)$ for those of the I(1) regressors, so that the respective rows of P_T itself are $o_p(1)$ and $O_p(1)$. Finally, similarly to M^{-1}, M_T^{-1} is $O_p(1)$ for products of I(0) regressors, but $O_p\left(T^{-1}\right)$ for products of I(1) regressors and cross-products, which gives the result. □

Proof of Corollary 2. The consistency in Theorem 1 is sufficient for the proof of the corollary, if $g(\theta)$ is continuous at $\theta_{\min}$, as I assume, and if $\theta_{\min}$ such that the first-order condition of the MSFE minimization problem is satisfied. Assumptions [N1]-[N2] and [N3'] are sufficient for the latter. □

Proof of Theorem 3. I employ the Wooldridge conditions for the proof of his Theorem 10.1 (Wooldridge, 1994, pg. 2711). Conditions (i)-(iii) of Wooldridge (1994) are satisfied by Assumption [N3] and by construction of $Q_T(\theta)$ – in particular by the linearity of the CoMiDaS regression in g.

It only remains to show the joint convergence of

$$(\nu_T^{-1/2}H_T(g(\theta_{\min}))\nu_T^{-1/2}, \nu_T^{-1/2}J_T(g(\theta_{\min}))).$$

The convergence of

$$\nu_T^{-1/2}H_T(g(\theta_{\min}))\nu_T^{-1/2} = \nu_T^{-1/2}TM_T\nu_T^{-1/2} \to_d H$$

is an immediate consequence of parts [a], [c], and [d] of Lemma A1.

The convergence of $\nu_T^{-1/2} J_T(g(\theta_{\min})) = -\nu_T^{-1/2} T P_T$ also follows from Lemma A1, but is more involved. Looking at each component of P_T separately,

$$M^{-1} = \begin{bmatrix} \Sigma_{pp}^{-1} + O\left(T^{-1}\right) & O\left(T^{-1}\right) \\ -T^{-1}\Omega_{11}^{-1}\Delta_{p1}'\Sigma_{pp}^{-1} + o\left(T^{-1}\right) & T^{-1}\Omega_{11}^{-1} + o\left(T^{-1}\right) \end{bmatrix}$$

and

$$M_T - M = \begin{bmatrix} O_p\left(T^{-1/2}\right) & O_p\left(1\right) \\ O_p\left(1\right) & O_p\left(T\right) \end{bmatrix}$$

from parts [b], [c], and [d] of Lemma A1. $T(M_T - M)M^{-1}$ simplifies to

$$\begin{bmatrix} \sum_t p_{0t}p_{0t}' - \Sigma_{pp} & 0 \\ \sum_t p_{1t}p_{0t}' - \Delta_{p1}' & T^{-1}\sum_t p_{1t}p_{1t}' - \Omega_{11} \end{bmatrix} \begin{bmatrix} \Sigma_{pp}^{-1} & 0 \\ -\Omega_{11}^{-1}\Delta_{p1}'\Sigma_{pp}^{-1} & \Omega_{11}^{-1} \end{bmatrix} + \begin{bmatrix} O_p\left(1\right) & O_p\left(1\right) \\ O_p\left(T\right) & O_p\left(T\right) \end{bmatrix}$$

using these asymptotic rates.

Partitioning $P_T = (P_{0T}', P_{1T}')'$ conformably with $p_t = (p_{0t}', p_{1t}')'$, it follows that

$$T^{1/2}P_{0T} = T^{-1/2}E'\sum\nolimits_t (b_{0t}b_{0t}' - \Sigma_{00})\kappa \to_d (E' \otimes \kappa')\mathbf{N}\left(0, \Xi\right)$$

where the convergence follows from part [b] of Lemma A1.

Further, P_{1T} may be written as

$$P_{1T} = T^{-1}\sum\nolimits_t (p_{1t}b_{0t}' - \Delta_{01}')\kappa - \left(T^{-2}\sum\nolimits_t p_{1t}p_{1t}'\Omega_{11}^{-1} - I\right)\zeta$$

so that convergence

$$P_{1T} \to_d \int B_1 dB_0'\kappa - \left(\int B_1 B_1'\Omega_{11}^{-1} - I\right)\zeta$$

follows from parts [c] and [d] of Lemma A1. Finally, since the convergences are joint, condition (iv) of Wooldridge (1994) is satisfied and the stated result is obtained. □

Proof of Corollary 4. The proof is similar to that of the preceding theorem, but [N2'] and [N3'] are needed to ensure conditions (i) and (ii) of Wooldridge's (1994) Theorem 10.1. I next verify condition (iv) and then condition (iii) last.

To analyze the score, note that

$$(\nu_T^*)^{-1}J_T^*(\theta) = \begin{bmatrix} T^{-1/2}\frac{\partial g_0'(\theta)}{\partial\theta_0} & 0 \\ T^{-1}\frac{\partial g_0'(\theta)}{\partial\theta_1} & T^{-1}\frac{\partial g_1'(\theta)}{\partial\theta_1} \end{bmatrix} \begin{bmatrix} TP_{0T} \\ TP_{1T} \end{bmatrix} = \begin{bmatrix} \frac{\partial g_0'(\theta)}{\partial\theta_0}T^{1/2}P_{0T} \\ \frac{\partial g_1'(\theta)}{\partial\theta_1}P_{1T} + \frac{\partial g_0'(\theta)}{\partial\theta_1}P_{0T} \end{bmatrix}$$

comes from the partitioning of $\theta = (\theta_0', \theta_{10}', \theta_{11}')'$ discussed above. Convergence of $T^{1/2}P_{0T}$ and P_{1T} is the same as in the proof of the preceding theorem, so that $(\nu_T^*)^{-1}J_T^*(\theta_{\min}) \to_d J^*(\theta_{\min})$.

The first term of the Hessian $(\nu_T^*)^{-1}(\partial g'(\theta)/\partial\theta)TM_T(\partial g(\theta)/\partial\theta')(\nu_T^*)^{-1}$ is

$$\left[\begin{array}{cc} \frac{\partial g_0'(\theta)}{\partial\theta_0}M_{00,T}\frac{\partial g_0(\theta)}{\partial\theta_0'} & 0 \\ 0 & \frac{\partial g_1'(\theta)}{\partial\theta_1}T^{-1}M_{11,T}\frac{\partial g_1(\theta)}{\partial\theta_1'} \end{array}\right] + O_p\left(T^{-1/2}\right)$$

using parts [a] and [c] of Lemma A1. The limiting distribution $H(\theta_{\min})$ at $\theta_{\min}$ then follows from parts [a] and [d]. To show that the remaining term is negligible, note that $(TP_T' \otimes I)$ is a $g \times gn$ matrix

The $ng \times g$ matrix of second derivatives $\partial vec\left(\partial g(\theta)'/\partial\theta\right)/\partial\theta'$ has only zeros in the block consisting of the last ng_1 rows and first g_0 columns, due to the partitioning of $\theta = (\theta_0', \theta_{10}', \theta_{11}')'$. Scaling by $(\nu_T^*)^{-1}$ allows

$$(\partial vec\left(\partial g'(\theta)/\partial\theta\right)/\partial\theta')(\nu_T^*)^{-1} = \left[\begin{array}{cc} T^{-1/2}V_{00}(\theta) & T^{-1}V_{01}(\theta) \\ 0 & T^{-1}V_{11}(\theta) \end{array}\right],$$

where the matrices $V_{00}(\theta)$ $(ng_0 \times g_0)$, $V_{11}(\theta)$ $(ng_1 \times g_1)$, and $V_{01}(\theta)$ $(ng_0 \times g_1)$ are defined by the second derivatives. Further, $(TP_T' \otimes I)$ may be written as

$$(TP_T' \otimes I) = \left[\begin{array}{cc} (TP_{0T}' \otimes I_g) & (TP_{1T}' \otimes I_g) \end{array}\right],$$

so that $(TP_T' \otimes I)(\partial vec\left(\partial g'(\theta)/\partial\theta\right)/\partial\theta')(\nu_T^*)^{-1}$ is

$$\left[\begin{array}{cc} (T^{1/2}P_{0T}' \otimes I_g)V_{00}(\theta) & (P_{0T}' \otimes I_g)V_{01}(\theta) + (P_{1T}' \otimes I_g)V_{11}(\theta) \end{array}\right],$$

which is $O_p(1)$ from the previous theorem. Premultiplying by $(\nu_T^*)^{-1}$ renders the entire term $O_p\left(T^{-1/2}\right)$, so that it is asymptotically negligible. Condition (iv) of Wooldridge (1994) is thus satisfied.

Finally, condition (iii) requires that an increasing sequence μ_T exists, such that $\mu_T(\nu_T^*)^{-1} \to 0$ and

$$\max_{\theta\in\Theta_T}\left\|\mu_T^{-1}\left(H_T(\theta_{\min}) - H_T(\theta_*)\right)\mu_T^{-1}\right\| = o_p(1), \tag{15}$$

where $\Theta_T \equiv \{\theta_* \in \Theta : ||\mu_T(\theta_* - \theta_{\min})|| \leq 1\}$ is a sequence of shrinking neighborhoods of $\theta_{\min}$.

The matrix inside the norm in (15) is

$$\begin{aligned} &\mu_T^{-1}(h'(\theta_{\min})TM_Th(\theta_{\min}) - h'(\theta_*)TM_Th(\theta_*))\mu_T^{-1} \\ &- \mu_T^{-1}\left(\begin{array}{c} (T((g(\theta_{\min}) - \alpha)'M_T - N_T') \otimes I)k(\theta_{\min}) \\ -(T((g(\theta_*) - \alpha)'M_T - N_T') \otimes I)k(\theta_*) \end{array}\right)\mu_T^{-1}, \end{aligned} \tag{16}$$

where $h(\theta) \equiv \partial g(\theta)/\partial\theta'$ and $k(\theta) \equiv \partial vec(\partial g'(\theta)/\partial\theta)/\partial\theta'$. Condition (iii) requires that there exists a sequence (μ_T) such that (16) is $o_p(1)$.

Similarly to Chang *et al.* (2009), let $\mu_T = (\nu_T^*)^{1-\delta}$ and $\theta_* = \theta_{\min} + (\nu_T^*)^{1-\delta}$ for arbitrarily small $\delta > 0$. Moreover, let

$$g(\theta_*) = g(\theta_{\min}) + R_0, \quad h(\theta_*) = h(\theta_{\min}) + R_1, \quad \text{and} \quad k(\theta_*) = k(\theta_{\min}) + R_2 \tag{17}$$

with R_0, R_1, and R_2 defined as conformable vectors or matrices. The elements of each of these are at most $O_p(T^{-1/2+\delta/2})$ from the mean value theorem and the existence of third derivatives of $g(\theta)$ assumed by [N2'].

The first term of (16) may be written as

$$(\nu_T^*)^{\delta-1}(h'(\theta_{\min})TM_TR_1 + R_1'TM_Th(\theta_{\min}) + R_1'TM_TR_1')(\nu_T^*)^{\delta-1}$$

by expanding $h'(\theta_*)TM_Th(\theta_*)$ using (17). Since R_1 is no larger than $O(T^{-1/2+\delta/2})$, ν_T^δ is no larger than $O(T^\delta)$, and from the convergence of the first term of the Hessian above, the entire first term of (16) is no larger than $O_p\left(T^{-1/2+5\delta/2}\right)$. There exists some small $\delta > 0$ such that this is $o_p(1)$.

The second term of (16) may be written as

$$(\nu_T^*)^{\delta-1}((TP_T \otimes I)R_2 + (R_0'TM_T \otimes I)(k(\theta_{\min}) + R_2))(\nu_T^*)^{\delta-1}$$

using (17). By the same logic, this term is also $o_p(1)$, so that (15) and, consequently, condition (iii) of Wooldridge (1994) are satisfied. □

Proof of Proposition 5. Rewriting (13) as $\hat{\eta}_{t+1} = p_t'W^*\varsigma + e_t$, an F-test of the whole regression takes the form $qF = T(\hat{\varsigma}-\varsigma)'W^{*\prime}M_TW^*(\hat{\varsigma}-\varsigma)/\hat{\sigma}_e^2$, where $\hat{\varsigma}$ is the least squares estimator of ς and $\hat{\sigma}_e^2$ is the usual variance estimator of (e_t). The distribution for a test on the subvector ς_1 is based on that for the whole regression, but with reduced rank.

The least squares estimator may be written as $\hat{\varsigma} = (W^{*\prime}M_TW^*)^{-1}W^{*\prime}\lambda_T$ using the restricted estimator in (9) and the NLS residuals in (12). Furthermore, $\lambda_T - \lambda$ may be written as

$$\lambda_T - \lambda = (N_T - N) - (M_T - M)(g(\hat{\theta}_{NLS}) - \alpha) - M(g(\hat{\theta}_{NLS}) - \alpha - \tau_*),$$

using the restricted estimator and the minimum MSFE parameter vector. When optimized properly, the second term becomes $(M_T - M)\tau_*$ and the third term becomes zero. The coefficient ς is equal to

$$\varsigma = (W^{*\prime}MW^*)^{-1}W^{*\prime}N - (W^{*\prime}MW^*)^{-1}W^{*\prime}M(w(\theta_{\min})\beta_{\min} - w\beta)) = (W^{*\prime}MW^*)^{-1}W^{*\prime}\lambda$$

using the minimum MSFE parameter vector. Subtracting ς from $\hat{\varsigma}$ yields

$$\hat{\varsigma} - \varsigma = (W^{*\prime}M_TW^*)^{-1}W^{*\prime}((N_T - N) - (M_T - M)\xi)$$

after some algebra, where $\xi \equiv \tau_* + W^*(W^{*\prime}MW^*)^{-1}W^{*\prime}\lambda$.

Consider the case of part [a] with only I(0) regressors. Using the definitions of M_T and N_T, the above expression may be further rewritten as

$$\hat{\varsigma} - \varsigma = (W^{*\prime}M_TW^*)^{-1}W^{*\prime}E'\left(T^{-1/2}\sum\nolimits_t(b_{0t}b_{0t}' - \Sigma_{00})\right)\kappa$$

with $\kappa = (1, -\xi')'$. Using this notation, it is clear that

$$T^{1/2}(\hat{\varsigma} - \varsigma) \to_d (W^{*\prime}\Sigma_{pp}W^*)^{-1}W^{*\prime}(E' \otimes \kappa')\mathbf{N}(0, \Xi)$$

by part [b] of Lemma A1. In order to show the chi-squared distribution, the exact variance is necessary. Some rather tedious but straightforward algebra using standard results for fourth moments of multivariate normal distributions shows that

$$(E' \otimes \kappa')\Xi(E \otimes \kappa) = (\sigma_\varepsilon^2 + \xi'\Sigma_{pp}\xi - 2\xi'\sigma_{p\varepsilon})\Sigma_{pp} + (\sigma_{p\varepsilon} - \Sigma_{pp}\xi)(\sigma_{p\varepsilon} - \Sigma_{pp}\xi)'$$

Under the null, $\tau_* = 0$ and $\lambda = N = \sigma_{p\varepsilon}$, so that

$$W^{*\prime}(E' \otimes \kappa')\Xi(E \otimes \kappa)W^* = (\sigma_\varepsilon^2 - \sigma_{\varepsilon p}W^*(W^{*\prime}\Sigma_{pp}W^*)^{-1}W^{*\prime}\sigma_{p\varepsilon})W^{*\prime}\Sigma_{pp}W^*$$

and the asymptotic variance of the estimator is therefore

$$(\sigma_\varepsilon^2 - \sigma_{\varepsilon p}W^*(W^{*\prime}\Sigma_{pp}W^*)^{-1}W^{*\prime}\sigma_{p\varepsilon})(W^{*\prime}\Sigma_{pp}W^*)^{-1}.$$

To get a chi-squared distribution, the first factor must equal the probability limit of $\hat{\sigma}_e^2$.

The estimator $\hat{\sigma}_e^2$ is equal to

$$\hat{\sigma}_e^2 = T^{-1}\sum\nolimits_t \hat{e}_t^2 = T^{-1}\sum\nolimits_t \left(\varepsilon_{t+1} - p_t'M_T^{-1}(N_T - \lambda_T) - p_t'W^*(W^{*\prime}M_TW^*)^{-1}W^{*\prime}\lambda_T\right)^2$$

using the expression for $\hat{\varsigma}$ above and the restricted estimator in (9). Expanding the square yields

$$\hat{\sigma}_e^2 = T^{-1}\sum\nolimits_t \varepsilon_{t+1}^2 - N_T'M_T^{-1}N_T + \lambda_T'(M_T^{-1} - W^*(W^{*\prime}M_TW^*)^{-1}W^{*\prime})\lambda_T. \qquad (18)$$

Under the null, the probability limit is in fact $\sigma_e^2 = \sigma_\varepsilon^2 - \sigma_{\varepsilon p}W^*(W^{*\prime}\Sigma_{pp}W^*)^{-1}W^{*\prime}\sigma_{p\varepsilon}$.

Finally, the degrees of freedom of the limiting chi-squared distribution for the whole regression comes from the rank $q+1$ of $W^{*\prime}\Sigma_{pp}W^*$. The degrees of freedom for the test of the subvector ς_1 comes from the difference in rank between $W^{*\prime}\Sigma_{pp}W^*$ and $w'\Sigma_{pp}w$, which is q. This completes the proof of part [a].

More generally, the n-vector (p_t) is assumed to be partitioned into I(0) and I(1) components, but $W^{*\prime}$ takes $q+1$ linear combinations defined by the q variables added. These linear combinations are themselves cointegrated regressors, and may be handled along the lines of Park and Phillips (1989). Define a $(q+1)\times(q+1)$ orthogonal matrix $A = (A_0, A_1)$ such that $AA' = I$ and such that $A_0'W^{*\prime}p_t \equiv z_{0t} \sim I(0)$ and $A_1'W^{*\prime}p_t \equiv z_{1t}$. These matrices rotate the regressor space so that the first component lies in the span of the $(m-1)$-dimensional cointegrating space and the second lies in the unidimensional space orthogonal to the cointegrating space.

Much of the notation used in the remainder of the proof is recycled from Section 3, with the difference being the rotation. I redefine $b_t \equiv (\varepsilon_{t+1}, z_{0t}', \triangle z_{1t})'$ as a $(q+2)$-vector, $b_{0t} \equiv (\varepsilon_{t+1}, z_{0t}')'$, and $b_{1t} \equiv \triangle z_{1t}$, so that the limiting variance from an invariance principle using (b_t) is

$$\Omega = \begin{bmatrix} \omega_\varepsilon^2 & \omega_{\varepsilon z} & \omega_{\varepsilon 1} \\ \omega_{z\varepsilon} & \Omega_{zz} & \omega_{z1} \\ \omega_{1\varepsilon} & \omega_{1z} & \omega_{11} \end{bmatrix} \quad \text{or} \quad \Omega = \begin{bmatrix} \Omega_{00} & \Omega_{01} \\ \Omega_{10} & \Omega_{11} \end{bmatrix}$$

in place of Ω in Section 3. All of the relevant variances in this proof are partitioned in the same way. Brownian motions are redefined accordingly.

I redefine the normalization matrix $\nu_T \equiv diag(T^{1/2}I_q, T)$ to be a $(q+1)\times(q+1)$ diagonal matrix. I redefine $\kappa \equiv (1, -\lambda_0^{*\prime}\Sigma_{zz}^{-1})'$ as a $(q+1)$-vector, $\zeta \equiv (\lambda_1^* - \delta_{z1}'\Sigma_{zz}^{-1}\lambda_0^*)$ as a scalar, where $\lambda^* \equiv A'W^{*\prime}\lambda$ is a $(q+1)$-vector partitioned as $\lambda^* = (\lambda_0^{*\prime}, \lambda_1^*)'$. I redefine E' as the unitary matrix that selects all but the first row of the following matrix of $(q+1)$ rows.

In part [b], there is at least one I(0) combination of $(W^{*\prime}p_t)$ – i.e., $z_{0t} \neq 0$. The estimator may be written as

$$\nu_T A'(\hat{\varsigma} - \varsigma) = H_T^{-1}J_T$$

where $H_T \equiv \nu_T^{-1}AW^{*\prime}TM_TW^*A'\nu_T^{-1}$ and $J_T \equiv \nu_T^{-1}A'W^{*\prime}T((N_T - N) - (M_T - M)\xi)$ are redefined from Section 3 with

$$\xi \equiv \tau_* + W^*A(A'W^{*\prime}MW^*A)^{-1}\lambda^*$$

redefined from part [a] of this proof. Under the null, $\tau_* = 0$ and

$$H_T = \begin{bmatrix} T^{-1}\sum_t z_{0t}z_{0t}' & T^{-3/2}\sum_t z_{0t}z_{1t} \\ T^{-3/2}\sum_t z_{1t}z_{0t}' & T^{-2}\sum_t z_{1t}^2 \end{bmatrix} \rightarrow_d \begin{bmatrix} \Sigma_{zz} & 0 \\ 0 & \int B_1^2 \end{bmatrix} \equiv H,$$

and

$$J_T = \begin{bmatrix} J_{0T} \\ J_{1T} \end{bmatrix} \rightarrow_d \begin{bmatrix} A_0'W^{*\prime}(E' \otimes \kappa')\mathbf{N}(0, \Xi) \\ \int B_1 dB_0'\kappa - (\int B_1^2\omega_{11}^{-1} - 1)\zeta \end{bmatrix} \equiv J,$$

where

$$J_{0T} \equiv T^{-1/2}E'\sum\nolimits_t (b_{0t}b_{0t}' - \Sigma_{00})\kappa$$
$$J_{1T} \equiv T^{-1}\sum\nolimits_t (z_{1t}b_{0t}' - \delta_{01}')\kappa + T^{-2}\sum\nolimits_t (z_{1t}^2\omega_{11}^{-1} - T)\zeta$$

using some algebra and limits along the same lines as the proof of Theorem 3. The variance matrix Ξ is defined exactly as in Section 3.

The test statistic may be rewritten as

$$qF = (\hat{\varsigma} - \varsigma)'A\nu_T\nu_T^{-1}A'W^{*\prime}TM_TW^*A\nu_T^{-1}\nu_TA'(\hat{\varsigma} - \varsigma)/\hat{\sigma}_e^2$$

and it follows from above that $\nu_TA'(\hat{\varsigma} - \varsigma) \rightarrow_d H^{-1}J$ and $\nu_T^{-1}A'W^{*\prime}TM_TW^*A\nu_T^{-1} \rightarrow_d H$.

In order for J to be a multivariate normal vector, the I(1) regressor must be strictly exogenous. This assumption implies that the I(0) regressors, defined by first differences of the I(1) regressor, are also strictly exogenous. However, I do not assume the I(0) regressors to be strictly exogenous. This allowance serves two purposes: (i) it leaves open the possibility for additional I(0) regressors, and, more importantly, (ii) it shows that only one element of the multivariate normal vector is not robust to the assumption.

The variance of the first q elements of J is given by $A_0'W^{*\prime}(E' \otimes \kappa')\Xi(E \otimes \kappa)W^*A_0 = (\sigma_\varepsilon^2 - \sigma_{\varepsilon z}\Sigma_{zz}^{-1}\sigma_{z\varepsilon})\Sigma_{zz}$ along exactly the same lines as the proof of part [a]. (This reduces to $\sigma_\varepsilon^2\Sigma_{zz}$ when the I(0) regressors are strictly exogenous.) The covariances of each of these elements with the remaining element of J is zero from the independence in Lemma A1. Under the assumptions of both strict exogeneity and that the error sequence is an mds, the distribution of the last element of J reduces to $\sigma_\varepsilon(\int B_1^2)^{1/2}\mathbf{N}(0,1)$. Under these assumptions, $\sigma_\varepsilon^2 - \sigma_{\varepsilon z}\Sigma_{zz}^{-1}\sigma_{z\varepsilon} = \sigma_\varepsilon^2$, so that the distribution of J is $\sigma_\varepsilon H^{1/2}\mathbf{N}(0,1)$.

The asymptotics for $\hat{\sigma}_e^2$ are similar to part [a]. Under the null, $\lambda_T = N_T$, so that in $N_T' W^* (W^{*\prime} M_T W^*)^{-1} W^{*\prime} N_T$ is the second term of (18). This term may be rewritten as

$$T^{1/2} N_T' W^* A \nu_T^{-1} (\nu_T^{-1} A' W^{*\prime} T M_T W^* A \nu_T^{-1})^{-1} \nu_T^{-1} A' W^{*\prime} T^{1/2} N_T$$

and it is clear that $\nu_T^{-1} A' W^{*\prime} T M_T W^* A \nu_T^{-1} \to_d H$ as above. The remaining factors, $\nu_T^{-1} A' W^{*\prime} T^{1/2} N_T$ and its transpose, have limits consisting of $\sigma_{z\varepsilon}$ and $\sigma_{\varepsilon z}$ as the first q rows and columns, respectively, with the single remaining element of each being zero. These zeros eliminate the last row and column of H, so that the result is $\hat{\sigma}_e^2 \to_p \sigma_\varepsilon^2 - \sigma_{\varepsilon p} W^* A_0 (A_0' W^{*\prime} \Sigma_{pp} W^* A_0)^{-1} A_0' W^{*\prime} \sigma_{p\varepsilon}$, or more simply σ_ε^2 since $\sigma_{\varepsilon p} = 0$ under these assumptions. The degrees of freedom of the chi-squared limit follows similarly to that in part [a]. □

Appendix B: Proofs of LLN and CLT for Covariances

(This appendix is not intended for publication and is included to support my claim in Lemma A1. I closely follow results and proofs of Phillips and Solo, 1992, pp. 978-980 and 990-993, contributing only by extending their results to covariances, a necessity for the multivariate extension. Phillips and Solo covered autocovariances, but not general covariances.)

A representative element of the matrix polynomial $\Xi_{rs}(z)$ may be written as $c_{rs}(z) = \sum_{i=0}^{\infty} \psi_{jk,i+r} \psi_{uv,i+s} z^k$, so that a representative element of the vector in (14), up to premultiplication by $(I \otimes a')$, is

$$\begin{aligned} \sum\nolimits_t \xi_{\upsilon t} \xi_{ut} = {} & \sum\nolimits_t \left[c_{00}(1)(\upsilon_t u_t - \sigma_{\upsilon u}) + \upsilon_t u_{t-1}^c + u_t \upsilon_{t-1}^c \right] \\ & - \sum\nolimits_t \left[(1-L)\, \zeta_{00,t} + \sum\nolimits_{r=1}^{\infty} (1-L)\, \zeta_{0r,t} + \sum\nolimits_{r=1}^{\infty} (1-L)\, \zeta_{r0,t} \right] \end{aligned} \tag{19}$$

where $\sigma_{\upsilon u} \equiv \mathbf{E}[\upsilon_t u_t]$, $\tilde{c}$ is defined relative to c in the same way as $\tilde{\Xi}$ is defined relative to Ξ, u_t and υ_t are arbitrary scalar elements of $v_{0,t}$, $u_{t-1}^c \equiv \sum_{r=1}^{\infty} c_{0r}(1)\, u_{t-r}$, $\upsilon_{t-1}^c \equiv \sum_{r=1}^{\infty} c_{r0}(1)\, \upsilon_{t-r}$, and $\zeta_{rs,t} \equiv \tilde{c}_{rs}(L)\, \upsilon_{t-r} u_{t-s}$.

The autocovariance structure of $(\xi_{\upsilon t})$ and (ξ_{ut}) defined above is such that $\mathbf{E}[\xi_{\upsilon t} \xi_{u,t-k}] = \sigma_{\upsilon u} c_{k0}(1)$ and $\mathbf{E}[\xi_{ut} \xi_{\upsilon,t-k}] = \sigma_{u\upsilon} c_{0k}(1)$. Variances of (υ_{t-1}^c) and (u_{t-1}^c) are $\sigma_\upsilon^2 \sum_{r=1}^{\infty} c_{r0}^2(1)$ and $\sigma_u^2 \sum_{r=1}^{\infty} c_{0r}^2(1)$. I denote these by $\sigma_{\upsilon c}^2$ and σ_{uc}^2 and note that they are finite by Assumptions [A3] and [A4] and an obvious substitution in the proof of Lemma 3.6 of Phillips and Solo (1992). Similarly, the covariances of these series, $\sigma_{u\upsilon c}^2 = \sigma_{u\upsilon} \sum_{r=1}^{\infty} c_{0r}(1)\, c_{r0}(1)$ is also finite under these assumptions.

The following Lemma is completely analogous to Lemma 5.9 of Phillips and Solo (1992).

Lemma B1. *Under Assumptions [A3]-[A4],* $\mathbf{E}(\sum_{r=1}^{\infty} \zeta_{0r,T})^2, \mathbf{E}(\sum_{r=1}^{\infty} \zeta_{r0,T})^2 < \infty$.

Proof of Lemma B1. I prove only the first result and appeal to symmetry for the second. Note that

$$\mathbf{E}(\sum\nolimits_{r=1}^{\infty} \zeta_{0r,T})^2 = \sum\nolimits_{r=1}^{\infty} \sum\nolimits_{i=0}^{\infty} \sum\nolimits_{r'=1}^{\infty} \sum\nolimits_{i'=0}^{\infty} \tilde{c}_{0r,i} \tilde{c}_{0r',i'} \mathbf{E}[\upsilon_{T-i} \upsilon_{T-i'} u_{T-i-r} u_{T-i'-r'}],$$

which is only non-zero when $r = r'$ and $i = i'$, by the iid assumption and the fact that $r, r' > 0$. Hence, $\mathbf{E}(\sum_{r=1}^{\infty} \zeta_{0r,T})^2 = \sigma_v^2 \sigma_u^2 \sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \tilde{c}_{0r,i}^2$, and $\sigma_v^2 \sigma_u^2 < \infty$ by Assumption [A4], even if $(\upsilon_t) = (u_t)$. The rest of the proof follows that of Lemma 5.9 of Phillips and Solo (1992) by showing that $\sum_{r=1}^{\infty} \sum_{i=0}^{\infty} \tilde{c}_{0r,i}^2 < \infty$ under Assumption [A3]. □

Lemma LLN. *Under Assumptions [A3]-[A4], $T^{-1} \sum_t \xi_{\upsilon t} \xi_{ut} \to_p 0$ as $T \to \infty$.*

Proof of Lemma LLN. The proof requires that the six terms of (19) converge to zero when divided by T. I show that four of them converge and appeal to symmetry for the remaining two. That $T^{-1} \sum_{t=1}^{T} \upsilon_t u_t \to_p \sigma_{\upsilon u}$ is straightforward from an LLN for martingale difference sequences. The proof that $T^{-1} \sum_t \upsilon_t u_{t-1}^c \to_p 0$ follows along the same lines as Phillips and Solo (1992, pg. 990, equation 42). I need only that $\sigma_{\upsilon c}^2 < \infty$, which is shown above. Similarly, $T^{-1} \sum_t \sum_{r=1}^{\infty} (1 - L)\, \zeta_{0r,t} \to_p 0$ follows along the same lines as Phillips and Solo (1992, pg. 990, equation 41) using Lemma B1. Finally, $T^{-1} \sum_t (1 - L)\, \zeta_{00,t} \to_p 0$ also follows along the same lines as Phillips and Solo (1992, pg. 991) under both Assumptions [A3] and [A4]. □

Lemma CLT. *Under Assumptions [A3]-[A4], $T^{-1/2} \sum_t \xi_{\upsilon t} \xi_{ut} \to_d \mathbf{N}(0, \Xi_{\upsilon u})$ as $T \to \infty$, where $\Xi_{\upsilon u} \equiv c_{00}^2(1) \mathbf{E}(\upsilon_t u_t - \sigma_{\upsilon u})^2 + \sigma_\upsilon^2 \sigma_{uc}^2 + \sigma_u^2 \sigma_{\upsilon c}^2 + \sigma_{\upsilon u} \sigma_{u\upsilon c} + \sigma_{u\upsilon} \sigma_{\upsilon uc}$.*

Proof of Lemma CLT. Similarly to the proof of the LLN above, the last three terms of (19) must converge to zero when divided by $T^{1/2}$, but I examine two and appeal to symmetry for the remaining one. The limiting distribution follows from the first three terms of (19).

That $T^{-1/2} \sum_t \sum_{r=1}^{\infty} (1 - L)\, \zeta_{0r,t} \to_p 0$ follows from $\mathbf{E}(\sum_{r=1}^{\infty} \zeta_{0r,T})^2 < \infty$ along the same lines as in the LLN above using Lemma B1. Similarly, that $(1 - L)\, T^{-1/2} \sum_t \zeta_{00,t} \to_p 0$ follows from $\mathbf{E}\zeta_{00,T} < \infty$, which may be verified under Assumptions [A3] and [A4]. The last three terms are thus $o_p\left(T^{1/2}\right)$ as required.

To get the limiting distribution from the first three terms is more involved, but follows along the same lines as Phillips and Solo (1992, pp. 992-3). Defining $Z_t \equiv c_{00}(1)(\upsilon_t u_t - \sigma_{\upsilon u}) + \upsilon_t u_{t-1}^c + u_t \upsilon_{t-1}^c$, I need to verify that (a) $T^{-1} \sum_{t=1}^{T} \mathbf{E}[Z_t^2 1(Z_t^2 > \varepsilon T)] \to 0$ and that (b) $T^{-1} \sum_t Z_t^2 \to_p \mathbf{E}Z_t^2$. I also need an exact representation of $\mathbf{E}Z_t^2$ to get the variance of the limiting Gaussian. As long as $\mathbf{E}Z_t^2 < \infty$, (a) follows from dominated convergence. The bound on $\mathbf{E}Z_t^2$ is confirmed below. Condition (b) requires $T^{-1} \sum_t (u_{t-1}^c)^2 \to_p \sigma_{uc}^2$ and $T^{-1} \sum_t (\upsilon_{t-1}^c)^2 \to_p \sigma_{\upsilon c}^2$ to be finite and time-invariant, which was already shown above.

Expand

$$\begin{aligned} \mathbf{E}Z_t^2 = {} & c_{00}^2(1) \mathbf{E}(\upsilon_t u_t - \sigma_{\upsilon u})^2 + \sigma_\upsilon^2 \sigma_{uc}^2 + \sigma_u^2 \sigma_{\upsilon c}^2 + \sigma_{\upsilon u} \sigma_{u\upsilon c} + \sigma_{u\upsilon} \sigma_{\upsilon uc} \\ & + 2c_{00}(1) \mathbf{E}(\upsilon_t u_t - \sigma_{\upsilon u}) \upsilon_t u_{t-1}^c + 2c_{00}(1) \mathbf{E}(\upsilon_t u_t - \sigma_{\upsilon u}) u_t \upsilon_{t-1}^c \end{aligned} \tag{20}$$

and note that all of the terms in the first line of (20) are finite, as shown above. Both terms in the second line of (20) contain an $\mathcal{F}_t$-measurable series, and it follows that they have an expected value of zero from the law of iterated expectations and the assumption of finite, time-invariant third moments in Assumption [A4]. □

The last step in the generalization is to extend the univariate covariance in Lemma CLT to the multivariate variance Ξ in Lemma A1. The matrix K_{00} consists of elements of the form $\mathbf{E}(\upsilon_{1t}u_{1t}-\sigma_{\upsilon u})(\upsilon_{2t}u_{2t}-\sigma_{\upsilon u})$. The scalar limit theory above considers explicitly the case of $\upsilon_{1t}=\upsilon_{2t}$ and $u_{1t}=u_{2t}$, but there is no loss of generality for different elements $\upsilon_{1t},\upsilon_{2t},u_{1t},u_{2t}$ of the vector v_{0t}. Since $c_{rs}(z)=\sum_{i=0}^{\infty}\psi_{jk,i+r}\psi_{uv,i+s}z^k$ was chosen for arbitrary j,k,u,v, it may represent any element of $\Xi_{rs}(z)\equiv\sum_{k=0}^{\infty}(\Psi_{k+r}\otimes\Psi_{k+s})z^k$. Extending to cross-products with other elements of $\Xi_{rs}(z)$ would simply require $c_{00}(1)d_{00}(1)$ in place of $c_{00}(1)$ in the variance above, where $d_{rs}(z)$ would have exactly the same properties as $c_{rs}(z)$. Thus, the multivariate extension for the first term of $\Xi_{\upsilon u}$ in Lemma CLT has a variance of $\Xi_{00}(1)K_{00}\Xi_{00}'(1)$.

The second and third terms of $\Xi_{\upsilon u}$ in Lemma CLT are $\sigma_\upsilon^2\sigma_u^2\sum_{r=1}^{\infty}c_{0r}^2(1)$ and $\sigma_u^2\sigma_\upsilon^2\sum_{r=1}^{\infty}c_{r0}^2(1)$ using the variances preceding Lemma B1. The multivariate extension of these terms to $\sum_{r=1}^{\infty}\Xi_{0r}(1)K_{0\bullet}\Xi_{0r}'(1)+\Xi_{r0}(1)K_{0\bullet}\Xi_{r0}'(1)$ follows along the same lines as the first term. For $\upsilon_{1t}\neq\upsilon_{2t}$ and $u_{1t}\neq u_{2t}$, covariances would replace the variances σ_υ^2 and σ_u^2 for cross-products, but these are finite since the variances are finite. They are incorporated in $K_{0\bullet}$ by its definition.

Finally, the last terms of $\Xi_{\upsilon u}$ in Lemma CLT may be written as

$$\sigma_{\upsilon u}\sigma_{u\upsilon}\sum_{r=1}^{\infty}c_{0r}(1)\,c_{r0}(1)+\sigma_{u\upsilon}\sigma_{\upsilon u}\sum_{r=1}^{\infty}c_{r0}(1)\,c_{0r}(1)$$

using the covariances preceding Lemma B1. The multivariate extension simply requires $d_{rs}(z)$ again and different (finite) covariances in place of $\sigma_{\upsilon u}$ and $\sigma_{u\upsilon}$ for $\upsilon_{1t}\neq\upsilon_{2t}$ and $u_{1t}\neq u_{2t}$. All of these covariances are incorporated in $K_{\bullet 0}$.

References

Almon, S. (1965). "The Distributed Lag Between Capital Appropriations and Expenditures," *Econometrica* 33, 178-196.

Alper, C.E., S. Fendoglu, and B. Saltoglu (2008) "Forecasting Stock Market Volatilities Using MIDAS Regressions: An Application to the Emerging Markets," MPRA Paper No. 7460.

Andreou, E., E. Ghysels, and A. Kourtellos (2009). "Should Macroeconomic Forecasters Use Daily Financial Data and How?" Department of Economics, University of North Carolina, mimeo.

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

Andreou, E., E. Ghysels, and A. Kourtellos (2010b). "Forecasting with Mixed-frequency Data," Department of Economics, University of North Carolina, mimeo.

Armesto, M.T., R. Hernández-Murillo, M.T. Owyang, and J. Piger (2009). "Measuring the Information Content of the Beige Book: A Mixed Data Sampling Approach," *Journal of Money, Credit and Banking* 41, 35-55.

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.

Bai, J., E. Ghysels, and J.H. Wright (2010). "State Space Models and MIDAS Regressions," Department of Economics, University of North Carolina, mimeo.

Barsky, R.B. and L. Kilian (2004). "Oil and the Macroeconomy Since the 1970's," *Journal of Economic Perspectives* 18, 115-134.

Benjanuvatra, S. (2009). "A Real-Time Comparison of Mixed-Frequency Models for Predicting GDP," University of York, mimeo.

Beveridge, S. and C.R. Nelson (1981). "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics* 7, 151-174.

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

Chambers, M.J. (2009). "Discrete Time Representations of Cointegrated Continuous-Time Models with Mixed Sample Data," *Econometric Theory* 25, 1030-1049.

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

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

Chang, Y., J.Y. Park, and P.C.B. Phillips (2001). "Nonlinear Econometrics Models with Cointegrated and Deterministically Trending Regressors," *The Econometrics Journal* 4, 1-36.

Chang, Y., J.I. Miller, and J.Y. Park (2009). "Extracting a Common Stochastic Trend: Theory with Some Applications," *Journal of Econometrics* 150, 231-247.

Chen, X. and E. Ghysels (2010). "News – Good or Bad – and Its Impact on Volatility Predictions over Multiple Horizons," *The Review of Financial Studies* 24, 46-81.

Christofferson, P.F. and F.X. Diebold (1998). "Cointegration and Long-Horizon Forecasting," *Journal of Business and Economic Statistics* 16, 450-458.

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.

Clements, M.P. and D.F. Hendry (1995). "Forecasting in Cointegrated Systems," *Journal of Applied Econometrics* 10, 127-146.

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

Frale, C. and L. Montefort (2010). "FaMIDAS: A Mixed Frequency Factor Model with MIDAS structure," Working Paper 788, Banca D'Italia.

Francis, N.R., E. Ghysels, and M.T. Owyang (2010). "The Low-Frequency Impact of Daily Monetary Policy Shocks," Working Paper 2011-009A, Federal Reserve Bank of St. Louis.

Ghysels, E., A. Rubia, and R. Valkanov (2009). "Multi-period Forecasts of Volatility: Direct, Iterated, and Mixed-data Approaches," Department of Economics, University of North Carolina, mimeo.

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

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). "There Is a Risk-return Tradeoff After All," *Journal of Financial Economics* 76, 509-548.

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

Ghysels, E., A. Sinko, and R. Valkanov (2007). "MIDAS Regressions: Further Results and New Directions," *Econometric Reviews* 26, 53-90.

Ghysels, E. and R. Valkanov (2010). "Forecasting Volatility with MIDAS," in *Volatility Models and Their Applications*, L. Bauwens, C. Hafner, and S. Laurent, eds. New York: Wiley, 1-23.

Ghysels, E. and J. Wright (2009). "Forecasting Professional Forecasters," *Journal of Business and Economic Statistics*, 27, 504-516.

Godfrey, L.G. and D.S. Poskitt (1975). "Testing the Restrictions of the Almon Lag Technique," *Journal of the American Statistical Association* 70, 105-108.

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

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

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

Hamilton, J.D. (2009). "Causes and Consequences of the Oil Shock of 2007-08," *Brookings Papers on Economic Activity* 1, Spring, 215-261.

Hansen, B.E. (2010). "Averaging Estimators for Autoregressions with a Near Unit Root," *Journal of Econometrics* 158, 142 155.

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

He, Y., S. Wang, and K.K. Lai (2010). "Global Economic Activity and Crude Oil Prices: A Cointegration Analysis," *Energy Economics* 32, 868-876.

Hogrefe, J. (2008). "Forecasting Data Revisions of GDP: A Mixed Frequency Approach," *Advances in Statistical Analysis* 92, 271-296.

Kilian, L. (2008). "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *The Review of Economics and Statistics* 90, 216-240.

Kilian, L. (2009). "Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market," *American Economic Review* 99, 1053-1069.

Klovland, J.T. (2002). "Business Cycles, Commodity Prices, and Shipping Freight Rates: Some Evidence from the Pre-WWI Period," SNF Report 48/02, Norwegian School of Economics and Business Administration.

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.

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

Maddison, A. (2010). "Statistics on World Population, GDP and Per Capita GDP, 1-2008 AD," Groningen Growth and Development Centre, Groningen University, http://www.ggdc.net/Maddison.

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

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

Miller, J.I. and S. Ni (2011). "Long-Term Oil Price Forecasts: A New Perspective on Oil and the Macroeconomy," *Macroeconomic Dynamics*, forthcoming.

Montefort, L. and G. Moretti (2010). "Real Time Forecasts of Inflation: The Role of Financial Variables," Working Paper 767, Banca D'Italia.

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

Park, J.Y. and P.C.B. Phillips (1988). "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory* 4, 468-497.

Park, J.Y. and P.C.B. Phillips (1989). "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory* 5, 95-131.

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

Phillips, P.C.B. and V. Solo (1992). "Asymptotics for Linear Processes," *Annals of Statistics* 20, 971-1001.

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

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

Tay, A.S. (2007). "Financial Variables as Predictors of Real Output Growth," School of Economics and Social Sciences, Singapore Management University, mimeo.

Wooldridge, J.M. (1994). "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, Vol. IV, R.F. Engle and D.L. McFadden, eds. Amsterdam: Elsevier, 2639-738.

Wu, D.-M. (1973). "Alternative Tests of Independence between Stochastic Regressors and Disturbances," *Econometrica* 41, 733-750.