Cointegration in a MIDAS Regression

H. Peter Boswijk*
Amsterdam School of Economics, University of Amsterdam,
and Tinbergen Institute

Philip Hans Franses
Erasmus School of Economics and Erasmus School of Law,
Erasmus University Rotterdam, and Tinbergen Institute

December 16, 2024

Abstract

Cointegration analysis in a mixed data sampling (MIDAS) regression often assumes that the underlying data generating process (DGP) also is a MIDAS model, that is, the variables are observable at different frequencies. In this paper we provide an alternative look at cointegration analysis in MIDAS regressions by assuming that the underlying DGP is a model at the high frequency for both the dependent and the independent variables, but that somehow only low-frequency observations of the dependent variable are available. We discuss the representation of the MIDAS model, given an autoregressive distributed lag model of order (1,1) at the high frequency, for any degree of aggregation of the dependent variable. We discuss a test for cointegration, which includes knowledge of the specific form of the implied moving average process in the MIDAS model, and we derive its associated asymptotic distribution. With simulations we examine the empirical performance of the test. We illustrate using quarterly total inflation as a function of monthly energy prices changes in the Netherlands.

Key words: Cointegration; Temporal Aggregation; MIDAS Regression
JEL codes: C22, C51

1 Introduction

MIDAS regressions involve a dependent variable observed at a low frequency and independent variables observed at a high frequency. A MIDAS regression (Ghysels *et al.*, 2007) often also includes lags of the low-frequency dependent variable on the right-hand side. An empirical example is quarterly growth in Gross Domestic Product (GDP) as the dependent variable and the monthly observed growth in industrial production as an explanatory variable.

*Corresponding author. Address: Amsterdam School of Economics, University of Amsterdam, PO Box 15867, 1001 NJ Amsterdam, The Netherlands. E-mail: `h.p.boswijk@uva.nl`

Cointegration analysis in a MIDAS regression often assumes that the underlying data generating process (DGP) also is a MIDAS model, that is, the variables are observable at different frequencies (see e.g. Ghysels *et al.*, 2007, Götz *et al.*, 2014, and Ghysels, 2016). In this paper, following Ghysels and Miller (2015), we provide an alternative look at cointegration analysis in MIDAS regressions by assuming that the underlying DGP is a model at the high frequency for both the dependent and the independent variables, but that somehow only low-frequency observations of the dependent variable are available. To illustrate, we consider two inflation variables, where we assume quarters as the low frequency of total inflation and months as the high frequency of energy prices changes.

In this paper we focus on an autoregressive distributed lag (ADL) model of order (1,1) at the highest frequency for two variables. We derive how cointegration at the high frequency translates to cointegration in the MIDAS regression. One result is that the MIDAS model has a moving average (MA) error, as was also observed in vector error correction models (VECMs) by Ghysels and Miller (2015). For this specific ADL(1,1) model we derive the explicit expression of the first-order autocorrelation coefficient of this MA(1) process. Next, we show that under the null hypothesis of no cointegration, this autocorrelation coefficient is a function of the aggregation factor. We also derive that when the degree of aggregation increases, the autocorrelation coefficient increases, converging to $\frac{1}{4}$.

As the MA(1) process is known under the null hypothesis of no cointegration, we incorporate it in our test for cointegration. We derive the relevant asymptotic theory, and with simulation experiments we show that the empirical size of the test matches the theory. With further simulation experiments we demonstrate that including knowledge of the MA(1) parameter increases the power of the test for cointegration, relative to tests where the MA term is ignored.

The outline of our paper is as follows. In Section 2 we discuss the representation of the MIDAS model, given the ADL(1,1) model at the high frequency, for any degree of aggregation of the dependent variable. In Section 3, we discuss the test for cointegration and its associated asymptotic distribution. With simulations we examine the empirical performance of the test. In Section 4 we illustrate the test for quarterly total inflation as a function of monthly energy prices changes. Section 5 concludes with some avenues for further research. An appendix contains mathematical proofs of the main results.

2 Model and representation

We consider a (partially latent) bivarate time series $\{(y_t, x_t)', t = 1, \ldots, n\}$, generated by

$$y_t = \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t, \quad (1)$$

$$x_t = x_{t-1} + \eta_t, \quad (2)$$

where $\{\varepsilon_t\}$ and $\{\eta_t\}$ are mutually uncorrelated white noise processes with variances σ_ε^2 and σ_η^2, respectively. The system consists of an ADL(1,1) model for y_t, conditional on x_t (1), combined with a random walk specification (2) for the exogenous variable x_t. If $|\alpha| < 1$ and $\beta_0 + \beta_1 \neq 0$, then the system implies that $(y_t, x_t)'$ is cointegrated of order (1,1), with cointegrating vector $(1, \gamma)'$, where $\gamma = (\beta_0 + \beta_1)/(1 - \alpha)$ is the long-run effect of x on y. The ADL equation can then be rewritten in error-correction model (ECM) form:

$$\Delta y_t = \beta_0 \Delta x_t + (\alpha - 1)(y_{t-1} - \gamma x_{t-1}) + \varepsilon_t. \tag{3}$$

On the other hand, if $\alpha = 1$ and $\beta_0 + \beta_1 = 0$, then the error correction term vanishes in (3), and y_t and x_t are integrated of order 1 but not cointegrated. The system can easily be generalized to allow for a vector of exogenous variables x_t, for higher-order lags of y_t and x_t (leading to an ADL(p,q) with $p > 1$ and/or $q > 1$), and for the presence of a constant and, if relevant, a linear trend term in (1)–(2). For notational convenience, results will be explicitly derived for the basic case, after we will discuss how these results generalize to such empirically relevant cases.

The time index $t = 1, 2, \ldots$ indicates the highest possible observation frequency. We consider the case where the time series x_t is observed at this highest frequency, but for y_t we only observe a temporally aggregated version at a lower frequency. More specifically, we will work under the assumption that we observe the average of y over m periods, at times t that are a multiple of m. It will be convenient to introduce the low-frequency time index $T = 1, \ldots, N$, and to denote the low-frequency aggregated observations by $Y_T = m^{-1} \sum_{i=0}^{m-1} y_{mT-i}$. We will assume that the number n of high-frequency observations is a multiple of m, such that $N = n/m$. For example, our empirical application considers the case where monthly observations on energy price changes are related to quarterly observed inflation, such that $m = 3$, and our sample period consists of n months and hence $N = n/3$ quarters.[1] Following Ghysels (2016) and Ghysels and Miller (2015), we will combine the low-frequency observations on y with the high-frequency observations on x by considering the stacked vector process

$$\mathbb{X}_T = \begin{pmatrix} X_{1,T} \\ \vdots \\ X_{m,T} \end{pmatrix} = \begin{pmatrix} x_{mT-(m-1)} \\ \vdots \\ x_{mT} \end{pmatrix}, \qquad T = 1, \ldots, N,$$

with $\mathbb{Y}_T = (Y_{1,T}, \ldots, Y_{m,T})'$ defined analogously, but observing only $Y_T = m^{-1} \sum_{i=1}^{m} Y_{i,T}$. The full set of observations therefore is given by $(Y_T, X_{1,T}, \ldots, X_{m,T})', T = 1, \ldots, N$ (in addition to starting values Y_0 and $X_{i,0}$).

Proposition 1 derives an ECM representation for Y_T conditional on $(X_{1,T}, \ldots, X_{m,T})$, which will be the basis for the inference procedures developed in the next section. Before presenting the result for general integer m, we provide some intuition for the result and its derivation in the specific case that $m = 2$. Given (1), the starting point is the implied vector autoregressive (VAR) representation for $\mathbb{Y}_T$ with exogenous variables $\mathbb{X}_T$, in simultaneous-equations form:

$$\begin{aligned} \begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix} \begin{pmatrix} Y_{1,T} \\ Y_{2,T} \end{pmatrix} &= \begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Y_{1,T-1} \\ Y_{2,T-1} \end{pmatrix} + \begin{pmatrix} \beta_0 & 0 \\ \beta_1 & \beta_0 \end{pmatrix} \begin{pmatrix} X_{1,T} \\ X_{2,T} \end{pmatrix} \\ &\quad + \begin{pmatrix} 0 & \beta_1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X_{1,T-1} \\ X_{2,T-1} \end{pmatrix} + \begin{pmatrix} E_{1,T} \\ E_{2,T} \end{pmatrix}, \end{aligned}$$

[1] In the MIDAS literature, usually the low-frequency observations are indicated by x_t, and hence high-frequency observations by $x_{t-i/m}, i = 0, \ldots, m-1$. Our notation avoids such fractional lags, but is otherwise equivalent.

where $E_{1,T} = \varepsilon_{2T-1}$ and $E_{2,T} = \varepsilon_{2T}$. By inverting the left-hand side matrix (and hence obtaining the reduced-form VAR representation), and subsequenty pre-multiplying the system by the averaging vector $(\frac{1}{2}, \frac{1}{2})$, we obtain the following equation for Y_T:

$$
\begin{aligned}
Y_T &= \tfrac{1}{2}(\alpha + \alpha^2)Y_{2,T-1} + \tfrac{1}{2}\left((1+\alpha)\beta_0 + \beta_1\right)X_{1,T} + \tfrac{1}{2}\beta_0 X_{2,T} + \tfrac{1}{2}(1+\alpha)\beta_1 X_{2,T-1} \\
&\quad + \tfrac{1}{2}\left((1+\alpha)E_{1,T} + E_{2,T}\right).
\end{aligned}
$$

The right-hand side variable $Y_{2,T-1}$ is unobserved, but can be related to the observed lags (and an unobserved lagged error) via

$$
\begin{aligned}
\tfrac{1}{2}(\alpha + \alpha^2)Y_{2,T-1} &= \tfrac{1}{2}\alpha^2(Y_{1,T-1} + Y_{2,T-1}) + \tfrac{1}{2}\alpha(Y_{2,T-1} - \alpha Y_{1,T-1}) \\
&= \alpha^2 Y_{T-1} + \tfrac{1}{2}\alpha(\beta_0 X_{2,T-1} + \beta_1 X_{1,T-1} + E_{2,T-1}), \qquad (4)
\end{aligned}
$$

which after substitution and rearrangement of terms leads to the final ECM:

$$
\begin{aligned}
\Delta Y_T &= (\alpha^2 - 1)Y_{T-1} + \delta X_{T-1} + \beta_{01}\Delta X_{1,T} + \beta_{02}\Delta X_{2,T} + U_T \\
&= (\alpha^2 - 1)(Y_{T-1} - \gamma X_{T-1}) + \beta_{01}\Delta X_{1,T} + \beta_{02}\Delta X_{2,T} + U_T, \qquad (5)
\end{aligned}
$$

where Δ is the low-frequency differencing operator (such that, e.g., $\Delta X_{1,T} = X_{1,T} - X_{1,T-1}$), where

$$
\delta = (1+\alpha)(\beta_0 + \beta_1), \qquad \gamma = \frac{\beta_0 + \beta_1}{1-\alpha}, \qquad \beta_{01} = \tfrac{1}{2}\left[\beta_0(1+\alpha) + \beta_1\right], \qquad \beta_{02} = \tfrac{1}{2}\beta_0,
$$

and where $U_T = \frac{1}{2}\left((1+\alpha)E_{1,T} + E_{2,T} + \alpha E_{2,T-1}\right)$. We observe that $\mathrm{cov}(U_T, U_{T-1}) = \frac{1}{4}\alpha\sigma_\varepsilon^2$, implying a first-order moving average (MA(1)) structure in the error of (5); this was also observed for MIDAS-VECM models by Ghysels and Miller (2015). This will be addressed in our inference procedures in the next section.

The MIDAS-ECM result for general m is presented next; the proof of this proposition is given in the appendix.

Proposition 1 *Let $\{(y_t, x_t)'\}_{t=1}^n$ be generated by (1)–(2), let m be a positive integer, and define $Y_T = m^{-1}\sum_{i=0}^{m-1} y_{mT-i}$, $X_T = m^{-1}\sum_{i=0}^{m-1} x_{mT-i}$ and $X_{i,T} = x_{mT-(m-i)}$ for $i = 1, \ldots, m$, $T = 1, \ldots, N = n/m$. Then Y_T admits the following ECM representation*

$$
\begin{aligned}
\Delta Y_T &= (\alpha^m - 1)Y_{T-1} + \delta X_{T-1} + \sum_{i=1}^{m} \beta_{0i}\Delta X_{i,T} + U_T \\
&= (\alpha^m - 1)(Y_{T-1} - \gamma X_{T-1}) + \sum_{i=1}^{m} \beta_{0i}\Delta X_{i,T} + U_T, \qquad (6)
\end{aligned}
$$

where[2]

$$
\delta = \sum_{i=0}^{m-1} \alpha^i(\beta_0 + \beta_1), \qquad \gamma = \frac{\beta_0 + \beta_1}{1-\alpha}, \qquad \beta_{0i} = \frac{1}{m}\left(\beta_0 + (\alpha\beta_0 + \beta_1)\sum_{j=1}^{m-i}\alpha^{j-1}\right), \qquad (7)
$$

[2] In (7), we use the convention that an empty summation of the form $\sum_{j=1}^{0}$ equals zero, so that $\beta_{0m} = \beta_0/m$.

and where

$$U_T = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=0}^{m-i}\alpha^j\right)E_{i,T} + \frac{1}{m}\sum_{i=2}^{m}\left(\sum_{j=m-i+1}^{m-1}\alpha^j\right)E_{i,T-1}.$$

Therefore, U_T is an MA(1) process with first-order autocorrelation coefficient

$$\rho = \frac{\sum_{i=2}^{m}\left(\sum_{j=0}^{m-i}\alpha^j\right)\left(\sum_{j=m-i+1}^{m-1}\alpha^j\right)}{\sum_{i=1}^{m}\left(\sum_{j=0}^{m-i}\alpha^j\right)^2 + \sum_{i=2}^{m}\left(\sum_{j=m-i+1}^{m-1}\alpha^j\right)^2}. \tag{8}$$

If $\alpha = 1$ and $\beta_0 + \beta_1 = 0$, then the error correction term vanishes from (6), and $\beta_{0i} = \beta_0/m$, such that the model reduces to

$$\Delta Y_T = \beta_0 \Delta X_T + U_T.$$

The first-order autocorrelation coefficient then becomes $\rho = (m^2-1)/(4m^2+2)$.

Writing the MA(1) process $U_T = \xi_T + \theta\xi_{T-1}$, where ξ_T is a white noise process with variance σ_ξ^2, the MA(1) coefficient θ can be obtained from the expression for ρ in Proposition 1 from

$$\theta = \frac{1-\sqrt{1-(2\rho)^2}}{2\rho}. \tag{9}$$

For example, for $m = 3$ and $\alpha = 0.5$, the implied MA parameter is $\theta = \frac{1}{6} = 0.167$. For $m = 3$ and $\alpha = 1$, this increases to $\theta = 0.221$. A higher value of m will lead to a higher coefficient θ, with a maximum value of $\theta = 2 - \sqrt{3} = 0.268$, corresponding to $\rho = \frac{1}{4}$, for $\alpha = 1$ and as $m \to \infty$. Although these MA effects may not be very strong, ignoring them in subsequent analysis will lead to distorted inference on cointegration, and in particular to size distortions in cointegration tests when using standard critical values; see, e.g., Boswijk and Franses (1992). Therefore, in the next section we develop statistical procedures taking into account MA effects, with the MA coefficient restricted to the value implied by α, which defines this paper's main contribution to the MIDAS cointegration literature.

3 Testing for cointegration

To obtain a test statistic for cointegration based on the model with an MA(1) term satisfying the restriction implied by Proposition 1, the starting point is the Gaussian conditional log-likelihood:

$$\ell(\alpha, \beta_0, \beta_1, \sigma_\xi^2) = -\frac{N}{2}\log(2\pi\sigma_\xi^2) - \frac{1}{2\sigma_{\xi^2}}\sum_{T=1}^{N}\xi_T(\alpha,\beta_0,\beta_1)^2, \tag{10}$$

where

$$\xi_T(\alpha,\beta_0,\beta_1) = (1+\theta(\alpha)L)^{-1}U_T(\alpha,\beta_0,\beta_1), \tag{11}$$

$$U_T(\alpha,\beta_0,\beta_1) = Y_T - \alpha^m Y_{T-1} - \delta X_{T-1} - \sum_{i=1}^{m}\beta_{0i}\Delta X_{i,T}, \tag{12}$$

with $\theta(\alpha)$ as implied by (8)–(9), and using the parameter definitions in (7). We use the conditional least-squares approach for the MA term, initalising $\xi_T = (1+\theta L)^{-1}U_T$ at zero, such that $\xi_T =$

$\sum_{i=0}^{T-1}(-\theta)^i U_{T-i}$. Note that σ_ξ^2 could be expressed in terms of σ_ε^2 and α, but this imposes no restriction on the parameters as long as σ_ε^2 varies freely; so we may parametrize the model in terms of σ_ξ^2 instead of σ_ε^2.

The likelihood ratio test statistic for $H_0 : \alpha = 1, \beta_0 + \beta_1 = 0$, then becomes

$$\mathrm{LR} = N \log \frac{\min_{\beta_0} \sum_{T=1}^{N} \xi_T(1, \beta_0, -\beta_0)^2}{\min_{\alpha,\beta_0,\beta_1} \sum_{T=1}^{N} \xi_T(\alpha, \beta_0, \beta_1)^2}. \tag{13}$$

The unrestricted minimization problem over α, β_0 and β_1 in the denominator requires numerical optimization. Because $U_T(1, \beta_0, -\beta_0) = \Delta Y_T - \beta_0 \Delta X_T$, the restricted estimator of β_0 can be simply obtained by a least-squares regression of $(1 + \theta(1)L)^{-1}\Delta Y_T$ on $(1 + \theta(1)L)^{-1}\Delta X_T$.

The large-sample asymptotic properties of LR are obtained in the next proposition.

Proposition 2 *Let $\{(y_t, x_t)'\}_{t=1}^n$ be generated by (1)–(2), with $\{(\varepsilon_t, \eta_t)'\}_{t=1}^n$ satisfying an invariance principle. Let* LR *be as defined in (13), using the notation defined in Section 2. Then as $N \to \infty$, with m fixed, and under $H_0 : \alpha = 1, \beta_0 + \beta_1 = 0$,*

$$\mathrm{LR} \xrightarrow{d} \int_0^1 dW_1(u)W(u)' \left(\int_0^1 W(u)W(u)'du \right)^{-1} \int_0^1 W(u)dW_1(u),$$

where $W(u) = (W_1(u), W_2(u))', u \in [0, 1]$, is a bivariate standard Brownian motion on $[0, 1]$. Under $H_1 : |\alpha| < 1$, $\mathrm{LR} = O_p(N)$.

We find that under the null hypothesis, LR has the same limiting distribution as the Wald test statistic, proposed by Boswijk and Franses (1992) and Boswijk (1994), would have in the absence of moving average errors (i.e., when applied to the original high-frequency data). Accurate asymptotic critical values and p-value functions for the limiting null distribution were obtained by MacKinnon *et al*. (1999). The proposition also implies that under the alternative hypothesis of cointegration, the test rejecting for large values of LR is consistent. Under this alternative, we may expect that inference on the long-run cointegration parameter γ, based on the (correctly specified) log-likelihood (10), will be asymptotically mixed normal (so that t-test statistics for hypotheses on γ may be compared with conventional standard normal critical values).

Proposition 2 does not explicitly characterize the asymptotic null distribution of the Boswijk (1994) Wald test statistic in the MIDAS regression (6), when no correction for moving average errors is applied. Analogously to the results for other cointegration tests in Theorems 4 and 5 of Ghysels and Miller (2015), we expect this null distribution to display bias terms and nuisance parameter dependencies, such that the use of the MacKinnon *et al*. (1999) critical values would result in size distortions.

The cointegration test proposed here can be generalized in various directions. First, the case where x_t is a vector-valued integrated (but not cointegrated) process, rather than a univariate random walk, is easily handled, with minor changes (in that case β_0, β_1, β_{0i}, γ and δ all become vectors of the same dimension k, and W becomes a $(1 + k)$-dimensional standard Brownian motion process). Secondly, allowing for a constant and possibly a linear trend term in the high-frequency DGP (1) will lead to corresponding extensions in (6); the resulting LR statistic should then be compared to critical values from the appropriate tables for such specifications of the deterministic components. Finally, in empirical applications we may need to allow for a lag length higher than 1 in the DGP. In that case, the same approach as in the previous section may be followed to obtain the corresponding MIDAS error correction model as a generalization of (6). However, the functional form of the coefficients of lagged differences and of the MA parameters would have to be derived for the specific case at hand.

Monte Carlo simulations of size and power

We end this section with a Monte Carlo simulation exercise to investigate the extent of the finite-sample size distortions of the Wald test in MIDAS ECM regressions, and of the effectivess of our proposed LR test to eliminate these size distortions. In addition, we investigate the effect of the degree of time aggregation, m, on the power of the tests.

As the DGP under the null hypothesis, we take (1)–(2) with $\alpha = 1, \beta_0 = 0.2$ and $\beta_1 = -0.2$, and with $(\varepsilon_t, \eta_t)'$ as an i.i.d. $N(0, I_2)$ sequence. We consider two (high-frequency) sample sizes $n \in \{120, 480\}$, meant to represent either 10 years or 40 years of monthly observations, and we consider the cases $m = 1$ (no aggregation), $m = 3$ (aggregation of y_t to quarterly averages) and $m = 6$ (aggregation to half-yearly averages). To investigate the power of the tests, we consider $\alpha = 0.9, \beta_0 = 0.2$ and $\beta_1 = -0.1$ for the case $n = 120$, and $\alpha = 0.975, \beta_0 = 0.2$ and $\beta_1 = -0.175$ for the larger sample size $n = 480$. This effectively means that we study local power, i.e., we take $\alpha = 1 - 12/n$ and choose β_1 such that $\gamma = (\beta_0 + \beta_1)/(1 - \alpha) = 1$ in all cases. We only use asymptotic critical values (which are invalid for the Wald test when $m > 1$), hence the Monte Carlo power is not size-corrected. Note that for a given n, increasing m implies decreasing the number of low-frequency observations $N = n/m$. For $m = 3$ and $m = 6$, we compare the size and power performance of the Wald test and our LR test (which explicitly incorporates the known MA component); for $m = 1$, the two testing principles effectively coincide, which is why we do not report any separate results for LR in that case. We perform two versions of all tests: with no deterministic components (denoted Wald and LR), and with an unrestricted intercept in the error correction model (denoted Wald_μ and LR_μ). All results are based on $100,000$ Monte Carlo replications.

From Table 1 we observe that the size distortions of the Wald test, caused by autocorrelation induced by aggregation, are limited, in particular when an intercept is included in the regression. If the test is performed with no intercept, there is more evidence of size distortions increasing with m (and hence with the autocorrelation coefficient ρ), in particular with the smaller sample size. In general, the size control of our LR test is good, with a worst case empirical size of 0.064 (corresponding to $n = 120$ and $m = 6$, hence aggregation to $N = 20$ half-yearly observations). The differences in the two testing approaches are more pronounced when considering test power. We observe that for a given sample size n and DGP, the power of the Wald test clearly decreases with m, whereas the power of the LR test is much less affected by the use of aggregated observations for y_t. Again, this effect is most noticeable in the smaller sample size. Although we have not provided a theoretical comparison of asymptotic local power of the two approaches, these simulation results suggest a considerable power gain from accounting for moving average effects in MIDAS ECM regressions.

Table 1: Monte Carlo size and power of MIDAS cointegration tests

Size

	$n = 120$				$n = 480$			
m	Wald	LR	Wald_μ	LR_μ	Wald	LR	Wald_μ	LR_μ
1	0.053		0.054		0.051		0.050	
3	0.065	0.056	0.048	0.048	0.059	0.051	0.041	0.049
6	0.075	0.064	0.051	0.048	0.061	0.053	0.040	0.046

Power

	$n = 120$				$n = 480$			
m	Wald	LR	Wald_μ	LR_μ	Wald	LR	Wald_μ	LR_μ
1	0.784		0.562		0.774		0.544	
3	0.567	0.729	0.362	0.503	0.657	0.761	0.422	0.530
6	0.287	0.654	0.183	0.449	0.571	0.743	0.347	0.512

Note: This table provides Monte Carlo simulation results for the Wald test for cointegration proposed by Boswijk (1994), and the LR test accounting for moving average effects in MIDAS regressions proposed in the present paper. Tests allowing for an unrestricted intercept are denoted Wald_μ and LR_μ. The high-frequency sample size is denoted by n, and the number of periods used for aggregation of the dependent variable is indicated by m. The top panel ("*Size*") shows rejection frequencies of the tests under $H_0 : \alpha = 1, \beta_0 + \beta_1 = 0$, using asymptotic 5% critical values. The bottom panel ("*Power*") shows rejection frequencies of the tests using local alternatives $\alpha = 1 - 12/n, \beta_0 + \beta_1 = 1 - \alpha$, using the same asymptotic 5% critical values. The number of Monte Carlo replications is $100,000$.

4 Empirical application

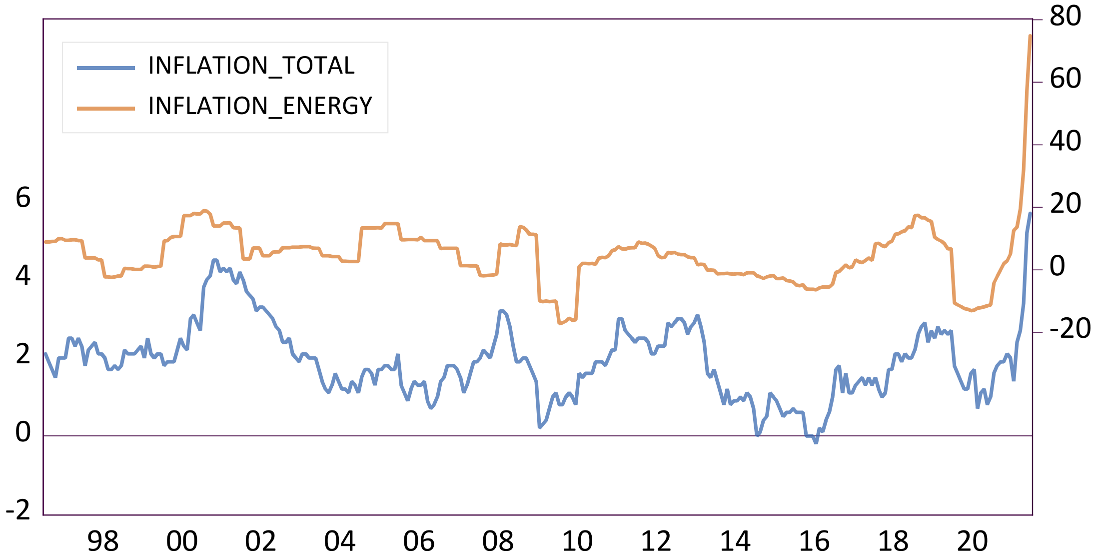
In this section we illustrate our approach to testing for cointegration in a MIDAS regression. We consider the monthly observations on total inflation and the monthly observations for inflation in energy prices (both year-on-year percentages) in the Netherlands[3] for the sample 1997M1 to 2021M12, see Figure 1. Obviously, the two variables show co-movement, and they can serve as an illustration. We denote total inflation as y_t and inflation in energy prices as x_t.

When we consider the regression in (1), where we include an intercept, we obtain for α, β_0 and β_1 the estimates 0.928, 0.048 and -0.042, respectively. The Wald test for the joint null hypothesis $\alpha = 1, \beta_0 + \beta_1 = 0$ has a value of 12.01. Comparing this with the 5% asymptotic critical value of 11.42, see Table IV in MacKinnon *et al.* (1999), we obtain evidence of the presence of cointegration amongst the inflation series at the 5% significance level (the p-value, computed using the procedures in MacKinnon *et al.* (1999), is 0.040). The cointegrating parameter is estimated as 0.086.

Next, we consider the situation where the total inflation series are aggregated to quarterly averages,

[3] CPI data from Statistics Netherlands, CBS Open data StatLine, have been obtained via `https://opendata.cbs.nl/statline/portal.html?_la=en&_catalog=CBS&tableId=83131ENG&_theme=1155`, selecting the expenditure categories "`000000 All items`" and "`045000 Electricity, gas and other fuels`".

Figure 1: Total inflation and inflation in energy prices in the Netherlands, 1997M1–2021M12



Data source: Statistics Netherlands.

so $m = 3$. When we consider the error correction model in (6), with the inclusion of an intercept, and we do not incorporate an MA(1) term, we obtain a Wald test statistic for the joint null hypothesis $\alpha = 1, \beta_0 + \beta_1 = 0$ equal to 8.06, with an asymptotic p-value of 0.168, indicating no evidence of cointegration at conventional significance levels. We see that this Wald test provides evidence of cointegration only at the 20% significance level (with a critical value of 7.53).

Evidence in favour of cointegration ever further diminishes when we consider the model in (6) with an intercept but with an unrestricted MA(1) term. The relevant Wald statistic is now 6.20. However, when we consider model (6) with an intercept and with the proper restriction on the MA(1) term imposed in estimation, the relevant likelihood ratio test statistic becomes equal to 10.63. With a p-value of 0.067, we now find significant evidence of cointegration at the 10% level (corresponding to a critical value of 9.54). The difference between the first and the third test, without accounting for the MA effect and with the restriction imposed on the MA(1) parameter, respectively, associates with the simulation results in the previous section, which showed an increase in power when the MA(1) term is properly restricted. In the last case, we obtain for α, β_0 and β_1 the estimates 0.932, 0.014 and -0.008, respectively, which indicates that the cointegrating parameter is estimated as 0.088, which is rather like the case without temporal aggregation.

5 Discussion

The basic premise in our cointegration analysis in a MIDAS regression model is that there is a high-frequency data generating process where the variables are observed at the same frequency but that the dependent variable is actually observed at a lower frequency. Given this data generating process, we derived the associated MIDAS regression, and we showed that the long-run cointegration parameter is preserved and that the MIDAS regression contains a moving average term. We also showed that under the null hypothesis of no cointegration, the parameter in the moving average term is known. We incorporated this knowledge in our test for cointegration in the MIDAS model and provided the relevant asymptotic theory for the case of two related variables. With simulations we demonstrated that our test has proper size and improved power over the situation where the MA term is ignored. An illustration to energy inflation and total inflation showed the merits of our approach.

A natural extension of our methodology involves more than two variables, while at the same time higher order autoregressive models can also be addressed. Basically, we expect that these extensions lead to similar insights as for the bivariate first order case, which is that we can learn from cointegration analysis of the MIDAS model what are the long-run and short-run adjustment parameters in the associated high-frequency model.

A particularly useful application of our method could concern nowcasting important macroeconomic variables based on high-frequency independent variables. In various countries in the world the statistics bureaus compile for example GDP data at a low frequency, perhaps even as low as only annually, whereas there is abundant high-frequency data available that can be scraped from internet sources. As such, our methodology can easily be implemented to nowcast GDP, perhaps even sequentially as time within a year proceeds.

Appendix: Proofs

Proof of Proposition 1. Recall the definition of $\mathbb{Y}_T = (Y_{1,T}, \ldots, Y_{m,T})' = (y_{mT-(m-1)}, \ldots, y_{mT})'$, and similarly define $\mathbb{X}_T$ and $\mathbb{E}_T$ from x_t and ε_t, respectively. The high-frequency DGP (1) can be expressed in structural VAR form (with exogenous regressors):

$$A_0\mathbb{Y}_T = A_1\mathbb{Y}_{T-1} + B_0\mathbb{X}_T + B_1\mathbb{X}_{T-1} + \mathbb{E}_T, \tag{A.1}$$

where

$$A_0 = I_m - \alpha L, \qquad A_1 = \alpha e_1 e_m', \qquad B_0 = \beta_0 I_m + \beta_1 L, \qquad B_1 = \beta_1 e_1 e_m',$$

and where I_m is the $m \times m$ identity matrix, e_i is the ith unit vector of dimension m, and L an $m \times m$ matrix with $L_{ij} = 1$ for $i = j + 1$, and zero otherwise (the lag transformation matrix). The reduced form VAR model is

$$\mathbb{Y}_T = \Phi\mathbb{Y}_{T-1} + \Gamma_0\mathbb{X}_T + \Gamma_1\mathbb{X}_{T-1} + \Psi\mathbb{E}_T, \tag{A.2}$$

where (defining $L^0 = I_m$)

$$\Phi = A_0^{-1}A_1 = \left(\sum_{i=1}^{m} \alpha^i e_i\right) e_m', \qquad \Gamma_0 = A_0^{-1}B_0 = \beta_0 I_m + (\alpha\beta_0 + \beta_1)\sum_{i=1}^{m-1} \alpha^{i-1}L^i,$$

$$\Psi = A_0^{-1} = \sum_{i=1}^{m} \alpha^{i-1}L^{i-1}, \qquad \Gamma_1 = A_0^{-1}B_1 = \beta_1\left(\sum_{i=1}^{m} \alpha^{i-1}e_i\right) e_m'.$$

We wish to work out the dynamics for $Y_T := m^{-1}\sum_{i=1}^{m} Y_{i,T}$, conditional on the high-frequency $X_{i,T}$ processes. Let $s_m = \sum_{i=1}^{m} e_i$, an m-vector of ones ("summing vector"), so that $Y_T = m^{-1}s_m'\mathbb{Y}_T$.

We find

$$mY_T = s_m'\Phi\mathbb{Y}_{T-1} + s_m'\Gamma_0\mathbb{X}_T + s_m'\Gamma_1\mathbb{X}_{T-1} + s_m'\Psi\mathbb{E}_T, \tag{A.3}$$

with

$$s_m'\Phi = \left(\sum_{i=1}^{m}\alpha^i\right)e_m', \qquad s_m'\Gamma_0 = \beta_0 s_m' + (\alpha\beta_0+\beta_1)\sum_{i=1}^{m-1}\left(\sum_{j=1}^{m-i}\alpha^{j-1}\right)e_i',$$

$$s_m'\Gamma_1 = \beta_1\left(\sum_{i=1}^{m}\alpha^{i-1}\right)e_m', \qquad s_m'\Psi = \sum_{i=1}^{m}\left(\sum_{j=0}^{m-i}\alpha^j\right)e_i'.$$

The lagged dependent term in (A.3) depends on $e_m'\mathbb{Y}_{T-1} = Y_{m,T-1}$, whereas we are looking for a representation involving only the low-frequency Y_{T-1}. To address this, we make use of

$$\left(\sum_{i=1}^{m}\alpha^i\right)Y_{m,T-1} - \alpha^m mY_{T-1} = \sum_{i=1}^{m-1}\alpha^i(Y_{m,T-1} - \alpha^{m-i}Y_{i,T-1}).$$

From (A.2), we have that

$$Y_{m,T-1} - \alpha^{m-i}Y_{i,T-1} = \xi_i'\mathbb{X}_{T-1} + \zeta_i'\mathbb{E}_{T-1},$$

where

$$\begin{aligned}\xi_i &= \alpha^{m-i-1}\beta_1 e_i' + (\alpha\beta_0+\beta_1)\sum_{j=i+1}^{m-1}\alpha^{m-j-1}e_j' + \beta_0 e_m' \\ &= \beta_0\sum_{j=i+1}^{m}\alpha^{m-j}e_j' + \beta_1\sum_{j=i}^{m-1}\alpha^{m-j-1}e_j',\end{aligned}$$

and

$$\zeta_i = \sum_{j=i+1}^{m}\alpha^{m-j}e_j'.$$

This leads to

$$mY_T = \alpha^m mY_{T-1} + s_m'\Gamma_0\mathbb{X}_T + \left(s_m'\Gamma_1 + \sum_{i=1}^{m-1}\alpha^i\xi_i'\right)\mathbb{X}_{T-1} + s_m'\Psi\mathbb{E}_T + \sum_{i=1}^{m-1}\alpha^i\zeta_i'\mathbb{E}_{T-1}.$$

After rearrangement of terms, this can be written as

$$\begin{aligned}\Delta Y_T &= (\alpha^m-1)Y_{T-1} + m^{-1}s_m'\Gamma_0\Delta\mathbb{X}_T + m^{-1}\left(s_m'(\Gamma_0+\Gamma_1) + \sum_{i=1}^{m-1}\alpha^i\xi_i'\right)\mathbb{X}_{T-1} \\ &\quad + m^{-1}s_m'\Psi\mathbb{E}_T + m^{-1}\sum_{i=1}^{m-1}\alpha^i\zeta_i'\mathbb{E}_{T-1},\end{aligned}$$

with

$$
\begin{aligned}
m^{-1}\left(s_m'(\Gamma_0+\Gamma_1)+\sum_{i=1}^{m-1}\alpha^i\xi_i'\right)\mathbb{X}_{T-1} &= \beta_0 X_{T-1}+m^{-1}(\alpha\beta_0+\beta_1)\sum_{i=1}^{m-1}\left(\sum_{j=1}^{m-i}\alpha^{j-1}\right)X_{i,T-1} \\
&\quad +m^{-1}\beta_1\left(\sum_{j=1}^{m}\alpha^{j-1}\right)X_{m,T-1} \\
&\quad +m^{-1}\beta_0\sum_{i=1}^{m-1}\alpha^i\sum_{j=i+1}^{m}\alpha^{m-j}X_{j,T-1} \\
&\quad +m^{-1}\beta_1\sum_{i=1}^{m-1}\alpha^i\sum_{j=i}^{m-1}\alpha^{m-j-1}X_{j,T-1} \\
&= \left(\sum_{i=0}^{m-1}\alpha^i\right)(\beta_0+\beta_1)X_{T-1} \\
&= \delta X_{T-1},
\end{aligned}
$$

and with

$$
\begin{aligned}
m^{-1}s_m'\Gamma_0\Delta\mathbb{X}_T &= \beta_0\Delta X_t+(\alpha\beta_0+\beta_1)\frac{1}{m}\sum_{i=1}^{m-1}\left(\sum_{j=1}^{m-i}\alpha^{j-1}\right)\Delta X_{i,T} \\
&= \sum_{j=1}^{m}\beta_{0j}\Delta X_{j,T}.
\end{aligned}
$$

This leads to the required result (6), with $U_T = m^{-1}s_m'\Psi\mathbb{E}_T + m^{-1}\sum_{i=1}^{m-1}\alpha^i\zeta_i'\mathbb{E}_{T-1}$. Note that $\delta = 0$ if $\alpha = 1$ and $\beta_0+\beta_1 = 0$; on the other hand, if $|\alpha| < 1$, then

$$
\delta = \frac{1-\alpha^m}{1-\alpha}(\beta_0+\beta_1) = (1-\alpha^m)\gamma,
$$

recalling the long-run parameter $\gamma = (\beta_0+\beta_1)/(1-\alpha)$.

To obtain the MA(1) structure of U_T, we use

$$
\sum_{i=1}^{m-1}\alpha^i\zeta_i' = \sum_{i=1}^{m-1}\alpha^i\sum_{j=i+1}^{m}\alpha^{m-j}e_j' = \sum_{i=2}^{m}\left(\sum_{j=m-i+1}^{m-1}\alpha^j\right)e_i',
$$

to find

$$
\begin{aligned}
\mathrm{var}(U_T) &= \frac{1}{m^2}\sum_{i=1}^{m}\left(\sum_{j=0}^{m-i}\alpha^j\right)^2+\frac{1}{m^2}\sum_{i=2}^{m}\left(\sum_{j=m-i+1}^{m-1}\alpha^j\right)^2, \\
\mathrm{cov}(U_T,U_{T-1}) &= \frac{1}{m^2}\sum_{i=2}^{m}\left(\sum_{j=0}^{m-i}\alpha^j\right)\left(\sum_{j=m-i+1}^{m-1}\alpha^j\right),
\end{aligned}
$$

and hence

$$
\rho = \mathrm{corr}(U_T,U_{T-1}) = \frac{\sum_{i=2}^{m}\left(\sum_{j=0}^{m-i}\alpha^j\right)\left(\sum_{j=m-i+1}^{m-1}\alpha^j\right)}{\sum_{i=1}^{m}\left(\sum_{j=0}^{m-i}\alpha^j\right)^2+\sum_{i=2}^{m}\left(\sum_{j=m-i+1}^{m-1}\alpha^j\right)^2}.
$$

The simplified expressions for ρ and β_{0i} in case $\alpha = 1, \beta_0 + \beta_1 = 0$ follow directly from the above expressions. For the case $|\alpha| < 1$, an alternative expression for ρ is

$$\rho = \frac{\alpha - m\alpha^m + m\alpha^{m+2} - \alpha^{2m+1}}{m - 2\alpha - m\alpha^2 + m\alpha^{2m} + 2\alpha^{2m+1} - m\alpha^{2m+2}}.$$

□

Proof of Proposition 2. The starting point is the assumed invariance principle:

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor un \rfloor} \begin{pmatrix} \varepsilon_t/\sigma_\varepsilon \\ \eta_t/\sigma_\eta \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W_1(u) \\ W_2(u) \end{pmatrix} = W(u), \qquad u \in [0,1].$$

Recall that $Y_T = m^{-1} \sum_{i=0}^{m-1} y_{mT-i}$, and $X_T = m^{-1} \sum_{i=0}^{m-1} x_{mT-i}$, $T = 1, \ldots, N = n/m$. Given that

$$X_T = x_0 + \frac{1}{m} \sum_{i=0}^{m-1} \sum_{t=1}^{mT-j} \eta_t = x_0 + \sum_{t=1}^{mT} \eta_t - \sum_{i=0}^{m-2} \left(\frac{m-1-i}{m} \right) \eta_{mT-i},$$

it follows that (for fixed m and as $n \to \infty$, hence $N \to \infty$)

$$\frac{1}{\sqrt{n}\sigma_\eta} X_{\lfloor uN \rfloor} = \frac{1}{\sqrt{n}\sigma_\eta} \sum_{t=1}^{m\lfloor uN \rfloor} \eta_t + o_p(1) \xrightarrow{d} W_2(u), \tag{A.4}$$

and similarly, under H_0,

$$\frac{1}{\sqrt{n}\sigma_\varepsilon} (Y_{\lfloor uN \rfloor} - \beta_0 X_{\lfloor uN \rfloor}) = \frac{1}{\sqrt{n}\sigma_\varepsilon} \sum_{t=1}^{m\lfloor uN \rfloor} \varepsilon_t + o_p(1) \xrightarrow{d} W_1(u). \tag{A.5}$$

Consider the reparametrization $(\alpha, \beta_0, \beta_1) \mapsto \psi = (\alpha, \delta, \beta_0)'$, where $\delta = (\beta_0 + \beta_1) \sum_{i=0}^{m-1} \alpha^i$, such that H_0 can be formulated as $H_0 : \alpha = 1, \delta = 0$. The LR statistic in the new parametrization is

$$\text{LR} = N \log \frac{\min_{\beta_0} S(1, 0, \beta_0)}{\min_{\alpha,\delta,\beta_0} S(\alpha, \delta, \beta_0)},$$

where

$$S(\alpha, \delta, \beta_0) = \frac{1}{N} \sum_{T=1}^{N} \left(\frac{\Delta Y_T - (\alpha^m - 1) Y_{T-1} - \delta X_{T-1} - \sum_{i=1}^{m} \beta_{0i}(\alpha, \delta, \beta_0) \Delta X_{i,T}}{1 + \theta(\alpha) L} \right)^2, \tag{A.6}$$

with with $\theta(\alpha)$ as implied by (8)–(9), and

$$\beta_{0i}(\alpha, \delta, \beta_0) = \frac{1}{m} \left(\beta_0 \alpha^{m-i} + \delta \frac{\sum_{j=0}^{m-i-1} \alpha^j}{\sum_{j=0}^{m-1} \alpha^j} \right).$$

To analyze the limiting null distribution, we require consistency of both the restricted estimator $\widetilde{\psi} = (1, 0, \widetilde{\beta}_0)'$ and the unrestricted estimator $\widehat{\psi} = (\widehat{\alpha}, \widehat{\delta}, \widehat{\beta}_0)$, under H_0. Because $\beta_{0i}(1, 0, \beta_0) = \beta_0/m$, such that $\sum_{i=1}^{m} \beta_{0i}(1, 0, \beta_0) \Delta X_{j,T} = \beta_0 \Delta X_T$, it follows directly from (A.6) that $\widetilde{\beta}_0$ solves

$$\frac{\partial S(1, 0, \widetilde{\beta}_0)}{\partial \beta_0} = -\frac{2}{N} \sum_{T=1}^{N} \frac{\Delta Y_T - \widetilde{\beta}_0 \Delta X_T}{1 + \theta(1) L} \times \frac{\Delta X_T}{1 + \theta(1) L} = 0,$$

where

$$\begin{aligned}\theta(1) &= \left.\frac{1-\sqrt{1-4\rho^2}}{2\rho}\right|_{\alpha=1} \\ &= \frac{1-\sqrt{1-4(m^2-1)^2/(4m^2+2)^2}}{2(m^2-1)/(4m^2+2)} \\ &= \frac{2m^2+1-m\sqrt{3m^2+6}}{m^2-1}.\end{aligned}$$

Therefore, $\widetilde{\beta}_0$ is the OLS estimator in the regression $(1+\theta(1)L)^{-1}\Delta Y_t = \beta_0(1+\theta(1)L)^{-1}\Delta X_T + \xi_T$, and given that ξ_T is uncorrelated with the $I(0)$ regressor, $\widetilde{\beta}_0$ is $\sqrt{N}$-consistent. For the unrestricted estimator, we analyze the gradients

$$\begin{aligned}\frac{\partial S(\psi)}{\partial \alpha} &= \frac{2}{N}\sum_{T=1}^{N}\xi_T(\psi)\times(1+\theta(\alpha)L)^{-1}\left(-m\alpha^{m-1}Y_{T-1} - \sum_{i=1}^{m}\frac{\partial\beta_{0i}(\alpha,\delta,\beta_0)}{\partial\alpha}\Delta X_{i,T}\right) \\ &\quad + \frac{2}{N}\sum_{T=1}^{N}\xi_T(\psi)\left(\frac{\partial(1+\theta(\alpha)L)^{-1}}{\partial\alpha}U_T(\psi)\right), \qquad \text{(A.7)}\\ \frac{\partial S(\psi)}{\partial \delta} &= \frac{2}{N}\sum_{T=1}^{N}\xi_T(\psi)\times(1+\theta(\alpha)L)^{-1}\left(-X_{T-1} - \sum_{i=1}^{m}\frac{\partial\beta_{0i}(\alpha,\delta,\beta_0)}{\partial\delta}\Delta X_{i,T}\right), \qquad \text{(A.8)}\\ \frac{\partial S(\psi)}{\partial \beta_0} &= \frac{2}{N}\sum_{T=1}^{N}\xi_T(\psi)\times(1+\theta(\alpha)L)^{-1}\left(-\sum_{i=1}^{m}\frac{\partial\beta_{0i}}{\partial\beta_0}\Delta X_{i,T}\right), \qquad \text{(A.9)}\end{aligned}$$

where $\xi_T(\psi)$ and $U_T(\psi)$ correspond to the functions (11)–(12) under the reparametrization $(\alpha,\beta_0,\beta_0)\mapsto\psi$. Using the fact that $(1+\theta L)^{-1} = 1+\sum_{i=1}^{\infty}(-\theta)^i L^i$, it follows that $\partial(1+\theta(\alpha)L)^{-1}/\partial\alpha$ is a power series in the lag operator with a zero weight at lag zero. This implies that the term in the second line of (A.7) has mean zero, when evaluated at the true value $\psi^0 = (1,0,\beta_0^0)'$ under H_0. Similarly, because $\xi_T(\psi^0)$ has zero covariance with $(1+\theta(1)L)^{-1}Y_{T-1}$, $(1+\theta(1)L)^{-1}X_{T-1}$ and $(1+\theta(1)L)^{-1}\Delta X_{i,T}$, it follows that the gradients in (A.7)–(A.9), when evaluated in ψ^0, have mean zero. Using the approach to derive the asymptotics in non-linear cointegration models developed by Saikkonen (1995), this property may be used to prove consistency of the unrestricted estimator $\widehat{\psi}$ (under a compactness assumption on the parameter space).

As $\widetilde{\psi}$ and $\widehat{\psi}$ are both consistent under H_0, the usual equivalence of the LR, Wald and LM statistics under H_0, following from a quadratic approximation of the log-likelihood, will apply. In the presence of different convergence rates for different parts of the parameter vector, as is the case in cointegration models, this still applies, in particular because the null hypothesis only restricts the parameters α and δ, which relate to coefficients of (non-cointegrated) $I(1)$ regressors. In other words, the restrictions do not combine parameters with different convergence rates. This leads to the following asymptotic expression that is valid under H_0:

$$\text{LR} = \frac{N}{2\sigma_\xi^2}\frac{\partial S(\widetilde{\psi})}{\partial\psi'}\left(\frac{\partial^2 S(\psi^0)}{\partial\psi\partial\psi'}\right)^{-1}\frac{\partial S(\widetilde{\psi})}{\partial\psi} + o_p(1), \qquad \text{(A.10)}$$

where the scale factor $N/(2\sigma_\xi^2)$ is due to the fact that $-\frac{1}{2}N\log S(\psi)$ is the concentrated log-likelihood (up to an additive constant), after concentrating out σ_ξ^2, implying that the score vector is proportional to

$-\frac{1}{2}NS(\psi)^{-1}\partial S(\psi)/\partial\psi$, and evaluation of $S(\psi)$ in either $\widetilde{\psi}$ or ψ^0 gives a consistent estimator of σ_ξ^2 under H_0.

The two non-zero elements of $\partial S(\widetilde{\psi})/\partial\psi$ satisfy

$$\begin{aligned}\frac{\partial S(\widetilde{\psi})}{\partial\alpha} &= -\frac{2m}{N}\sum_{T=1}^{N}\frac{Y_{T-1}}{1+\theta(1)L}\times\frac{\Delta Y_T-\beta_0\Delta X_T}{1+\theta(1)L}+o_p(1),\\ \frac{\partial S(\widetilde{\psi})}{\partial\delta} &= -\frac{2}{N}\sum_{T=1}^{N}\frac{X_{T-1}}{1+\theta(1)L}\times\frac{\Delta Y_T-\beta_0\Delta X_T}{1+\theta(1)L}+o_p(1),\end{aligned}$$

where we have used the fact that the terms involving the $I(1)$ regressors Y_{T-1} and X_{T-1} dominate the remainder terms involving $I(0)$ regressors; and that

$$\xi_T(\widetilde{\psi})=\frac{\Delta Y_T-\beta_0\Delta X_T}{1+\theta(1)L}+(\widetilde{\beta}_0-\beta_0)\frac{\Delta X_T}{1+\theta(1)L},$$

where the second term will be negligible in the limit expressions due to $\sqrt{N}$-consistency of $\widetilde{\beta}_0$.

Define the power series $c(L)=(1+\theta(1)L)^{-1}$. From the asymptotic properties of statistics based on linear processes, see Phillips and Solo (1992), we find that

$$\begin{aligned}\frac{1}{\sqrt{n}}c(L)X_{\lfloor uN\rfloor} &= \frac{1}{\sqrt{n}}\sum_{T=1}^{\lfloor uN\rfloor}c(L)\Delta X_T+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{T=1}^{\lfloor uN\rfloor}c(1)\Delta X_T+o_p(1)\\ &= c(1)\frac{1}{\sqrt{n}}X_{\lfloor uN\rfloor}+o_p(1)\xrightarrow{d}\sigma_\eta c(1)W_2(u),\end{aligned}$$

see (A.4), and similarly

$$\frac{1}{\sqrt{n}}\sum_{T=1}^{\lfloor uN\rfloor}c(L)(\Delta Y_T-\beta_0\Delta X_T)=c(1)\frac{1}{\sqrt{n}}\sum_{T=1}^{\lfloor uN\rfloor}(\Delta Y_T-\beta_0\Delta X_T)+o_p(1)\xrightarrow{d}\sigma_\varepsilon c(1)W_1(u),$$

see (A.5). Via the continuous mapping theorem, and the known results on weak convergence to stochastic integrals, this leads to

$$\begin{aligned}\frac{\partial S(\widetilde{\psi})}{\partial\alpha} &\xrightarrow{d} -2\sigma_\varepsilon^2c(1)^2m^2\int_0^1\left(W_1(u)+\beta_0\frac{\sigma_\eta}{\sigma_\varepsilon}W_2(u)\right)dW_1(u),\\ \frac{\partial S(\widetilde{\psi})}{\partial\delta} &\xrightarrow{d} -2\sigma_\varepsilon\sigma_\eta c(1)^2m\int_0^1 W_2(u)dW_1(u).\end{aligned}$$

It can be checked from the MA(1) structure of U_T that $mc(1)^2\sigma_\varepsilon^2=\sigma_\xi^2$, which, together with the notation $b=\sigma_\eta/\sigma_\varepsilon$, may be used to simplify the right-hand side expressions:

$$\begin{aligned}\frac{\partial S(\widetilde{\psi})}{\partial\alpha} &\xrightarrow{d} -2m\sigma_\xi^2\int_0^1\left(W_1(u)+\beta_0bW_2(u)\right)dW_1(u),\\ \frac{\partial S(\widetilde{\psi})}{\partial\delta} &\xrightarrow{d} -2b\sigma_\xi^2\int_0^1 W_2(u)dW_1(u).\end{aligned}$$

Similarly, we find for the elements of the Hessian matrix $\partial^2 S(\psi_0)/\partial\psi\partial\psi'$:

$$\begin{aligned}
\frac{1}{N}\frac{\partial^2 S(\psi^0)}{\partial\alpha^2} &\xrightarrow{d} 2m^2\sigma_\xi^2 \int_0^1 \left(W_1(u)+\beta_0 b W_2(u)\right)^2 du,\\
\frac{1}{N}\frac{\partial^2 S(\psi^0)}{\partial\delta^2} &\xrightarrow{d} 2b^2\sigma_\xi^2 \int_0^1 W_2(u)^2 du,\\
\frac{1}{N}\frac{\partial^2 S(\psi^0)}{\partial\alpha\partial\delta} &\xrightarrow{d} 2bm\sigma_\xi^2 \int_0^1 \left(W_1(u)+\beta_0 b W_2(u)\right) W_2(u) du.
\end{aligned}$$

Furthermore, the Hessian matrix will be asymptotically block-diagonal with respect to (α, δ) and β_0. Combining these results with (A.10), this leads, still under H_0, to the required result:

$$\begin{aligned}
\text{LR} &\xrightarrow{d} \int_0^1 dW_1(u) \begin{pmatrix} m(W_1(u)+\beta_0 b W_2(u)) \\ bW_2(u) \end{pmatrix}' \\
&\quad \times \left[\int_0^1 \begin{pmatrix} m(W_1(u)+\beta_0 b W_2(u)) \\ bW_2(u) \end{pmatrix}\begin{pmatrix} m(W_1(u)+\beta_0 b W_2(u)) \\ bW_2(u) \end{pmatrix}' du\right]^{-1} \\
&\quad \times \int_0^1 \begin{pmatrix} m(W_1(u)+\beta_0 b W_2(u)) \\ bW_2(u) \end{pmatrix} dW_1(u) \\
&= \int_0^1 dW_1(u)W(u)' \left(\int_0^1 W(u)W(u)'du\right)^{-1} \int_0^1 W(u)dW_1(u).
\end{aligned}$$

Under $H_1 : |\alpha| < 1$, we use the fact that

$$N^{-1}\text{LR} = \log\frac{S(\widetilde{\psi})}{S(\widehat{\psi})} = \log\frac{\widetilde{\sigma}_\xi^2}{\widehat{\sigma}_\xi^2}.$$

From consistency of $\widehat{\psi}$ under H_1 (which again can be proved using Saikkonen, 1995, given that the unrestricted model is correctly specified), it follows that

$$\widehat{\sigma}_\xi^2 = \frac{1}{N}\sum_{t=1}^{T}\widehat{\xi}_T^2 = \frac{1}{N}\sum_{t=1}^{T}\xi_T^2 + o_p(1) \xrightarrow{p} \sigma_\xi^2.$$

Since the restricted estimator is based on a misspecified model, we have

$$\widetilde{\sigma}_\xi^2 = \frac{1}{N}\sum_{t=1}^{T}\left(\frac{\Delta Y_T - \beta_0^*\Delta X_T}{1+\theta(1)L}\right)^2 + o_p(1) \xrightarrow{p} \text{var}\left(\frac{\Delta Y_T - \beta_0^*\Delta X_T}{1+\theta(1)L}\right) > \sigma_\xi^2,$$

where β_0^* is the pseudo-true value in the misspecified model. Therefore, with $c = \text{plim}\,\widetilde{\sigma}_\xi^2/\sigma_\xi^2 > 1,$ we have $N^{-1}\text{LR} \xrightarrow{p} \log c > 0$, and hence $\text{LR} = O_p(1)$.

□

References

Boswijk, H. P. (1994), "Testing for an unstable root in conditional and structural error correction models," *Journal of Econometrics*, 63, 37–60.

Boswijk, H. P. and P. H. Franses (1992), "Dynamic specification and cointegration," *Oxford Bulletin of Economics and Statistics*, 54, 369–381.

Ghysels, E. (2016), "Macroeconomics and the reality of mixed frequency data," *Journal of Econometrics*, 193, 294–314.

Ghysels, E., and J. I. Miller, (2015), "Testing for cointegration with temporally aggregated and mixed-frequency time series," *Journal of Time Series Analysis*, 36, 797–816.

Ghysels, E., A. Sinko and R. Valkanov (2007), "MIDAS regressions: Further results and new directions," *Econometric Reviews*, 26, 53–90.

Götz, T. B., A. Hecq and J.-P. Urbain (2014), "Forecasting mixed-frequency time series with ECM-MIDAS models," *Journal of Forecasting*, 33, 198–213.

MacKinnon, J. G., A. A. Haug and L. Michelis (1999), "Numerical distribution functions of likelihood ratio tests for cointegration," *Journal of Applied Econometrics*, 14, 563–577.

Phillips, P. C. B. and V. Solo (1992), "Asymptotics for linear processes," *The Annals of Statistics*, 20, 971–1001.

Saikkonen, P. (1995), "Problems with the asymptotic theory of maximum likelihood estimation in integrated and cointegrated systems," *Econometric Theory*, 11, 888–911.