Department of Finance
Faculty of Business and Economics

Working Paper Series

The underdamped oscillation MIDAS model

Jonathan Dark, David Harris

Working Paper No. 15/21

The underdamped oscillation MIDAS model*

Jonathan Dark[1*] David Harris[2]

[1]Department of Finance, University of Melbourne, Victoria 3010, Australia

[2]Department of Economics, University of Melbourne, Victoria 3010, Australia

*Corresponding author: jdark@unimelb.edu.au,

Tel: +61 3 8344 6866, Fax: +61 3 8344 6914

February 12, 2024

Abstract

We propose a flexible parametric MIDAS model that allows for overshooting by drawing on the literature for oscillating systems. Unlike exponential Almon or beta lags, weights can change sign. The model is parsimonious, easy to estimate, and includes a decay parameter that avoids endpoint restrictions. We establish consistency of model selection for oscillation based on suitable information criteria, and simulations support the approach in finite samples. The model is illustrated via out of sample forecasts of monthly changes in US inflation with a daily commodity price index, and the growth rate in US industrial production with the CBOE volatility index.

Keywords: MIDAS, underdamped oscillation, forecasting, overshooting

JEL classification: C22, C53

*We are particularly grateful for suggestions made by Peter Phillips and appreciate discussions held with Christopher Skeels, Morten Nielsen and participants at the Time Series and Forecasting Symposium 2019 and the Australia New Zealand Econometrics Study Group Meeting 2020. The authors have no declarations of interest. The data that support the findings of the paper are available on request.

1 Introduction

Since the seminal work of Ghysels et al. (2004), MIDAS models have become a popular method to analyse mixed frequency data.[1] Most applications employ exponential Almon or beta lags which are flexible, parsimonious and easy to estimate without the need for lag length optimisation. While they capture a variety of lag structures, the same sign is imposed on all lag weights which may be an issue if they oscillate or change sign.

It is well documented that economic activity is subject to short, medium and long term cycles (Ercolani, 2014). Continuous time macro-econometric systems (Bergstrom, 1966; Bergstrom et al., 1992) demonstrate that oscillating variables do not necessarily result in oscillating lag polynomials, as this requires second or higher order differential equations and complex roots. Oscillating lag polynomials are also consistent with the Dornbusch over-shooting model of exchange rates (Dornbusch, 1976) and its extension to commodity prices (Frankel, 1986).

The need for this flexibility is well understood when estimating distributed lag models (DLMs). The polynomial distributed lag (PDL) of Almon (1965) allows for sign changing weights, however it requires *choices* regarding lag length and polynomial order. Hendry et al. (1984) notes "the effects of an incorrect choice of polynomial and lag order are complex" and if either choice is incorrect estimates are biased and inconsistent.[2] Even if the lag length is known, a judgement regarding the trade off from a loss of efficiency (from too high a polynomial order) with biasedness and inconsistency (from too low a polynomial order) may be required (Godfrey and Poskitt, 1975). Further, PDL estimates do not guarantee damping of the lag weights and end-point restrictions often cause substantial loss (Trivedi, 1970).[3]

DLMs may have other functional forms that allow for sign changing weights, like the flexible fourier series (Hamlen and Hamlen Jr, 1978), splines (Corradi and Gambetta, 1976) or smoothness priors (Shiller, 1973). Fourier series require a choice regarding the number of harmonic terms. Similar issues exist with splines, which require selection of the number of knots, and the smoothness priors (SP) approach which requires a regularisation parameter.

[1] See Ghysels (2013) and Andreou et al. (2011) for reviews.

[2] Amemiya and Morimune (1974) show that the optimal lag length and polynomial order depend on the values of the data generating process (DGP), the number of lags, sample size, ratio of the error variance to the dependent variable variance, and degree of autocorrelation in the dependent variable.

[3] Whilst statistical tests may be used, the lag and polynomial order are unknown and there are different end point restrictions, so many tests are required.

Estimates from all approaches may be sensitive to changes in lag length, so lag length optimisation is required. There is also no guarantee that weights will damp.

MIDAS models have employed some of these flexible forms, however they suffer from similar issues. Breitung and Roling (2015) employ the SP approach to estimate a non-parametric (NP) MIDAS model. Even though estimation employs least squares, it is computationally intensive and requires judgement. Estimation is required over a range of lag lengths, with the smoothness parameter for each lag length determined via numerical optimisation of a modified AIC. Limitations of the modified AIC are acknowledged by the authors noting that "experimentation" is required. The PDL has also been applied to MIDAS (Ghysels, 2013; Pettenuzzo et al., 2016), where choice of lag length, polynomial order and the possibility of endpoint restrictions apply.

We propose a parametric MIDAS model that addresses these limitations, adopting a form used for underdamped oscillation in the physics literature. The flexible form can capture shapes associated with the exponential Almon and beta lags, but also allows the weights to change sign. Unlike previous polynomials that allow for sign changing weights (fourier series, spline, PDL or SP), little judgement, restriction testing or computational effort is required. The specification is parsimonious (with four parameters plus a constant), and easy to estimate. One of the parameters captures the rate of decay which helps avoid non-damping lag weights and endpoint restrictions. Like the exponential Almon and beta polynomials, the model determines the number of lags.

Asymptotic theory establishes consistency of the use of information criterion for model selection between the proposed underdamped oscillation MIDAS model (UO-MIDAS), a restricted version without oscillation, and no predictability. The approach is supported by monte carlo simulation, which also supports the finite sample properties of the non-linear least squares estimator and forecast performance relative to benchmarks. For illustrative purposes we consider two data sets. The first forecasts US inflation using a daily commodity price index, because commodity price overshooting sees the short term effect exceed the long run effect (Bergstrom et al., 1992; Frankel, 1986). The second forecasts US industrial production (IP) using the CBOE volatility index (VIX), given the overshooting of industrial production in response to a rise in uncertainty (Bloom, 2009). For both data sets the PDL, NP and UO MIDAS models capture overshooting, however UO MIDAS is much easier to

implement and provides superior out of sample (OOS) forecasts.

The paper is structured as follows. Section 2 provides theoretical support for oscillating lag polynomials. Section 3 briefly discusses the MIDAS literature and introduces the new model. Section 4 develops the asymptotic theory and section 5 assesses finite sample performance. Section 6 is the empirical application. Concluding remarks follow.

2 Theoretical support for oscillating lag polynomials

Bergstrom (1966) uses first order differential equations to characterise a macro-econometric system where variables adjust in response to deviations from a partial equilibrium level. Dependent variables respond to stimulus from other variables with exponentially decaying lags, with subsequent work permitting richer forms of lag distribution via *second order* differential equations (Bergstrom et al., 1992). Both models study the system near the equilibrium vector x^* by linearising it around that equilibrium via $A = \frac{\partial f(x)}{\partial x}|_{x=x^*}$, where A is the Jacobian matrix, $Dy = Ay$ where y is the vector of variables and D denotes the change. When the initial state is close to x^*, if the eigenvalues of A have negative real parts, the system is stable around x^* and all trajectories approach x^* as $t \to \infty$. If at least one eigenvalue has a positive real part the system is unstable, while complex eigenvalues result in an oscillating variable trajectory.[4]

Oscillating variables do not guarantee oscillating lag distributions as they require second (or higher) order differential equations and complex roots. Most equations in Bergstrom et al. (1992) have two speed of adjustment parameters (γ_1 and γ_2) that determine the lag distribution. Most equations are expressed as $D^2 y_i = -\gamma_1 D y_i + \gamma_2(x - y_i)$, where y_i is the i^{th} dependent variable and x is a function of other variables. If roots of the characteristic equation $z^2 + \gamma_1 z + \gamma_2 = 0$ are real, lag polynomials will not oscillate, but complex roots will result in oscillation.

To see why, consider the general solution to a second order linear homogeneous differential equation with characteristic equation $z^2 + \gamma_1 z + \gamma_2 = 0$ and complex roots $z = l + \alpha i$ and $z = l - \alpha i$. Derivation of the general solution is well understood (and is available in most

[4]Recent work often finds Hopf bifurcation boundaries with regular oscillation (Barnett et al., 2015).

undergraduate mathematics texts), with the solution equal to

$$y = e^{lt}[A\cos(\alpha t) + B\sin(\alpha t)] \tag{1}$$

where y denotes the value of the series at time t. The derivation of equation 1 employs Euler's equation $e^{iq} = cos(q) + isin(q)$ which establishes the link between complex roots and oscillation in the series. Given the standard solution characterises the dynamics of the series (y) over time, Bergstrom's application to a lag polynomial replaces y with the polynomial weight $w(j)$ for lag j

$$w(j) = e^{lj}[A\cos(\alpha j) + B\sin(\alpha j)] \tag{2}$$

In section 3.2 we will see this solution characterises the lag polynomial for our proposed model.

Of the 14 equations in the system, only inflation and exchange rates overshoot which is consistent with the Dornbusch model of exchange rates (Dornbusch, 1976) and its extension to commodities (Frankel, 1986). Like exchange rates, commodity prices are flexible so the commodity market over-reacts with short term effects greater than long run effects. These results motivate our proposed model below.

3 MIDAS Models

Section 3.1 provides a brief overview of MIDAS, outlining some of the alternative specifications. We then introduce the proposed model in Section 3.2.

3.1 Previous literature

The MIDAS model for a single explanatory variable and h step ahead forecasting is

$$Y_t = \beta_0 + \beta_1 W(L^{1/m};\theta)x_{t-h}^{(m)} + \varepsilon_t \tag{3}$$

where $W(L^{1/m};\theta) = \Sigma_{j=1}^{K} w(j;\theta)L^{(j-1)/m}$, K is the lag length and $L^{s/m}x_{t-1}^{(m)} = x_{t-1-s/m}^{(m)}$. Here t is the low frequency time unit (say monthly) and m is the higher sampling frequency (say daily), where $L^{1/m}$ operates at the higher frequency. This has been extended to the p

order autoregressive(AR)-MIDAS model

$$Y_t = \beta_0 + \sum_{i=1}^{p} \lambda_i Y_{t-h-i+1} + \beta_1 W(L^{1/m}; \theta) x_{t-h}^{(m)} + \varepsilon_t \tag{4}$$

as well as a common factor structure [5]

$$Y_t = \beta_0 + \sum_{i=1}^{p} \lambda_i Y_{t-h-i+1} + \beta_1 W(L^{1/m}; \theta)(1 - \sum_{i=1}^{p} \lambda_i L^i) x_{t-1}^{(m)} + \varepsilon_t. \tag{5}$$

The most common polynomial is the exponential Almon lag (herein the Almon lag)

$$w(j; \theta) = \frac{\exp(\theta_1 j + \theta_2 j^2)}{\sum_{j=1}^{K} \exp(\theta_1 j + \theta_2 j^2)}. \tag{6}$$

An alternative is the beta lag

$$w(j; \theta) = \frac{f(\frac{j}{K}, \theta_1; \theta_2)}{\sum_{j=1}^{K} f(\frac{j}{K}, \theta_1; \theta_2)} \tag{7}$$

where $f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ and $\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx$. Both polynomials impose the same sign on all lag coefficients which may be unsuitable if there is overshooting. More flexible approaches like the polynomial distributed lag (PDL) of Almon (1965) let the weights change sign

$$w(j; \theta) = \sum_{k=0}^{P} \theta_k j^k \tag{8}$$

however as mentioned above, there are significant costs if an incorrect polynomial order (P), lag length (K) or end-point restriction is imposed. Stepwise weights (Forsberg and Ghysels, 2007) require determination of the number of steps and may not be parsimonious. By estimating a coefficient for each lag, the unrestricted MIDAS model (Foroni et al., 2015) requires a low m. The non-parametric model (NP-MIDAS) (Breitung and Roling, 2015) overcomes this but requires a regularisation parameter. Further, all of these models require a choice of lag length and none guarantee polynomial decay.[6]

[5] Clements and Galvão (2008) are critical of the basic AR(p)-MIDAS specification because it artificially generates a periodic response of Y_t to $x_t^{(m)}$.

[6] Many other extensions to the model have occurred, including specifications that employ machine learning and textual data (Babii et al., 2022).

3.2 The proposed MIDAS model

We propose a parametric form that draws on the physics literature for oscillating systems. The model is parsimonious, easy to estimate, and includes a parameter that captures polynomial decay. This latter feature lets the data determine the number of lags, and helps avoid non-damping lag polynomials and endpoint restrictions. An autoregressive version of the underdamped oscillation MIDAS model (UO-AR(p)-MIDAS) is

$$Y_t = \beta_0 + \sum_{i=1}^{p} \lambda_i Y_{t-h-i+1} + \sum_{j=1}^{K} w(j;\theta) x^{(m)}_{t-1-(j-1)/m} + \varepsilon_t \tag{9}$$

where

$$w(j;\theta) = e^{-\frac{\gamma}{2}j}[A\cos(\alpha j) + B\sin(\alpha j)]. \tag{10}$$

The common factor version of the model (UO-comfac-MIDAS) applies the polynomial (Equation (10)) to the common factor structure (Equation (5)). The model without auto-regressive dynamics (UO-MIDAS) applies the same polynomial to Equation (3). UO MIDAS (no hyphen) denotes all versions of the model.

The UO polynomial has four parameters $\theta = (\gamma, A, B, \alpha)$, where γ controls the rate of geometric damping, A and B control the oscillation amplitude or signal strength, and α the oscillation frequency. Figure 1 plots $e^{-\frac{\gamma}{2}j}$ by lag for γ values between 0.025 to 0.4. For a lag length of one year γ may be ≈ 0.025 for daily data and ≈ 0.4 for monthly data.[7] Unlike the Almon and Beta polynomials which normalise lag weights and summarise the overall response via β_1, the UO polynomial is not normalised. This is because even if the net effect of $x^{(m)}_t$ is zero, the oscillating polynomial may still be useful for forecasting. Instead, a higher signal to noise ratio (SNR) is achieved by increasing A and B.[8] The model's flexibility therefore comes at the cost of not being able to summarise the overall response of Y_t to $x^{(m)}_{t-h}$.

(Insert Figure 1)

Table 1 reports parameter values for 8 DGPs, and Figure 2 plots their polynomials up to lag 250. DGP1 is based on empirical estimates below, and is consistent with overshooting.

[7] In physics the damping ratio is $\gamma = c/\sqrt{km}$, where c is a friction coefficient, m the mass of the object and k the elastic constant corresponding to Hooke's law. Equation 10 applies the polynomial associated with the standard solution i.e the real part of the complex root $l = -\gamma/2$ however $l = -\gamma*$ could also be used (where $\gamma* = \gamma/2$).

[8] Holding α and γ constant, an "x" fold increase in A and B re-scales the y axis by a factor of "x", without changing the polynomial shape.

If $\alpha = 0$, B is unidentified and there is no oscillation with geometric decay (DGP2). An increase(decrease) in α, increases(decreases) oscillation (DGP3-4). The relative weights on A and B move the location of the hump (DGP4) and an initial negative response occurs if $B < 0$ (DGP5-6). A decrease in γ slows the rate of decay (DGP7) and γ increases dampen the oscillation irrespective of α. Lags similar to empirical Almon and beta polynomial estimates can also be generated (DGP8).

Table 1: UO-MIDAS model parameters

	DGP1	DGP2	DGP3	DGP4	DGP5	DGP6	DGP7	DGP8
β_0	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
γ	0.04	0.10	0.04	0.06	0.10	0.04	0.005	0.15
A	0.01	0.01	0.01	0.001	0.001	0.01	0.01	0.01
B	0.01	0.00	0.01	0.01	-0.01	-0.01	0.01	0.10
α	0.09	0.00	0.15	0.05	0.09	0.09	0.05	0.05

(Insert Figure 2)

The UO polynomial is similar to the flexible fourier form (FFF)

$$w(j;\theta) = \sum_{i=1}^{Q} \left[\theta_{1,i} \sin\left(\frac{2ij\pi}{K}\right) + \theta_{2,i} \cos\left(\frac{2ij\pi}{K}\right) \right] \tag{11}$$

which has been employed as a trend function in VARs (Enders and Jones, 2016), volatility (Baillie and Morana, 2009), and unit root testing (Enders and Lee, 2012), as well as a lag polynomial in DLMs (Hamlen and Hamlen Jr, 1978). Q may be determined via information criteria (IC), however its order and shape can vary significantly depending on the IC used. When modeling a trend, K is the number of observations and there is no need for damping. When fitting a lag polynomial, weights are expected to damp and estimates can be very sensitive to the lag length. Estimation therefore requires joint determination of Q, K and the possibility of end-point restrictions.

4 Model selection

An identification problem may arise in the UO-MIDAS model for certain coefficient values. A leading example is that the coefficients $w(j;\theta)$ are monotonic in j if $\alpha = 0$, in which case B is not identified. A similar lack of identification of B occurs if $\alpha = \pi$, albeit $w(j;\theta)$ are not monotonic in this case. Another example is that $A = B = 0$ will result in $w(j;\theta) = 0$ for every j, implying X_t is not useful for forecasting Y_t and γ and α are not identified. In this

section we address these possible identification issues for UO-MIDAS using model selection based on an information criterion. We propose the Schwarz (SIC) or Hannan-Quinn criterion (HQIC) for model selection between UO-MIDAS, and versions with no oscillation ($\alpha = 0$) or no predictability ($w(j;\theta) = 0$ for all $j \in 1, 2, ...K$). This is shown to provide consistent selection of the correct model.[9]

To develop the approach, re-express the UO-MIDAS model as

$$Y_t = X_{t-1}' W(\theta) + \varepsilon_t$$

where

$$X_{t-1} = \begin{pmatrix} x_{t-1} \\ x_{t-1-1/m} \\ \vdots \\ x_{t-1-(K-1)/m} \end{pmatrix}, \quad W(\theta) = \begin{pmatrix} w(1;\theta) \\ w(2;\theta) \\ \vdots \\ w(K;\theta) \end{pmatrix}, \quad \theta = \begin{pmatrix} \gamma \\ A \\ B \\ \alpha \end{pmatrix}$$

Denote the parameter space for θ by $\Theta = \Gamma \times \mathcal{A} \times \mathcal{B} \times [-\pi, \pi]$. The true parameter values are denoted $\theta_0 = (\gamma_0, A_0, B_0, \alpha_0)'$. Additional regressors, such as intercept and lags of Y_t, are omitted from this representation but would be included in any estimation as required without changing what follows. The range for α is restricted to $[0, 2\pi - \varepsilon]$ for small $\varepsilon > 0$ for identification purposes — the non-negativity is because the functions cos and sin are even and odd respectively, and the upper limit of 2π is because cos and sin have periodicity of 2π. The relevant parameter configurations for the model selection step are as follows.

(a) The unrestricted parameter space is denoted $\Theta_a = \Gamma \times \mathcal{A} \times \mathcal{B} \times [0, 2\pi - \varepsilon]$ with free parameter count for the IC below given by $p_a = 5$ (including intercept).

(b) If $A = B = 0$ then $w(j;\gamma, 0, 0, \alpha) = 0$ for every j, implying that X_{t-1} has no predictive power for Y_t, and γ and α are not identified. The restricted parameter space for this

[9] Physicists avoid these issues because the oscillating variable is typically *observable*. On measuring the decay ratio ((height of the second peak)/(height of the first peak)) and period of oscillation (time between the first and second peaks) UO parameters are calculated analytically. In the UO-MIDAS model, the oscillation is a *latent relation* between variables, so these methods are not possible. An alternative more computationally intensive approach would employ a Davies procedure (Davies, 1987; Hansen, 1996). If η denotes the parameters under the null and ϕ the unidentified nuisance parameters, when ϕ is identified the test statistic is based on the expected score $\partial l(\eta, \phi)/\partial \eta$ only being equal to zero when η and ϕ are at their true values. When ϕ is unidentified, the expected value of the score is zero when η is at its true value and ϕ is an arbitrary constant (Davies, 1987). The score can be calculated over a range of values for ϕ and the maximum determined. Unlike the regular test, the null is not χ^2_1. It should be noted that in the Almon and beta MIDAS models β_1 and θ are not separately identified under the null: θ_1 and θ_2 are identified if $\beta_1 \neq 0$ and β_1 is identified if the lag weights sum to unity.

situation is denoted $\Theta_b = \{0\} \times \{0\} \times \{0\} \times \{0\}$ with free parameter count for the IC given by $p_b = 1$.

(c) If $\alpha = 0$ then $w(j; \gamma, A, B, 0) = e^{-\frac{\gamma}{2}j}A$, implying that B is not identified. In this case X_{t-1} has predictive power for Y_t with monotonic non-oscillating weights. The restricted parameter space is denoted $\Theta_c = \Gamma \times \mathcal{A} \times \{0\} \times \{0\}$ with free parameter count for the IC given by $p_c = 3$.

(d) If $\alpha = \pi$ then $w(j; \gamma, A, B, \pi) = e^{-\frac{\gamma}{2}j}A(-1)^j$, implying that B is not identified. In this case X_{t-1} has predictive power for Y_t with oscillating weights. The restricted parameter space is denoted $\Theta_d = \Gamma \times \mathcal{A} \times \{0\} \times \{\pi\}$ with free parameter count for the IC given by $p_d = 3$.

Hypothesis testing to select between cases (a)–(d) would be complicated by unidentified nuisance parameters in (b)–(d). Nevertheless an approach based on a standard information criteria can give consistent model selection between (a)–(d). The (quasi) Gaussian log-likelihood for the model can be expressed

$$\ell_n(\theta) = -\frac{n}{2}\log(1 + 2\pi) - \frac{n}{2}\log\hat{\sigma}_n^2(\theta)$$

where $\hat{\sigma}_n^2(\theta)$ is the usual variance

$$\hat{\sigma}_n^2(\theta) = \frac{1}{n}\sum_{t=1}^{n}(Y_t - X_{t-1}'W(\theta))^2.$$

The IC has the penalised log-likelihood form

$$\xi_m = -2\max_{\theta \in \Theta_m} \ell_n(\theta) + p_m q_n, \quad m = a, b, c, d.$$

This is the Schwarz criterion if $q_n = \log n$, the Akaike criterion if $q_n = 2$, and the HQ criterion if $q_n = 2\log\log n$. The selected model from a, b, c, d is the one with the smallest value of ξ_m. The following technical assumptions are made to prove Theorem 1, which gives the consistency of model selection between cases (a)–(d).

Assumptions

(A) Θ is compact.

(B) θ_0 lies in the interior of Θ.

(C) (X_t, ε_t) is strictly stationary and ergodic with finite fourth moment.

(D) ε_t is a stationary martingale difference sequence with $E(\varepsilon_t^2) = \sigma_0^2$ and finite fourth moment

(E) $E(X_t X_t') = \Sigma_{XX}$ exists and is non-singular.

The compactness of Θ can be imposed by placing bounds on the parameters, for example $\gamma \in \Gamma = [0, \overline{\gamma}]$, $A \in \mathcal{A} = [\underline{A}, \overline{A}]$, $\alpha \in [0, 2\pi - \varepsilon]$, $B \in \mathcal{B} = [\underline{B}, \overline{B}]$. For restricted estimation imposing $\alpha = 0$, define the restricted parameter space $\Theta_\alpha = \Gamma \times \mathcal{A} \times \mathcal{B} \times \{0\}$.

Theorem Under Assumptions (A)–(E), $q_n \to \infty$ and $n^{-1} q_n \to 0$

(a) If $(A_0, B_0) \neq (0, 0)$ and $\alpha_0 \neq 0$ or π then $P(\xi_a < \xi_b, \xi_c, \xi_d) \to 1$.

(b) If $A_0 = B_0 = 0$ then $P(\xi_b < \xi_a, \xi_c, \xi_d) \to 1$.

(c) If $(A_0, B_0) \neq (0, 0)$ and $\alpha_0 = 0$ then $P(\xi_c < \xi_a, \xi_b, \xi_d) \to 1$.

(d) If $(A_0, B_0) \neq (0, 0)$ and $\alpha_0 = \pi$ then $P(\xi_d < \xi_a, \xi_b, \xi_c) \to 1$.

The proof is provided in the Appendix. The implication is that the SIC ($c_n = \log n$) or HQIC ($c_n = 2 \log \log n$) can be used to consistently select between cases (a)–(d). The AIC ($c_n = 2$) is not covered by the theorem, as it does not provide consistent model selection.

5 Finite sample properties of the UO-MIDAS model

First we examine the non-linear least squares (NLLS) estimator when the model is correctly specified. Section 5.2 then examines use of the SIC and HQIC when selecting between UO-MIDAS, UO-MIDAS with $\alpha = 0$, and no predictability ($w(j; \theta) = 0$ for all $j \in \{1, .., K\}$). Section 5.3 considers polynomial fit and OOS forecasts relative to MIDAS benchmarks (Almon, NP, PDL). We employ a similar setup to Andreou et al. (2010) where the DGP is

$$Y_t = \beta_0 + W(L^{1/m}; \theta) x_t^{(m)} + \varepsilon_t \tag{12}$$

$\varepsilon_t \sim N.i.i.d$ (0,0.125), the sample size is $T = 100, 200, 500$, and $x_t^{(m)}$ is sampled m times between $t-1$ and t with a sample size of mT. Each DGP has $m = 3, 40, 100$; low, medium and high signal to noise ratios (SNRs); and three different specifications for $x_t^{(m)}$ i) $x_t^{(m)} \sim N.i.i.d$ (0,1); ii) $x_t^{(m)} = \sigma_t^{(m)} e_t^{(m)}$ where $(\sigma_t^{(m)})^2 = 0.05 + 0.1(x_{t-1}^{(m)})^2 + 0.85(\sigma_{t-1}^{(m)})^2$ and $e_t^{(m)} \sim N.i.i.d$ (0,1); and iii) $x_t^{(m)} = 0.5 + \varsigma_1 x_{t-1}^{(m)} + e_t^{(m)}$, where $e_t^{(m)} \sim N.i.i.d$ (0,1). For most DGPs $\varsigma_1 = 0.9$, but we also consider $x_t^{(m)} = 0.5 + \varsigma_1 x_{t-1}^{(m)} + \varsigma_2 x_{t-2}^{(m)} + e_t^{(m)}$, where $\varsigma_1 = 0.5$ and $\varsigma_2 = -0.3$. Each simulation performs 500 replications.[10]

Table 2 summarises the five UO-MIDAS DGPs considered. Low SNRs correspond to the first five DGPs in Table 1. Medium (high) SNRs increase A and B by a factor of 5(10).[11] True is the true lag and max, min and steps are discussed in section 5.3.

Table 2: Monte carlo: UO-MIDAS model parameters

	DGP1	DGP2	DGP3	DGP4	DGP5
Parameters					
β_0	0.50	0.50	0.50	0.50	0.50
γ	0.04	0.10	0.04	0.06	0.10
α	0.09	0.00	0.15	0.05	0.09
Low signal					
A	0.01	0.01	0.01	0.001	0.001
B	0.01	0.00	0.01	0.01	-0.01
Medium signal					
A	0.05	0.05	0.05	0.005	0.005
B	0.05	0.00	0.05	0.05	-0.05
High signal					
A	0.10	0.10	0.10	0.01	0.01
B	0.10	0.00	0.10	0.10	-0.10
Lags					
True	130	90	130	120	80
Max	180	140	180	170	130
Min	80	40	80	70	30
Steps	10	10	10	10	10

The graphs in Figure 2 represent the parameters for the weak SNR. True is the number of lags used for data generation. In section 5.3, max represents the lag for the UO-MIDAS and Almon-MIDAS models. The NP-MIDAS and PDL-MIDAS models estimate an optimal lag, by considering lags between max and min with increments equal to the step size.

Findings are largely insensitive to m and $x_t^{(m)}$ so we focus on $m = 40$ and $x_t^{(m)} \sim i.i.d$. The exception is when $x_t^{(m)} \sim AR(1)$ in a limited number of circumstances, which will be discussed below. Given the large number of experiments, illustrative examples are reported with more details available on request.[12]

[10] We don't consider UO-MIDAS models with more than one $x_t^{(m)}$, as most studies with > 1 predictor either regress Y_t against a principle component or employ forecast combinations e.g Andreou et al. (2013).

[11] A and B are set to approximate the strength of the low, medium and high SNRs in Andreou et al. (2010). The remaining DGPs in Table 1 are not considered as DGP 6 has a similar shape to DGPs 1, 3, 4 and 5, DGP 7 damps too slowly for economic applications, and DGP 8 is similar to the Almon lag.

[12] We also consider the effects of an incorrect lag. Like the exponential Almon and beta polynomials, parameters adjust so the polynomial and forecasts are insensitive to overstating the lag length.

5.1 Finite sample properties of NLLS estimator

UO-MIDAS is generally well estimated via NLLS. Table 3 presents results for DGP 1 which show that when $x_t^{(m)} \sim i.i.d$ and $x_t^{(m)} \sim AR(2)$, estimates have a small bias and appear $\sqrt{T}$ consistent. When $x_t^{(m)} \sim AR(1) : \varsigma = 0.9$ and the SNR low or medium, the small bias and $\sqrt{T}$ consistency remains. The decrease in RMSE (relative to $x_t^{(m)} \sim i.i.d$ and $x_t^{(m)} \sim AR(2)$), suggests persistence in $x_t^{(m)}$ helps with identification. This also applies when the SNR is high and $T = 100$, however when $T \geq 200$ increases in T *increase* bias and RMSE.

Table 3: Finite sample properties of UO-MIDAS model: DGP 1

DGP			Bias					RMSE				
SNR	m	T	β_0	A	B	α	γ	β_0	A	B	α	γ
$x_t^{(m)} \sim i.i.d$												
low	40	100	0.0005	−0.0021	0.0685	0.0024	0.0150	0.1132	0.0970	0.7893	0.1676	0.2873
		200	0.0007	−0.0004	0.0004	0.0003	0.0041	0.0949	0.0720	0.0978	0.1166	0.1846
		500	0.0002	0.0000	0.0001	−0.0002	0.0006	0.0758	0.0496	0.0540	0.0778	0.1020
med	40	100	0.0003	0.0000	−0.0005	−0.0002	−0.0001	0.1128	0.0750	0.0800	0.0511	0.0657
		200	0.0007	0.0001	−0.0003	−0.0001	0.0000	0.0949	0.0634	0.0664	0.0429	0.0542
		500	0.0002	0.0000	0.0001	0.0000	0.0001	0.0757	0.0489	0.0520	0.0336	0.0427
high	40	100	0.0003	0.0000	−0.0004	−0.0001	−0.0001	0.1128	0.0749	0.0794	0.0358	0.0462
		200	0.0007	0.0001	−0.0002	0.0000	0.0000	0.0949	0.0633	0.0663	0.0303	0.0383
		500	0.0002	0.0000	0.0001	0.0000	0.0000	0.0757	0.0489	0.0520	0.0238	0.0301
$x_t^{(m)} \sim AR(1)$: $\varsigma_1 = 0.9$												
low	40	100	0.0026	0.0000	0.0000	−0.0001	−0.0001	0.1991	0.0304	0.0289	0.0438	0.0518
		200	0.0020	0.0000	0.0000	0.0000	0.0000	0.1650	0.0259	0.0244	0.0369	0.0437
		500	−0.0007	0.0000	0.0000	0.0000	0.0000	0.1328	0.0199	0.0194	0.0293	0.0338
med	40	100	0.0024	0.0000	0.0000	0.0000	0.0000	0.1988	0.0305	0.0288	0.0196	0.0232
		200	0.0019	0.0000	0.0000	0.0000	0.0000	0.1648	0.0259	0.0243	0.0164	0.0195
		500	−0.0008	0.0000	0.0000	0.0000	0.0000	0.1328	0.0199	0.0193	0.0131	0.0151
high	40	100	0.0023	0.0000	0.0000	0.0000	0.0000	0.1988	0.0305	0.0288	0.0138	0.0164
		200	0.0334	−0.0004	−0.0003	0.0001	−0.0002	0.5977	0.0708	0.0626	0.0311	0.0467
		500	1.4297	−0.0007	−0.0205	0.0015	−0.0038	1.6232	0.1005	0.1956	0.0559	0.0858
$x_t^{(m)} \sim AR(2)$: $\varsigma_1 = 0.5$, $\varsigma_2 = -0.3$.												
low	40	100	0.0014	−0.0013	0.0108	0.0029	0.0102	0.1943	0.0851	0.4518	0.1472	0.2391
		200	0.0015	−0.0002	0.0003	0.0002	0.0025	0.1553	0.0607	0.0804	0.0981	0.1636
		500	−0.0004	0.0000	0.0001	−0.0001	0.0003	0.1233	0.0438	0.0473	0.0678	0.0884
med	40	100	0.0022	0.0001	−0.0004	−0.0001	−0.0001	0.1852	0.0663	0.0710	0.0452	0.0585
		200	0.0017	0.0001	−0.0002	−0.0001	0.0000	0.1526	0.0560	0.0591	0.0380	0.0483
		500	−0.0004	0.0000	0.0001	0.0000	0.0000	0.1224	0.0435	0.0464	0.0299	0.0380
high	40	100	0.0022	0.0001	−0.0004	−0.0001	0.0000	0.1850	0.0662	0.0707	0.0318	0.0412
		200	0.0016	0.0001	−0.0002	0.0000	0.0000	0.1526	0.0559	0.0590	0.0268	0.0342
		500	−0.0005	0.0000	0.0001	0.0000	0.0000	0.1225	0.0435	0.0464	0.0212	0.0268

Despite this, when $\varsigma = 0.9$, $m = 40$ and $T = 500$, the median estimate of all parameters for DGP1 is close to the true value (in brackets): $B_0 = 0.5006(0.5)$, $A = 0.1000(0.1)$, $B = 0.0998(0.1)$, $\gamma = 0.0900(0.9)$, $\alpha = 0.0399(0.04)$. Importantly the increase in bias and RMSE as T increases, was only observed in DGP 1 and 2 for the high SNR when $x_t^{(m)} \sim AR(1) : \varsigma = 0.9$. DGPs 3-5 have a small bias and appear $\sqrt{T}$ consistent for *all* experiments. Results suggest if the SNR is high and $x_t^{(m)}$ persistent, NLLS may have difficulty

identifying the intercept (β_0) from the polynomial (in particular B). Section 5.3 however will show this doesn't adversely affect polynomial fit and OOS forecasts.[13]

Table 4 reports the bias and RMSE for DGP2 where $\alpha = 0$. B has a large positive bias and RMSE, and α also has a positive though much smaller bias. For a given SNR and $x_t^{(m)}$, an increase in T decreases bias and RMSE for all parameters. When $x_t^{(m)} \sim i.i.d$, increasing the SNR decreases bias and RMSE and estimates appear $\sqrt{T}$ consistent. However when $\zeta_1 = 0.9$, an increase in the SNR *increases* bias and RMSE of B which is consistent with DGP1 (Table 3). The effect of a large positive bias in B is minimal though as $\alpha \to 0$, because $\sin(\alpha)$ and therefore $B\sin(\alpha) \to 0$. This is reflected in Table 5 that reports the number of replications where $-0.001 \leq \alpha \leq 0.001$. The significant pile up around zero means (like DGP1), there is no adverse affect on polynomial fit and OOS forecasts in section 5.3.

Table 4: Finite sample properties of UO-MIDAS model: DGP 2

DGP			Bias					RMSE				
SNR	m	T	β_0	A	B	α	γ	β_0	A	B	α	γ
$x_t^{(m)} \sim i.i.d$												
low	40	100	−0.0002	−0.0243	0.3310	0.0578	0.0471	0.1099	0.6712	1.5509	0.3282	0.5929
		200	−0.0001	−0.0025	0.2009	0.0434	0.0551	0.0939	0.1569	0.8928	0.3109	0.5636
		500	0.0004	−0.0017	0.1453	0.0191	0.0220	0.0742	0.0670	0.7343	0.1962	0.3046
med	40	100	−0.0002	−0.0025	0.3388	0.0074	0.0092	0.1098	0.0912	1.1027	0.1223	0.2077
		200	−0.0001	−0.0019	0.2155	0.0065	0.0068	0.0941	0.0737	0.9303	0.1109	0.1827
		500	0.0004	−0.0010	0.0960	0.0056	0.0034	0.0742	0.0566	0.6997	0.0991	0.1477
high	40	100	−0.0002	−0.0021	0.2331	0.0047	0.0033	0.1099	0.0881	1.1230	0.0943	0.1609
		200	−0.0001	−0.0016	0.1510	0.0045	0.0027	0.0940	0.0718	1.0028	0.0898	0.1392
		500	0.0003	−0.0009	0.0424	0.0044	0.0007	0.0742	0.0555	0.7229	0.0858	0.1108
$x_t^{(m)} \sim AR(1)$: $\varsigma_1 = 0.9$												
low	40	100	0.0036	−0.0005	0.0188	0.0080	0.0149	0.2441	0.0365	0.1498	0.1135	0.1627
		200	0.0053	−0.0004	0.0173	0.0063	0.0124	0.2085	0.0302	0.1414	0.1017	0.1462
		500	0.0038	−0.0002	0.0166	0.0052	0.0113	0.1642	0.0237	0.1296	0.0909	0.1336
med	40	100	0.0149	−0.2007	0.5950	0.0072	0.0261	0.4206	1.7792	1.6774	0.2727	0.5463
		200	0.0065	−0.0004	0.4243	0.0022	0.0059	0.2200	0.0374	1.6185	0.0604	0.1039
		500	0.0055	−0.0003	0.2466	0.0019	0.0055	0.1905	0.0412	1.2165	0.0554	0.1039
high	40	100	0.0175	−0.0721	0.6691	0.0032	0.0127	0.5145	1.2649	1.9925	0.1659	0.4178
		200	0.0122	−0.0006	0.3314	0.0016	0.0045	0.3633	0.0724	1.3918	0.0509	0.1067
		500	0.0175	−0.0007	0.1873	0.0018	0.0038	0.6532	0.0976	1.1338	0.0662	0.1164

5.2 HQIC and SIC for model selection

We now consider the HQIC and SIC for model selection between: Model 1 UO-MIDAS, Model 2 UO-MIDAS with $\alpha = 0$ and Model 3 $w(j;\theta) = 0$ for $j \in \{1, ..., K\}$. We employ

[13] High persistence in $x_t^{(m)}$ means highly correlated regressors, so our findings are consistent with forecasts not being compromised when RHS variables suffer from multi-collinearity. We also consider first differencing $x_t^{(m)}$ when estimating UO-MIDAS with $\varsigma = 0.9$. The experiment is otherwise identical to section 5.3, where Almon, NP and PDL MIDAS models do not difference $x_t^{(m)}$. Differencing significantly worsens UO-MIDAS polynomial fit and OOS MSE, with NP-MIDAS clearly dominant.

Table 5: No of replications with $\alpha \approx 0$ UO-MIDAS model: DGP 2

		Low SNR			Medium SNR			High SNR		
	m	T=100	T=200	T=500	T=100	T=200	T=500	T=100	T=200	T=500
iid	3	104	147	200	258	266	281	273	283	297
	40	139	176	223	275	281	281	304	307	284
	100	128	160	183	253	254	267	290	281	265
AR(1)	3	249	242	244	303	291	305	277	277	253
$\varsigma_1 = 0.9$	40	244	255	243	282	283	271	279	280	259
	100	245	235	233	295	279	266	273	262	260

Reports the number of replications (out of 500), where $-0.001 \leq \alpha \leq 0.001$.

the 5 UO-MIDAS DGPs and compare OOS forecasts from UO-MIDAS to approaches that use the HQIC or SIC to select model 1, 2 or 3. For each of the 500 replications, OOS MSE is calculated by fixing the estimated parameters and conditionally updating one step ahead forecasts over 100 OOS observations. The 500×1 vectors of losses are used to determine the model confidence set (MCS) (Hansen et al., 2011).[14]

The HQIC and SIC select the true model $\approx 100\%$ of the time for *most* experiments. With the exception of DGP2 (where model 2 is the DGP), there was no statistically significant difference between forecasts. For DGP2, the HQIC and SIC generally select model 2 98% to 100% of the time. OOS forecasts using the SIC slightly dominated those based on the HQIC, with both clearly dominating a strategy that only used model 1. Results therefore support model selection using either criterion.

We therefore only report the *minority* of exceptions where IC selected the true model $< 95\%$ of the time. We report results for $T = 100$ which are similar to $T = 200, 500$. Table 6 reports the percentage of times the HQIC and SIC selected each model. The final three columns report OOS MSE for three strategies: UO only employs model 1 and HQIC and SIC choose between models 1, 2 and 3. Most exceptions occur when the SNR is low, except DGP1 with a high SNR and $\zeta_1 = 0.9$ (consistent with section 5.1). Even though the HQIC selects the true model more often than the SIC, this doesn't necessarily improve forecasts. Approximately 50% of the time, model selection using IC performs as well or better than only using model 1. So if forecasting with a weak SNR, it may be optimal to employ a mis-specified, more parsimonious model.

[14]We also calculate OOS MAE and polynomial fit via the sum of squared and absolute deviations between the fitted and actual polynomial. Unreported MCS results are similar.

Table 6: Model selection when HQIC/SIC correct $< 95\%$ of replications $T = 100$

					% reps model selected		MSE		
DGP	SNR	x_t	m	Model	HQIC	SIC	UO	HQIC	SIC
1	low	iid	40	M1	87%	71%	0.0166***	0.0169	0.0173
				M2	9%	18%			
				M3	4%	11%			
	high	AR(1)	40	M1	94%	94%	0.0163***	0.0163***	0.0163***
				M2	6%	6%			
				M3	0%	0%			
2	low	iid	40	M1	4%	1%	0.0167	0.0167***	0.0167
				M2	78%	63%			
				M3	18%	36%			
3	low	iid	40	M1	92%	80%	0.0167***	0.0170	0.0174
				M2	4%	5%			
				M3	4%	15%			
4	low	iid	40	M1	26%	6%	0.0167*	0.0167*	0.0166***
				M2	40%	29%			
				M3	34%	65%			
5	low	iid	40	M1	13%	3%	0.0166	0.0163	0.0161***
				M2	20%	8%			
				M3	67%	89%			

Columns 6 and 7 report the percentage of times the HQIC and SIC select M1 = UO-MIDAS (unconstrained), M2 = UO-MIDAS ($\alpha = 0$), M3 = no predictability. M1 is the true model for all DGPs except for DGP2 which is M2. AR(1) parameter $\zeta_1 = 0.9$. The last three columns report OOS MSE using three strategies: UO employs M1 every replication, HQIC and SIC use information criteria to select between M1, M2 and M3. ***,**,* denotes inclusion in the MCS at the 25% ($M^*_{0.25}$), 10% ($M^*_{0.10}$) and 5% ($M^*_{0.10}$) levels respectively. For all other DGPs, the HQIC and SIC generally select the correct model $\approx 100\%$ of the time and there is no difference in MSE.

5.3 UO-MIDAS model performance relative to benchmarks

The performance of the UO-MIDAS model is now evaluated relative to the Almon-MIDAS, NP-MIDAS and PDL-MIDAS models. Section 5.3.1 considers the five UO-MIDAS DGPs and section 5.3.2 considers four Almon-MIDAS DGPs.

5.3.1 UO-MIDAS DGP

For the 5 UO-MIDAS DGPs, we estimate the UO-MIDAS model and its benchmarks (Almon, NP, PDL).[15] We evaluate polynomial fit and OOS forecast performance in the same way as section 5.2, and determine which of the four models is in the MCS.

Table 7 presents squared deviations by SNR for $T = 100$ when $x_t^{(m)} \sim i.i.d$ and $x_t^{(m)} \sim$

[15]Unreported results show that UO-MIDAS performance is robust to overstating the lag length, so UO and Almon-MIDAS are estimated via NLLS with the maximum K. For NP-MIDAS we follow Breitung and Roling (2015) and initially apply the maximum K in Table 2 and determine the regularisation parameter via numerical optimisation of a modified AIC. K is then reduced by the step size in Table 2 and the model re-estimated. This continues until the minimum K and the adjusted R^2 used to select the final model. PDL-MIDAS estimation commences with the maximum K and polynomial orders from $P = 3$ to 10. The best model is selected via an AIC that penalises according to P. The lag is then reduced by the step size and the procedure repeated. From the set of best models for each K, the final model is selected using an AIC that penalises according to K.

$AR(1) : \varsigma = 0.9$. Results for OOS MSE and $T = 200, 500$ are similar. For most DGPs, SNRs, aggregation levels and sample sizes, UO-MIDAS is the only model in the MCS and PDL-MIDAS is never included. Limited exceptions indicate that UO-MIDAS dominance is less likely if all of the following applies: little or no oscillation, a low SNR and $x_t^{(m)} \sim i.i.d.$ Under these circumstances Almon or NP-MIDAS *may* provide the best fit and OOS forecast.[16] Further, when the SNR is strong, $T \geq 200$ and $\varsigma = 0.9$, UO-MIDAS generally outperformed the alternatives. Therefore the parameter bias (in section 5.1) rarely results in a poor polynomial fit or forecast. Caution is recommended though, as Table 7 reveals the fit was poor for DGP5 when the SNR was strong and $\varsigma = 0.9$. Finally, when $x_t^{(m)}$ is persistent and the SNR high, an increase in T may see NP-MIDAS forecast slightly better than UO-MIDAS, however both models outperform Almon and PDL-MIDAS.

Table 7: Squared deviations×100 in lag weights: UO-MIDAS DGP, T=100

DGP		$x_t^{(m)} \sim i.i.d$				$x_t^{(m)} \sim AR(1) : \varsigma = 0.9$			
Signal	m	UO	Alm	NP	PDL	UO	Alm	NP	PDL
DGP1									
low	40	0.096***	0.163	0.106	0.295	0.002***	0.136	0.013	0.136
medium	40	0.075***	2.253	0.285	2.304	0.007***	2.616	0.209	3.331
high	40	0.086***	8.903	0.714	8.676	0.022***	12.348	0.824	13.301
DGP2									
low	40	0.090	0.075	0.062***	0.133	0.001***	0.014	0.003	0.002
medium	40	0.056	0.047***	0.091	0.105	0.001***	0.247	0.006	0.004
high	40	0.052	0.043***	0.110	0.114	0.001***	0.964	0.009	0.013
DGP3									
low	40	0.117***	0.235	0.144	0.452	0.002***	0.215	0.015	0.170
medium	40	0.078***	3.471	0.341	4.283	0.007***	4.701	0.194	3.856
high	40	0.088***	13.763	0.766	16.424	0.020***	19.254	0.751	15.412
DGP4									
low	40	0.093	0.078	0.068***	0.161	0.001***	0.024	0.004	0.016
medium	40	0.074***	0.099	0.117	0.172	0.001***	0.753	0.014	0.015
high	40	0.069***	0.251	0.148	0.143	0.002***	4.128	0.039	0.038
DGP5									
low	40	0.094	0.095	0.069***	0.175	0.002***	0.020	0.006	0.023
medium	40	0.075***	0.667	0.144	0.323	0.002***	0.522	0.012	0.258
high	40	0.068***	2.444	0.172	0.611	1.505	2.323	0.020***	0.742

Reports the sum of squared deviations between the actual (UO-MIDAS model) and estimated lag weights by SNR (low, medium, high). Models in the MCS at $\hat{M}^*_{75\%}$,$\hat{M}^*_{90\%}$, $\hat{M}^*_{95\%}$ are identified by ***,**, and * respectively. MCS results for OOS forecasting are similar.

5.3.2 Almon-MIDAS DGP

We now perform a similar experiment to section 5.3.1, with data generated via the four Almon-MIDAS DGPs in Andreou et al. (2010); DGP 1: slow decay $\theta = (0.0007, -0.006)$,

[16]The limited exceptions are: i) DGP 2 when $x_t^{(m)} \sim i.i.d$ where Almon-MIDAS provides the best fit and forecasts; ii) DGP4 and 5 when $x_t^{(m)} \sim i.i.d$ and the SNR low, where Almon and NP-MIDAS perform well and iii) DGP5 when $\varsigma = 0.9$ and the SNR high, where NP-MIDAS is the best.

DGP 2: fast decay $\theta = (0.0007, -0.05)$, DGP 3: inverse U $\theta = (0.3, -0.006)$, and DGP 4: near flat $\theta = (0, -0.0005)$. We set $\beta_0 = 0.5$ and consider low, medium and high SNRs via $\beta_1 = 0.6, 1.5$ and 3 respectively with $m = 3, 40, 100$; and $T = 100, 200, 500$.[17] [18]

Table 8 reports illustrative results when $x_t^{(m)} \sim i.i.d$, $m = 40$ and $T = 100$. MCS results for polynomial fit mirror OOS MSE. Almon-MIDAS is very often the only model in the MCS, but as the SNR increases UO-MIDAS may provide comparable performance (DGP2, 4).[19] Despite this, polynomial estimates from all models provide very good approximations that improve with increases in T. Almon-MIDAS may therefore provide the best polynomial fit and OOS MSE, however it's only slightly better than UO-MIDAS which slightly outperforms NP and PDL-MIDAS. The exception is DGP3 (inverse U) where Almon-MIDAS is clearly superior.

Table 8: Almon-MIDAS DGP, $x_t^{(m)} \sim i.i.d$, T=100

DGP		Squared deviations×100				Out of sample MSE			
Signal	m	UO	Alm	NP	PDL	UO	Alm	NP	PDL
DGP 1									
low	40	0.0660	0.0535***	0.1001	0.1072	0.0162	0.0160***	0.0165	0.0166
med	40	0.0709	0.0526***	0.1523	0.1389	0.0162	0.0160***	0.0170	0.0169
high	40	0.0908	0.0522***	0.1796	0.1395	0.0164	0.0160***	0.0173	0.0169
DGP 2									
low	40	0.0679	0.0492***	0.0927	0.1290	0.0165	0.0163***	0.0168	0.0171
med	40	0.0722	0.0483***	0.1114	0.1522	0.0166	0.0163***	0.0170	0.0174
high	40	0.0970***	0.9332	0.1413	0.1517	0.0168***	0.0257	0.0173	0.0174
DGP 3									
low	40	0.1453	0.0568***	0.2856	0.6201	0.0173	0.0164***	0.0187	0.0221
med	40	0.5465	0.0531***	0.2423	1.6237	0.0213	0.0163***	0.0182	0.0323
high	40	1.9961	0.0526***	0.2333	3.6090	0.0358	0.0163***	0.0181	0.0524
DGP 4									
low	40	0.0601***	0.0591***	0.0670	0.0787	0.0162***	0.0162***	0.0163*	0.0165
med	40	0.0612***	0.0592***	0.0821	0.0923	0.0162***	0.0163***	0.0165	0.0166
high	40	0.0676***	0.1286	0.0935	0.1119	0.0163***	0.0171	0.0166	0.0168

Reports the sum of the squared deviations between the actual and estimated lag weights and OOS MSE by SNR (low, medium, high). Models in the MCS at $\hat{M}^*_{75\%}$,$\hat{M}^*_{90\%}$, $\hat{M}^*_{95\%}$ are identified by ***,**, and * respectively.

5.4 Concluding comments

In the presence of oscillating and non-oscillating lag polynomials, UO-MIDAS is well estimated via NLLS. The HQIC and SIC are shown to effectively select between UO-MIDAS

[17] The Almon (as opposed to the beta) polynomial is used given its ubiquity, ease of estimation, and smaller square errors (relative to the true weights) in finite samples (Andreou et al., 2013).

[18] Almon and UO-MIDAS apply the following lags: DGP1 lag 40, DGP2 lag 20, DGP3 lag 60, DGP4 lag 120, while NP and PDL-MIDAS employ lag optimisation in the following way: DGP1 lags 15-40 step size 1, DGP2 lags 8-20 step size 1, DGP3 lags 20-60 step size 1 and DGP4 lags 50-120 step size 5.

[19] MCS p values excluding Almon-MIDAS generally support UO-MIDAS over the alternatives, with the main exception being DGP 3 where NP-MIDAS dominates. PDL-MIDAS also performs well for DGP 1 and 2 when the signal is strong and $T = 500$.

models with and without oscillation and no predictability. Minor exceptions occurred when the SNR was weak, however even when the incorrect model was selected, forecast performance was as good or better than the correctly specified model $\approx 50\%$ of the time. High levels of polynomial oscillation with a highly persistent $x_t^{(m)}$, may result in parameter identification problems with increases in T, however this generally has little effect on forecast performance. Larger sample sizes increase the oscillations in $x_t^{(m)}$, and this may cause these identification issues via an interaction with the oscillating polynomial. Exploring this issue via near integrated asymptotics is left for further research.

6 Empirical application

We illustrate the proposed UO MIDAS model with two data sets. The first, DS1, forecasts the one month ahead change in the US monthly inflation rate using daily GSCI commodity index returns. The second, DS2, forecasts the annualised monthly growth rate in US industrial production (IP) using the daily CBOE Volatility Index (VIX).

For both data sets autocorrelation functions suggest an AR(2) process, so we estimate an AR(2) benchmark plus the following 13 models: UO, Almon, Beta, NP and PDL MIDAS with and without AR(2) dynamics (including the common factor version for UO, Almon and Beta). We apply the following nomenclature: UO-AR(2)-MIDAS, UO-comfac-AR(2)-MIDAS, and UO-MIDAS (no AR term) with UO MIDAS (no hyphen) representing all UO MIDAS models.

UO, Almon and Beta MIDAS are estimated via NLLS, while NP and PDL MIDAS employ least squares. Following Breitung and Roling (2015), NP MIDAS selects the lag via the adjusted R^2 and PDL MIDAS selects the lag and polynomial order via the AIC. We calculate standard errors for parameters via the HAC sandwich estimator and employ the square root of the diagonal elements of $\hat{J}\hat{\Sigma}\hat{J}'$ for the lag polynomial weights where[20]

$$\sqrt{T}(\hat{\psi}(\theta) - \psi(\theta)) \xrightarrow{d} N(0, J\Sigma J') \tag{13}$$

[20] $\Sigma = cA^{-1}BA^{-1}$, where $c = \frac{T}{T-(5+p)}$ is a degrees of freedom correction, $A = \partial^2 f(\theta^*)/[\partial\theta^*\partial\theta^{*'}]$ is the Hessian and B is a kernel weighted outer-product estimate $B = \sum_{j=-\infty}^{\infty} k(\frac{j}{b_T})\Gamma(j)$, where $k(\frac{j}{b_T})$ is a kernel estimator, $\Gamma(j) = \Gamma(-j)$ and $\Gamma(j) = \sum_{t=j+1}^{T} \frac{\partial f(\theta^*)_t}{\partial\theta^*} \frac{\partial f(\theta^*)'_{t-j}}{\partial\theta^*}$. We apply a Bartlett kernel with a bandwidth (b_T) equal to the rule of thumb estimate $0.75T^{1/3}$.

$\psi(\theta^*)$ is the $K \times 1$ weight vector, $\hat{\psi}(\theta^*)$ its estimate, and $J = \partial\psi(\theta^*)/\partial\theta^{*'}$ the Jacobian.[21]

6.1 DS1: US inflation forecasting using daily GSCI commodity returns

We examine commodity returns as a predictor of inflation because they are determined in auction markets and respond rapidly to shocks. Commodity prices do not improve 12 month inflation forecasts (Stock and Watson, 1999) and there is little evidence of cointegration (Pecchenino, 1992), however they granger cause inflation over the short run (Kyrtsou and Labys, 2006) and successfully forecast inflation one month ahead (Monteforte and Moretti, 2013; Breitung and Roling, 2015).

We consider monthly US Consumer price indices (P_t) from January 1972 to August 2018. Inflation rates ($\pi_t = 100 \times ((P_t \div P_{t-1}) - 1)$) are I(1) so we model their first differences.[22] We employ a 25 year rolling window and set $K = 125$ for the Almon, Beta and UO MIDAS models. For NP and PDL MIDAS we optimise K between 25 to 125.

Lag polynomials change shape at the end of 2006. Figure 3 presents plots for the first 120 estimation windows (conditional on the information at December 1996 to November 2006), and the remaining 140 windows (conditional on the information at December 2006 to July 2018). Figures plot the average lag weight and the 5th and 95th percentiles for the UO-AR(2)-MIDAS and Almon-AR(2)-MIDAS models across the estimation windows. Illustrative models for both sub-samples are reported in Table 9 along with model selection tests in Table 10, which select the UO-AR(2)-MIDAS model over its restricted versions (this also applied for the UO-AR(2)-comfac and UO-MIDAS models).

(Insert Figure 3)

Results show an initial positive relation between commodity returns and the change in the inflation rate. For all MIDAS models, the impact decays to zero after approximately five weeks (25 trading days). The first sub-sample has a weaker initial response, and the

[21] Numerical optimisation of non-linear MIDAS models may encounter estimation issues (Foroni et al., 2015). We estimate UO MIDAS by maximising the negative of the squared residuals using the sequential quadratic programming algorithm in Oxmetrics. Starting values for the AR(2) parameters are equal to partial acf values. Previous research (below) suggests a 6-12 month lag, so we start γ at 0.06, with our remaining starting values close to zero.

[22] First differences are de-seasonalised using the Census X-13 methodology in Eviews with seasonal dummies, additive outliers, and the TRAMA auto ARIMA selection. Breakpoint unit root tests on the inflation rate (with innovation or additive outliers with or without a trend) strongly reject the null of a unit root on allowing for a break at June 1982. We therefore also estimate all models using the inflation rate from July 1982. Our reported MIDAS models using first differences however provide better forecasts of the inflation *level*, as none of the MIDAS models in levels are in the MCS at 10%.

Table 9: Inflation MIDAS models

	UO-AR(2)		UO-comfac		UO			Almon-AR(2)		Beta-AR(2)	
	coeff	std err	coeff	std err	coeff	std err		coeff	std err	coeff	std err
1st sub-sample (Dec 96 to Nov06): representative models at April 30, 2003											
β_0	-0.010	0.009	-0.010	0.009	-0.007	0.007	β_0	-0.014	0.010	-0.014	0.010
λ_1	-0.328^{***}	0.070	-0.335^{***}	0.071	–	–	λ_1	-0.364^{***}	0.072	-0.360^{***}	0.071
λ_2	-0.220^{***}	0.074	-0.227^{***}	0.067	–	–	λ_2	-0.268^{***}	0.058	-0.271^{***}	0.058
γ	0.055	0.037	0.027^{*}	0.015	0.029^{*}	0.015		–	–	–	–
A	0.004	0.006	0.007	0.005	0.007	0.006	θ_1	0.133^{*}	0.080	2.298^{*}	1.178
B	0.023^{***}	0.008	0.018^{***}	0.004	0.018^{***}	0.005	θ_2	-0.005^{**}	0.002	16.503^{**}	7.147
α	0.094^{***}	0.010	0.095^{***}	0.006	0.095^{***}	0.008	β_1	0.349^{***}	0.060	0.338^{***}	0.061
LL	-11.197		-11.167		-12.719			-11.388		-11.655	
2nd sub-sample (Dec06 to Jul18): representative models as at June 30, 2008											
β_0	0.001	0.009	-0.005	0.009	-0.004	0.008	β_0	-0.016	0.010	-0.016	0.010
λ_1	-0.383^{***}	0.052	-0.301^{***}	0.059	–	–	λ_1	-0.318^{***}	0.054	-0.313^{***}	0.053
λ_2	-0.362^{***}	0.062	-0.379^{***}	0.070	–	–	λ_1	-0.379^{***}	0.060	-0.387^{***}	0.060
γ	0.043^{***}	0.009	0.038^{***}	0.008	0.049^{***}	0.007		–	–	–	–
A	0.033^{***}	0.006	0.030^{***}	0.006	0.036^{***}	0.007	θ_1	0.045	0.068	1.357^{***}	0.358
B	-0.018^{*}	0.010	-0.006	0.004	-0.005	0.015	θ_2	-0.005^{**}	0.002	14.930^{***}	3.327
α	0.035^{***}	0.006	0.055^{***}	0.005	0.065^{***}	0.015	β_1	0.440^{***}	0.067	0.430^{***}	0.066
LL	-10.475		-10.887		-12.992			-11.469		-11.622	

Reports illustrative estimates for MIDAS models fit to monthly changes in US inflation rates against daily GSCI commodity index returns. Models are estimated using a 25 year rolling window which commences January 1972 to December 1996. ***,**,* denotes statistical significance at 1%,5% and 10% respectively using HAC standard errors. LL denotes the log likelihood.

Table 10: Model selection via SIC and HQIC: UO-AR(2)-MIDAS

	April 2003		June 2008	
	SIC	HQIC	SIC	HQIC
UO	-3.1550	-3.2069	-3.2217	-3.2735
$\alpha = 0$	-3.1510	-3.1880	-3.1430	-3.1800
No predictability	-3.1061	-3.1283	-2.9951	-3.0174

Reports Schwarz information criteria (SIC) and Hannan-Quinn information criteria (HQIC) for the UO-AR(2)-MIDAS model, its restricted version $\alpha = 0$ and $w(j;\theta) = 0$ for $j \in \{1, ..., K\}$. Model ordering for the UO-AR(2)-comfac-MIDAS model is identical. The UO-MIDAS model also had the lowest SIC and HQIC, however when $\alpha = 0$ the model failed to converge and had the highest SIC/HQIC.

lag weights for UO, NP and PDL MIDAS models suggest mild overshooting. Figure 4 for example shows that as at April 2003, overshooting is statistically significant for the UO-AR(2)-comfac model, but its significance at the 5% level is less clear for the UO-AR(2) model. In the second sub-sample there is a larger initial response and stronger evidence of overshooting.

(Insert Figure 4)

Table 11 presents MSE and MAE for the one month ahead forecasts over the entire OOS period: January 1997 to August 2018, as well as the two sub-samples: January 1997 to December 2006 and January 2007 to August 2018. Over the entire OOS period, results strongly support the UO MIDAS model. At the 25% level, UO-AR(2)-MIDAS and UO-comfac-AR(2)-MIDAS are the only models in the MCS for the MAE. The MSE only includes one other model that also captures overshooting (PDL-AR(2)-MIDAS).

On considering the first sub-sample, most AR(2)-MIDAS models (including UO MIDAS) are in the MCS at 25%. Therefore even if overshooting is weak or not present, UO MIDAS forecasts are not compromised. Consistent with the higher level of overshooting in the second sub-sample, our results provide strong support for UO MIDAS, as Almon and Beta MIDAS are never in the MCS. UO MIDAS with AR(2) dynamics dominates, with some NP and PDL MIDAS models in the MCS at the 10% but not the 25% level.

Table 11: Out of sample forecasts: change in US inflation rate

	MSE			MAE		
	Jan 97-Aug 18	Jan 97-Dec 06	Jan 07-Aug 18	Jan 97-Aug 18	Jan 97-Dec 06	Jan 07-Aug 18
AR(2) models						
UO	0.0609**	0.0594**	0.0619**	0.1828**	0.1881**	0.1782**
Alm	0.0707*	0.0621**	0.0781	0.1923	0.1921**	0.1924
Beta	0.0716*	0.0628**	0.0791	0.1942	0.1920**	0.1961
NP	0.0725*	0.0681**	0.0763	0.2034	0.2058**	0.2014
PDL	0.0640**	0.0626**	0.0651*	0.1912	0.1933**	0.1893*
AR(2)	0.0839	0.0736	0.0928	0.2084	0.2026**	0.2135
Common factor models						
UO	0.0624**	0.0611**	0.0635**	0.1857**	0.1907**	0.1813**
Alm	0.0984	0.0967	0.0999	0.2317	0.2451	0.2203
Beta	0.0948	0.0913	0.0979	0.2296	0.2395	0.2211
Models with no AR term						
UO	0.0716	0.0760	0.0679*	0.2024	0.2188	0.1883**
Alm	0.0841	0.0846	0.0836	0.2131	0.2265	0.2016
Beta	0.0847	0.0852	0.0843	0.2140	0.2254	0.2043
NP	0.0685*	0.0725**	0.0651**	0.2021	0.2129*	0.1928
PDL	0.0700*	0.0723	0.0680*	0.2032	0.2165	0.1918*

Reports OOS MSE and MAE for monthly changes in US inflation over the entire OOS period, January 1997 to August 2018, and two sub-periods, January 1997 to December 2006 and January 2007 to August 2018. **,* denotes inclusion in the MCS at the 25% ($M^*_{0.25}$) and 10% ($M^*_{0.10}$) levels respectively. Note $M^*_{0.25} \subset M^*_{0.10}$.

6.2 DS2: US industrial production forecasting using the daily VIX

Now we model the one month ahead annualised growth rate in US industrial production (IP) as a function of the VIX (a proxy for uncertainty). We are motivated by the monthly structural VAR and theoretical model of the firm by Bloom (2009), which shows that a rise in the VIX initially decreases IP over the next 6 months. IP then overshoots with most of the recovery occurring over the following six months. Instead we use the MIDAS framework to regress monthly IP against a daily VIX and perform OOS forecasting. We employ IP from March 1992 to May 2021, with 20 year rolling windows and $K = 250$.

Like inflation, we identify a shift in the polynomials. Our first sub-sample consists of the first 98 estimation windows conditional on the information from February 2012 to March 2020, the second sub-sample is over the remaining 13 estimation windows. We report illustrative models for each sub-sample in Table 12 and model selection tests in Table 13. Out of the 14 models considered, UO-AR(2)-MIDAS consistently provides the best fit. The SIC and HQIC also clearly select UO-AR(2)-MIDAS over its restricted versions ($\alpha = 0$ and $w(j;\theta) = 0$ for all $j \in \{1, ..., K\}$).[23] Models without AR dynamics provide the best forecasts below, so we also report UO-MIDAS, Almon-MIDAS and Beta-MIDAS models. UO-comfac estimates are similar to those reported, but Almon-comfac and Beta-comfac models were dropped as they often failed to converge.

Figure 5 plots the average lag weight and the 5th and 95th percentiles for the UO-AR(2)-MIDAS models over the sub-samples.[24] Polynomials in the first sub-sample are similar to Bloom (2009) and suggest a rise in the VIX decreases IP over the first 50 lags (2.5 months), with overshooting occurring over the next 9 to 10 months. In the second sub-sample there is a much larger decrease in IP, a quicker recovery, and a shorter but more pronounced period of overshooting. This is consistent with the "V shaped recovery" after the initial shock from COVID-19. Further, over the second sub-period, all Beta MIDAS models estimate a *negative* relation between IP and the VIX ($\beta_1 < 0$), however 9 out of 13 Almon-MIDAS models estimate a *positive* response ($\beta_1 > 0$). The Beta-MIDAS models therefore capture the initial decrease in IP, while most of the Almon-MIDAS models (with $\beta_1 > 0$) capture

[23] The same ordering applied for the UO-AR(2)-comfac and UO-MIDAS models in February 2019. For September 2020, UO-MIDAS still had the lowest SIC/HQIC, but the model with $\alpha = 0$ didn't converge.

[24] Plots for the other UO MIDAS models are similar as are the 95% confidence intervals around the lag polynomials for each month.

the overshooting.

(Insert Figure 5)

Table 12: Industrial production MIDAS models

	UO-AR(2)		UO			Almon		Beta	
	coeff	std err	coeff	std err		coeff	std err	coeff	std err
1st sub-sample (Feb 12 to Mar 20): representative models at February 2019									
β_0	3.002**	1.369	1.175	3.580	β_0	7.565***	1.430	7.565***	1.449
λ_1	0.081	0.069	–	–	-	–	–	–	–
λ_2	0.140*	0.074	–	–	-	–	–	–	–
γ	0.030***	0.007	0.039**	0.018		–	–	–	–
A	−0.017***	0.005	−0.015**	0.007	θ_1	0.254	0.170	4.235**	1.912
B	0.054***	0.011	0.028***	0.008	θ_2	−0.007**	0.004	61.632	43.609
α	0.005***	0.001	0.010***	0.003	β_1	−0.326***	0.073	−0.326***	0.074
LL	-10402.5		-11574.1			-11593.6		-11581.5	
2nd sub-sample (Apr 20 to Apr 21): representative models at September 2020									
β_0	3.622	2.698	4.174***	1.202	β_0	−2.522***	0.301	7.7952***	3.0402
λ_1	0.466***	0.165	–	–	-	–	–	–	–
λ_2	−0.251	0.170	–	–	-	–	–	–	–
γ	0.069***	0.024	0.034*	0.021		–	–	–	–
A	−0.102	0.078	−0.064	0.048	θ_1	27.376***	2.8e−05	3.340***	1.184
B	0.035	0.044	0.009	0.024	θ_2	−0.264***	2.8e−05	168.280***	1.184
α	0.068***	0.014	0.055***	0.005	β_1	0.182***	0.064	−0.343*	0.176
LL	-23904.3		-29449.7			-36260.5		-34664.9	

Reports illustrative MIDAS models fit to one month ahead US industrial production growth rates (annualised) against a daily VIX. Models are estimated using a 20 year rolling window commencing March 1992 to February 2012. ***,**,* denotes statistical significance at 1%,5% and 10% respectively, using HAC standard errors. LL denotes the log likelihood.

Table 13: Model selection via SIC and HQIC: UO-AR(2)-MIDAS

	February 2019		September 2020	
	SIC	HQIC	SIC	HQIC
UO	3.9290	3.8684	4.7610	4.7004
$\alpha = 0$	4.0212	3.9779	4.8592	4.8159
No predictability	3.9851	3.9391	4.8764	4.8504

Reports Schwarz information criteria (SIC) and Hannan-Quinn information criteria (HQIC) for the UO-AR(2)-MIDAS model, its restricted version $\alpha = 0$ and $w(j;\theta) = 0$ for $j \in \{1, ..., K\}$. The same ordering applied for the UO-AR(2)-comfac-MIDAS model as well as UO-MIDAS in February 2019. For September 2020, UO-MIDAS had the lowest SIC/HQIC, but the model with $\alpha = 0$ did not converge and had the highest SIC/HQIC.

Table 14 reports the one month ahead OOS MSE and MAE for the entire period (March 2012 to May 2021) and the second sub-period (May 2020 to May 2021). The first sub-period results are similar to the entire OOS period. UO-MIDAS has the lowest MSE and MAE across the entire OOS period. The low power of MSE means that all models are in the MCS at 25%, however UO-MIDAS is the only model in the MAE MCS. For the second subsample most models are in the MSE MCS at 10%, however for the MAE only UO-MIDAS and Almon-MIDAS are included. Given the Almon-MIDAS can only capture the initial decrease or overshooting, its inclusion in the MCS is more likely a result of low power associated with

the short OOS period.[25]

Table 14: Out of sample forecasts: US industrial production growth rate

	MSE		MAE	
	Mar 12-May 21	May 20-May 21	Mar 12-May 21	May 20-May 21
AR(2) models				
UO	313.51***	1761.2*	9.051	29.650
Alm	319.66***	1859.2	9.533	32.277
Beta	313.61***	1792.8	9.413	31.733
NP	305.16***	1816.7	9.005	28.935
PDL	282.51***	1632.8**	9.045	29.259
AR(2)	308.93***	1610.9**	8.948	28.483
UO:com	306.87***	1704.3**	8.981	28.934
Models with no AR term				
UO	259.19***	1460.2***	8.148***	25.470***
Alm	275.64***	1594.5***	8.600	25.908***
Beta	283.19***	1652.8***	8.743	27.006
NP	300.55***	1823.1*	8.966	29.920
PDL	281.89***	1745.0**	8.796	29.421

Reports OOS MSE and MAE for the monthly growth rate in US industrial production (annualised) over the entire OOS period March 2012 to May 2021, as well as a subset from May 2020 to May 2021. The first subset (March 2012 to April 2020) is similar to the overall result. **,* denotes inclusion in the MCS at the 25% ($M^*_{0.25}$), 10% ($M^*_{0.10}$) levels respectively. Note $M^*_{0.25} \subset M^*_{0.10}$.

7 Conclusion

We proposed a flexible parametric MIDAS model that can model oscillating lag polynomials. The model is parsimonious, easy to estimate, and can approximate shapes commonly captured by the exponential Almon and Beta lag. The model includes a parameter that captures the rate of decay, so endpoint restrictions are not required. We show that the SIC and HQIC can consistently select between the proposed model, a version with no oscillation, and no predictability. Empirical forecasting applications supported overshooting, with the model outperforming a variety of MIDAS benchmarks. Further research may consider other variables that overshoot (like exchange rates), as well as applying the UO polynomial to other models like GARCH-MIDAS (Engle et al., 2013), DCC-MIDAS (Colacito et al., 2011) and VAR-MIDAS (Ghysels, 2016).

[25] Bloom (2009) shows US employment has a similar though much weaker, marginally significant response to the VIX. UO-MIDAS models (annualised growth rate in non farm employee numbers (s.a) against a daily VIX) find similar results, with forecasts no better than Almon and Beta-MIDAS.

Bibliography

Almon, S., 1965. The distributed lag between capital appropriations and expenditures. Econometrica: Journal of the Econometric Society , 178–196.

Amemiya, T., Morimune, K., 1974. Selecting the optimal order of polynomial in the Almon distributed lag. The Review of Economics and Statistics , 378–386.

Andreou, E., Ghysels, E., Kourtellos, A., 2010. Regression models with mixed sampling frequencies. Journal of Econometrics 158, 246–261.

Andreou, E., Ghysels, E., Kourtellos, A., 2011. Forecasting with mixed-frequency data, in: The Oxford handbook of economic forecasting.

Andreou, E., Ghysels, E., Kourtellos, A., 2013. Should macroeconomic forecasters use daily financial data and how? Journal of Business & Economic Statistics 31, 240–251.

Babii, A., Ghysels, E., Striaukas, J., 2022. Machine learning time series regressions with an application to nowcasting. Journal of Business & Economic Statistics 40, 1094–1106.

Baillie, R.T., Morana, C., 2009. Modelling long memory and structural breaks in conditional variances: An adaptive FIGARCH approach. Journal of Economic Dynamics and Control 33, 1577–1592.

Barnett, W.A., Chen, G., et al., 2015. Bifurcation of macroeconometric models and robustness of dynamical inferences. Foundations and Trends® in Econometrics 8, 1–144.

Bergstrom, A., 1966. Monetary phenomena and economic growth: a synthesis of neoclassical and keynesian theories. The Economic Studies Quarterly (Tokyo. 1950) 17, 1–8.

Bergstrom, A.R., Nowman, K., Wymer, C., 1992. Gaussian estimation of a second order continuous time macroeconometric model of the UK. Economic Modelling 9, 313–351.

Bloom, N., 2009. The impact of uncertainty shocks. Econometrica 77, 623–685.

Breitung, J., Roling, C., 2015. Forecasting inflation rates using daily data: A nonparametric MIDAS approach. Journal of Forecasting .

Clements, M.P., Galvão, A.B., 2008. Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. Journal of Business & Economic Statistics 26, 546–554.

Colacito, R., Engle, R.F., Ghysels, E., 2011. A component model for dynamic correlations. Journal of Econometrics 164, 45–59.

Corradi, C., Gambetta, G., 1976. The estimation of distributed lags by spline functions. Empirical Economics 1, 41–51.

Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.

Dornbusch, R., 1976. Expectations and exchange rate dynamics. Journal of Political Economy 84, 1161–1176.

Enders, W., Jones, P., 2016. Grain prices, oil prices, and multiple smooth breaks in a VAR. Studies in Nonlinear Dynamics & Econometrics 20, 399–419.

Enders, W., Lee, J., 2012. A unit root test using a fourier series to approximate smooth breaks. Oxford bulletin of Economics and Statistics 74, 574–599.

Engle, R.F., Ghysels, E., Sohn, B., 2013. Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics 95, 776–797.

Ercolani, J.S., 2014. Cyclical activity and gestation lags in investment. The Manchester School 82, 620–630.

Foroni, C., Marcellino, M., Schumacher, C., 2015. Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. Journal of the Royal Statistical Society: Series A (Statistics in Society) 178, 57–82.

Forsberg, L., Ghysels, E., 2007. Why do absolute returns predict volatility so well? Journal of Financial Econometrics 5, 31–67.

Frankel, J.A., 1986. Expectations and commodity price dynamics: The overshooting model. American Journal of Agricultural Economics 68, 344–348.

Ghysels, E., 2013. Matlab toolbox for mixed sampling frequency data analysis using midas regression models. Available on MATLAB Central at http://www. mathworks. com/matlabcentral/fileexchange/45150-midas-regression .

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

Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The MIDAS touch: Mixed data sampling regression models .

Godfrey, L., Poskitt, D., 1975. Testing the restrictions of the Almon lag technique. Journal of the American Statistical Association 70, 105–108.

Hamlen, S.S., Hamlen Jr, W.A., 1978. Harmonic alternatives to the Almon polynomial technique. Journal of Econometrics 7, 57–66.

Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica: Journal of the econometric society , 413–430.

Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79, 453–497.

Hendry, D.F., Pagan, A.R., Sargan, J.D., 1984. Dynamic specification. Handbook of econometrics 2, 1023–1100.

Kyrtsou, C., Labys, W.C., 2006. Evidence for chaotic dependence between US inflation and commodity prices. Journal of Macroeconomics 28, 256–266.

Monteforte, L., Moretti, G., 2013. Real-time forecasts of inflation: The role of financial variables. Journal of Forecasting 32, 51–61.

Pecchenino, R.A., 1992. Commodity prices and the CPI: Cointegration, information, and signal extraction. International Journal of Forecasting 7, 493–500.

Pettenuzzo, D., Timmermann, A., Valkanov, R., 2016. A MIDAS approach to modeling first and second moment dynamics. Journal of Econometrics 193, 315–334.

Shiller, R.J., 1973. A distributed lag estimator derived from smoothness priors. Econometrica: Journal of the Econometric Society , 775–788.

Stock, J.H., Watson, M.W., 1999. Forecasting inflation. Journal of Monetary Economics 44, 293–335.

Trivedi, P., 1970. A note on the application of Almon's method of calculating distributed lag coefficients. Metroeconomica 22, 281–286.

Figure 1: Polynomial decay as a function of γ

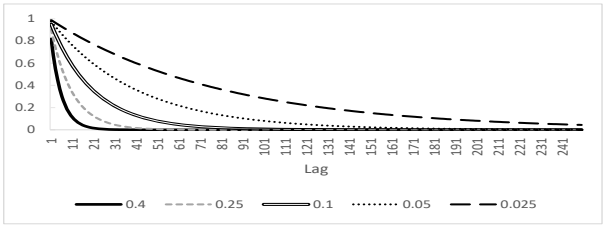


Figure 2: UO-MIDAS model lag polynomials

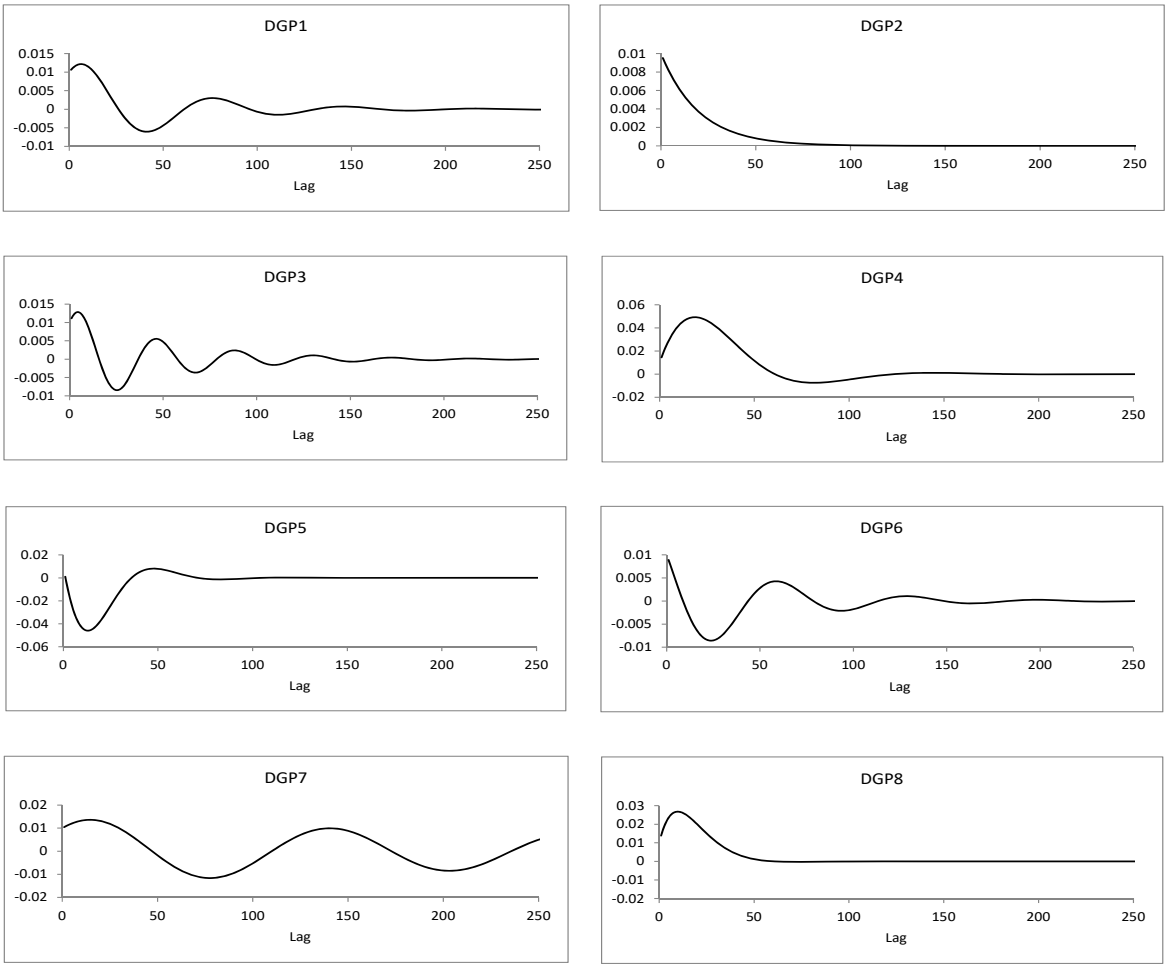
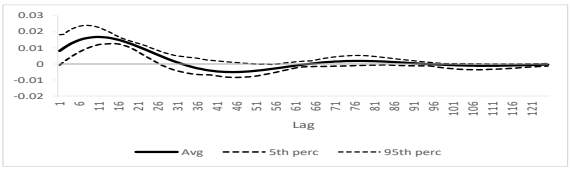
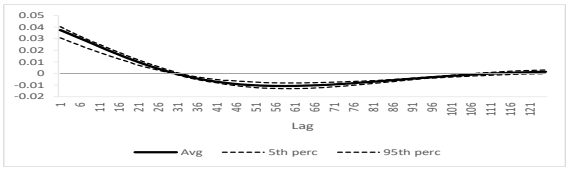


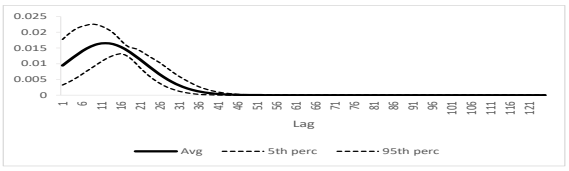
Figure 3: Average inflation lag polynomials with 5th and 95th percentiles



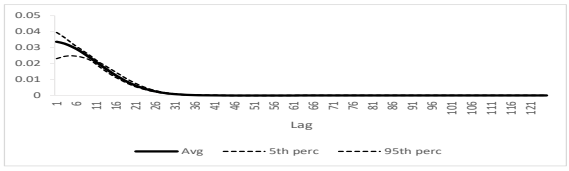
(a) UO-AR(2) Dec96-Nov06



(b) UO-AR(2) Dec06-Jul18

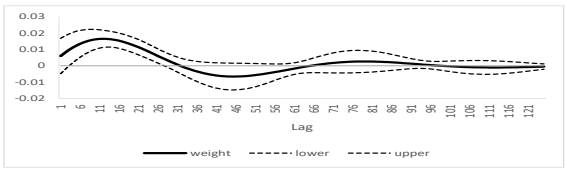


(c) Alm-AR(2) Dec96-Nov06

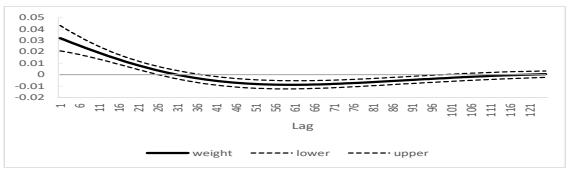


(d) Alm-AR(2) Dec06-Jul18

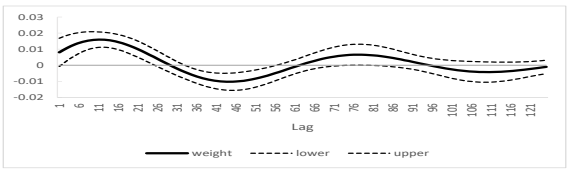
Figure 4: Illustrative inflation lag polynomials with 95% confidence intervals



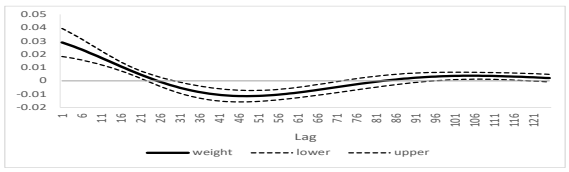
(a) UO-AR(2)-MIDAS April 2003



(b) UO-AR(2)-MIDAS June 2008

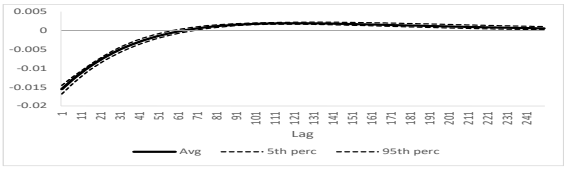


(c) UO-comfac-MIDAS April 2003

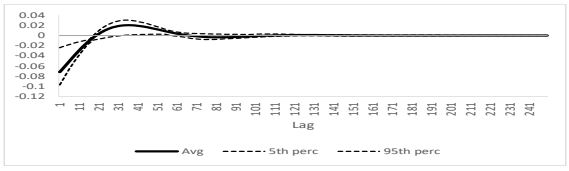


(d) UO-comfac-MIDAS June 2008

Figure 5: Average industrial production lag polynomials with 5th and 95th percentiles



(a) UO-AR(2) Feb12-Mar20



(b) UO-AR(2) Apr20-Apr21

A Appendix: Proof of Theorem

The following steps show that the probability of selecting a model that imposes an *incorrect* restriction converges to zero, that is:

(a)* If $(A_0, B_0) \neq (0,0)$ and $\alpha_0 \neq 0$ or π then $P(\xi_a < \xi_b, \xi_c, \xi_d) \to 1$.

(c)* If $(A_0, B_0) \neq (0,0)$ and $\alpha_0 = 0$ then $P(\xi_c < \xi_b, \xi_d) \to 1$.

(d)* If $(A_0, B_0) \neq (0,0)$ and $\alpha_0 = \pi$ then $P(\xi_d < \xi_b, \xi_c) \to 1$.

To show these, consider any two models i, j with $p_i \geq p_j$. The difference between their ICs can be written

$$\begin{aligned} \xi_i - \xi_j &= \left(-2 \max_{\theta \in \Theta_i} \ell_n(\theta) + p_i q_n\right) - \left(-2 \max_{\theta \in \Theta_j} \ell_n(\theta) + p_j q_n\right) \\ &= -\lambda_{i,j} + (p_i - p_j) q_n \end{aligned} \tag{14}$$

where $\lambda_{i,j}$ is the likelihood ratio statistic for testing the null hypothesis of model j against the alternative of model i. If $\lambda_{i,j} \xrightarrow{p} +\infty$ at rate $O_p(n)$ then $\xi_i - \xi_j \xrightarrow{p} -\infty$ and hence $P(\xi_i - \xi_j < 0) \to 1$.

For any model $i = a, b, c, d$ the log-likelihood maximisation can be expressed

$$\max_{\theta \in \Theta_i} \ell_n(\theta) = -\frac{n}{2} \log(1 + 2\pi) - \frac{n}{2} \log \min_{\theta \in \Theta_i} \hat{\sigma}_n^2(\theta),$$

so the likelihood ratio statistic $\lambda_{i,j}$ is

$$\lambda_{i,j} = 2 \left(\max_{\theta \in \Theta_i} \ell_n(\theta) - \max_{\theta \in \Theta_j} \ell_n(\theta) \right) = n \log \frac{\min_{\theta \in \Theta_j} \hat{\sigma}_n^2(\theta)}{\min_{\theta \in \Theta_i} \hat{\sigma}_n^2(\theta)}. \tag{15}$$

In the following we use the notation $W_0 = (w(1;\theta_0), \ldots, w(K;\theta_0))'$. The error variance function can be written

$$\begin{aligned} \hat{\sigma}_n^2(\theta) = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2 &- 2 \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t X_{t-1}' \left(W(\theta) - W_0\right) \\ &+ (W(\theta) - W_0)' \frac{1}{n} \sum_{t=1}^{n} X_{t-1} X_{t-1}' (W(\theta) - W_0) \end{aligned}$$

so that, under Assumptions (C) and (D),

$$\hat{\sigma}_n^2(\theta) \xrightarrow{p} \sigma_0^2 + (W(\theta) - W_0)'\Sigma_{XX}(W(\theta) - W_0) := \sigma^2(\theta)$$

pointwise for each θ where $\Sigma_{XX} = E(X_{t-1}X_{t-1}')$. For any model defined by parameter space Θ_m, compactness of Θ_m and continuity of $W(\theta)$ implies this convergence is uniform in θ and a minimiser exists, so that

$$\begin{aligned}\min_{\theta\in\Theta_m} \hat{\sigma}_n^2(\theta) &\xrightarrow{p} \min_{\theta\in\Theta_m} \sigma^2(\theta) \\ &= \sigma_0^2 + \min_{\theta\in\Theta_m} (W(\theta) - W_0)'\Sigma_{XX}(W(\theta) - W_0) \\ &:= \sigma_0^2 + \omega_m'\Sigma_{XX}\omega_m.\end{aligned}$$

If $\theta_0 \in \Theta_m$ then minimisation is obtained by $\theta = \theta_0$ and hence $\omega_m = 0$, so that

$$\min_{\theta\in\Theta_m} \hat{\sigma}_n^2(\theta) \xrightarrow{p} \sigma_0^2.$$

If $\theta_0 \notin \Theta_m$ then $\omega_m \neq 0$ and hence

$$\min_{\theta\in\Theta_m} \hat{\sigma}_n^2(\theta) \xrightarrow{p} \sigma_0^2 + \omega_m'\Sigma_{XX}\omega_m > \sigma_0^2,$$

the final inequality from the positive definitiveness of Σ_{XX} (Assumption (E)).

Considering the expression for $\lambda_{i,j}$ in (15), if model i is defined so that $\theta_0 \in \Theta_i$ and model j so that $\theta_0 \notin \Theta_j$ then it follows that

$$\lambda_{i,j} = n\left(\log\frac{\sigma_0^2 + \omega_m'\Sigma_{XX}\omega_m}{\sigma_0^2} + o_p(1)\right)$$

and hence, from (14),

$$\xi_i - \xi_j = -n\left(\log\frac{\sigma_0^2 + \omega_m'\Sigma_{XX}\omega_m}{\sigma_0^2} + o_p(1)\right) + (p_i - p_j)q_n.$$

The first term diverges to $-\infty$ at rate $O_p(n)$ which the second term diverges to $+\infty$ at rate $o(n)$. Hence $\xi_i - \xi_j \xrightarrow{p} -\infty$ and

$$P(\xi_i - \xi_j < 0) \to 1.$$

Now to show that models that impose a *correct* restriction will be selected over less restricted models.

(b)** If $(A_0, B_0) = (0, 0)$ then $P(\xi_b < \xi_a.\xi_c, \xi_d) \to 1$.

(c)** If $(A_0, B_0) \neq (0, 0)$ and $\alpha_0 = 0$ then $P(\xi_c < \xi_a) \to 1$.

(d)** If $(A_0, B_0) \neq (0, 0)$ and $\alpha_0 = \pi$ then $P(\xi_d < \xi_a) \to 1$.

We will show (c)** in detail, with the others following similarly. For any B define the likelihood ratio process

$$\lambda_{a,c}^{(n)}(B) = 2\left(\max_{\gamma,A,\alpha} \ell_n(\gamma, A, \alpha, B) - \max_{\gamma,A} \ell_n(\gamma, A, 0, B)\right).$$

When $\alpha_0 = 0$, for each $B \in \mathcal{B}$ standard likelihood ratio asymptotics gives

$$\lambda_{a,c}^{(n)}(B) \xrightarrow{d} \lambda_{a,c}(B)$$

with $\lambda_{a,c}(B) \sim \chi_1^2$, and hence $O_p(1)$ pointwise for each B. We can then show

$$\sup_{B\in\mathcal{B}} \lambda_{a,c}^{(n)}(B) = O_p(1), \tag{16}$$

essentially from continuity and boundedness in B on the compact parameter space $\mathcal{B}$ and non-singularity of Σ_{XX}. Thus, since we assume $c_n \to +\infty$,

$$\xi_a - \xi_c = -\sup_{B\in\mathcal{B}} \lambda_n(B) + 2c_n \xrightarrow{p} +\infty$$

and hence $P(\xi_a - \xi_c > 0) \to 1$.

To elaborate on (16), define the notation $\eta = (\gamma, A)$ and, for any $B \in \mathcal{B}$, the unrestricted estimator

$$(\hat{\eta}(B), \hat{\alpha}(B)) = \arg\max_{\eta,\alpha} \ell_n(\eta, \alpha, B)$$

and restricted estimator (with $\alpha = 0$)

$$\tilde{\eta}(B) = \arg\max_{\eta} \ell_n(\eta, 0, B).$$

To construct the score and Hessian, define the derivatives

$$
\begin{aligned}
w_\alpha(j;\eta,\alpha,B) &= \frac{\partial w(j;\eta,\alpha,B)}{\partial \alpha} = e^{-\frac{\gamma}{2}j}[-Aj\sin(\alpha j) + Bj\cos(\alpha j)] \\
w_\gamma(j;\eta,\alpha,B) &= \frac{\partial w(j;\eta,\alpha,B)}{\partial \gamma} = -\frac{j}{2}e^{-\frac{\gamma}{2}j}[A\cos(\alpha j) + B\sin(\alpha j)] \\
w_A(j;\eta,\alpha,B) &= \frac{\partial w(j;\eta,\alpha,B)}{\partial A} = e^{-\frac{\gamma}{2}j}\sin(\alpha j),
\end{aligned}
$$

and then

$$
w_\eta(j;\eta,\alpha,B) = [w_\gamma(j;\eta,\alpha,B) \quad w_A(j;\eta,\alpha,B)]
$$

and

$$
W_\alpha(B) = \begin{pmatrix} w_\alpha(1;\eta_0,0,B) \\ \vdots \\ w_\alpha(K;\eta_0,0,B) \end{pmatrix}, \quad W_\eta(B) = \begin{pmatrix} w_\eta(1;\eta_0,0,B) \\ \vdots \\ w_\eta(K;\eta_0,0,B) \end{pmatrix}.
$$

The elements of the score evaluated at the true values $\eta = \eta_0$, $\alpha = 0$ (noting B is not identified when $\alpha = 0$) are denoted

$$
s_\alpha(B) = \left.\frac{\partial \ell_n(\eta,\alpha,B)}{\partial \alpha}\right|_{\eta=\eta_0,\alpha=0} = \frac{n}{\sigma_0^2}\frac{1}{n}\sum_{t=1}^{n} W_\alpha(B)' X_{t-1}\varepsilon_t \tag{17}
$$

and

$$
s_\eta(B) = \left.\frac{\partial \ell_n(\eta,\alpha,B)}{\partial \eta}\right|_{\eta=\eta_0,\alpha=0} = \frac{n}{\sigma_0^2}\frac{1}{n}\sum_{t=1}^{n} W_\eta(B)' X_{t-1}\varepsilon_t. \tag{18}
$$

The corresponding elements of the Hessian are

$$
\begin{aligned}
H_{\alpha\alpha}(B) &= \left.\frac{\partial^2 \ell_n(\eta,\alpha,B)}{\partial \alpha^2}\right|_{\eta=\eta_0,\alpha=0} \\
&= -\frac{n}{\sigma_0^2}\left(\frac{1}{n}\sum_{t=1}^{n}(W_\alpha(B)' X_{t-1})^2 + o_p(1)\right) \\
&= -\frac{n}{\sigma_0^2}\left(W_\alpha(B)'\Sigma_{XX}W_\alpha(B) + o_p(1)\right)
\end{aligned} \tag{19}
$$

and similarly

$$
H_{\eta\alpha}(B) = \left.\frac{\partial^2 \ell_n(\eta,\alpha,B)}{\partial \eta \partial \alpha}\right|_{\eta=\eta_0,\alpha=0} - \frac{n}{\sigma_0^2}\left(W_\eta(B)'\Sigma_{XX}W_\alpha(B) + o_p(1)\right) \tag{20}
$$

$$
H_{\eta\eta}(B) = \left.\frac{\partial^2 \ell_n(\eta,\alpha,B)}{\partial \eta \partial \eta'}\right|_{\eta=\eta_0,\alpha=0} - \frac{n}{\sigma_0^2}\left(W_\eta(B)'\Sigma_{XX}W_\eta(B) + o_p(1)\right) \tag{21}
$$

The usual expansion of the likelihood function is then

$$\ell_n(\tilde{\eta}(B),0;B) = \ell_n(\hat{\eta}(B),\hat{\alpha}(B);B) + \frac{1}{2}\begin{pmatrix}\hat{\eta}(B)-\tilde{\eta}(B)\\ \hat{\alpha}(B)\end{pmatrix}'\begin{pmatrix}H_{\eta\eta}(B) & H_{\alpha\eta}(B)\\ H_{\eta\alpha}(B) & H_{\alpha\alpha}(B)\end{pmatrix}\begin{pmatrix}\hat{\eta}(B)-\tilde{\eta}(B)\\ \hat{\alpha}(B)\end{pmatrix} + o_p(1) \tag{22}$$

with the remainder uniformly $o_p(1)$ in B over the compact set $\mathcal{B}$. Similarly the expansion for the unrestricted MLE is

$$\begin{pmatrix}\hat{\eta}(B)-\eta_0\\ \hat{\alpha}(B)\end{pmatrix} = -\begin{pmatrix}H_{\eta\eta}(B) & H_{\alpha\eta}(B)\\ H_{\eta\alpha}(B) & H_{\alpha\alpha}(B)\end{pmatrix}^{-1}\begin{pmatrix}s_\eta(B)\\ s_\alpha(B)\end{pmatrix} + o_p(1) \tag{23}$$

and for the restricted MLE

$$\begin{pmatrix}\hat{\eta}(B)-\eta_0\\ \hat{\alpha}(B)\end{pmatrix} = \begin{pmatrix}-H_{\eta\eta}(\eta_0,0;B)^{-1}s_\eta(B)\\ 0\end{pmatrix} + o_p(1) \tag{24}$$

Partitioned inversion of the unrestricted Hessian and difference between (23) and (24) gives

$$\begin{pmatrix}\hat{\eta}(B)-\tilde{\eta}(B)\\ \hat{\alpha}(B)\end{pmatrix} = \begin{pmatrix}H_{\eta\eta}(B)^{-1}H_{\eta\alpha}(B)\\ -1\end{pmatrix}H_{\alpha\alpha\cdot\eta}(B)^{-1}s_{\alpha\cdot\eta}(B) + o_p(1). \tag{25}$$

where

$$s_{\alpha\cdot\eta}(B) = s_\alpha(B) - H_{\alpha\eta}(B)H_{\eta\eta}(B)^{-1}s_\eta(B)$$
$$H_{\alpha\alpha\cdot\eta}(B) = H_{\alpha\alpha}(B) - H_{\alpha\eta}(B)H_{\eta\eta}(B)^{-1}H_{\eta\alpha}(B).$$

Substitution of (25) into (22) gives

$$\lambda_n(B) = -s_{\alpha\cdot\eta}(B)'H_{\alpha\alpha\cdot\eta}(B)^{-1}s_{\alpha\cdot\eta}(B) + o_p(1).$$

Now using (17)–(21) gives

$$\begin{aligned} s_{\alpha\cdot\eta}(B) &= \sqrt{n}\,\nu(B)'\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\Sigma_{XX}^{-1/2}X_{t-1}\frac{\varepsilon_t}{\sigma_0} + o_p(1) \\ &= \sqrt{n}\,\nu(B)'Z_n \end{aligned}$$

and

$$\begin{aligned} H_{\alpha\alpha\cdot\eta}(B) &= H_{\alpha\alpha}(B) - H_{\alpha\eta}(B)H_{\eta\eta}(B)^{-1}H_{\eta\alpha}(B) \\ &= -n\,\nu(B)'\nu(B) + o_p(1) \end{aligned}$$

where

$$\nu(B) = \sigma_0^{-1}\Sigma_{XX}^{1/2}(W_\alpha(B) - W_\eta(B)(W_\eta(B)'\Sigma_{XX}W_\eta(B))^{-1}W_\eta(B)'\Sigma_{XX}W_\alpha(B))$$

and

$$Z_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}V_t \xrightarrow{d} N(0, I_K)$$

where $V_t = \Sigma_{XX}^{-1/2}X_{t-1}\varepsilon_t/\sigma_0$ is a strictly stationary martingale difference sequence. Thus

$$\lambda_n(B) = \left(\frac{\nu(B)'Z_n}{(\nu(B)'\nu(B))^{1/2}} + o_p(1)\right)^2$$

which is clearly asymptotically χ_1^2 for each $B \in \mathcal{B}$. The compactness of $\mathcal{B}$, continuity of $W_\alpha(B)$ and $W_\eta(B)$, and non-singularity of Σ_{XX} further imply that $\sup_{B\in\mathcal{B}}\lambda_n(B) = O_p(1)$ as required for (16).

Combining (a)*, (c)*, (d)* and (b)**, (c)**, (d)** demostrates (a)–(d) in the statement of the Theorem.

fbe.unimelb.edu.au/finance