Dynamic panels with MIDAS covariates: Nonlinearity, estimation and fit

Lynda Khalaf, Maral Kichian, Charles Saunders, Marcel Voia

To cite this version:

Lynda Khalaf, Maral Kichian, Charles Saunders, Marcel Voia. Dynamic panels with MIDAS covariates: Nonlinearity, estimation and fit. Journal of Econometrics, 2021, 220 (2), pp.589-605. 10.1016/j.jeconom.2020.04.015 . hal-03528880

HAL Id: hal-03528880

https://univ-orleans.hal.science/hal-03528880

Submitted on 3 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0304407620301342
Manuscript_815d60d8a74cf3e1d7e6b6ee6063bc2b

Dynamic Panels with MIDAS Covariates: Nonlinearity, Estimation and Fit

Lynda Khalaf* Carleton University
Maral Kichian† University of Ottawa
Charles Saunders‡ University of Western Ontario
Marcel Voia§ University of Orléans

January 27, 2021

Abstract

This paper introduces Mixed Data Sampling (MIDAS) into the panel data context. To address the unidentified nuisance parameter problem, we propose to invert model specification tests for inference on the MIDAS parameter along with bounds tests for model coefficients. Illustrative identification, simulation and empirical analyses are conducted in the dynamic GMM framework. Our framework allows for departures from *i.i.d.* errors such as clustering and dynamic specifications. A simulation study and an application to a model of reserve holdings illustrate the usefulness of the proposed methods, and more broadly set a promising template for shrinkage approaches.

JEL Classification: C13, C15, C23, C33
Keywords: dynamic panel model, mixed data sampling, unidentified nuisance parameter, reserve holdings model

1 Introduction

A time dependent regression generally requires that all the variables have observations at the same frequency as the dependent variable (for example, annual). If a regressor is available at a higher frequency (e.g., monthly or weekly), common practice is to lower its frequency either by aggregating the higher-frequency data using equal weights, or by using the first, middle, or last high-frequency

*Economics Department and Centre for Monetary and Financial Economics (CMFE), Carleton University; Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Economics Department, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel: (613) 520-2600 ext. 8697; email: Lynda.Khalaf@carleton.ca.

†Economics Department and Graduate School of Public and International Affairs, University of Ottawa. 120 University, Social Sciences Building, Ottawa, Ontario, Canada, K1N 6N5. Tel: 613-562-5800 ext. 1872. Fax: 613-562-5241 email: mkichian@uottawa.ca.

‡Department of Economics, University of Western University. Mailing address: Department of Economics, Faculty of Social Science, University of Western Ontario, 1151 Richmond Street N., London, Ontario, N6A 5C2 Canada. email: csaund9@uwo.ca.

§Corresponding author. Laboratoire d'Économie d'Orléans, Faculté de Droit d'Économie et de Gestion, Rue de Blois - BP 26739, 45067 ORLÉANS Cedex 2. email: marcel.voia@univ-orleans.fr.

© 2020 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

observation as the representative value over the lower frequency interval. However, doing so may lead to a loss of information pertaining to the high frequency data, or it may introduce specification biases.

Ghysels, Santa-Clara, and Valkanov (2004) introduce the Mixed Data Sampling (MIDAS) framework as a means of dealing with such arbitrary aggregation schemes. MIDAS is by now well-known in the econometrics literature; see the special edition of the Journal of Econometrics (August 2016) for a recent overview and applications. See also Ghysels (2016), Götz, Hecq, and Smeekes (2016), Ghysels, Santa-Clara, and Valkanov (2004) for multivariate models, and more generally Clements and Galvão (2008), Kuzin, Marcellino, and Schumacher (2011), Bai, Ghysels, and Wright (2013), Guérin and Marcellino (2013), and Foroni and Marcellino (2013) for a comprehensive survey of recent advances.

Another way to deal with regressors available at higher frequencies is to include them *as is* in the regression, and to apply the unrestricted MIDAS (UMIDAS) estimation approach of Foroni, Marcellino, and Schumacher (2013). This option is only viable when resulting degrees of freedom are sufficiently large. Thus, a key benefit of MIDAS is that its polynomial aggregation structure limits degrees of freedom losses while allowing for possibly unequal weights. Commonly-used polynomials include the Almon distributed lag and the Beta distribution function. These allow for a variety of weighting schemes including equal weights, hump-shaped weights, and assigning more/less weight to recent versus older observations. This variety of weight profiles are obtained at a minimum cost of two additional parameters, which together constitute the MIDAS vector ($\boldsymbol{\theta}$).

To set focus consider a simple panel regression with a single high-frequency covariate that non-linearly embeds $\boldsymbol{\theta}$, and where the regression coefficient, β, is the parameter of interest. The presence of these parameter non-linearities complicate identification. In this context, $\boldsymbol{\theta}$ is weakly identified in the near-zero β sub-space. This is the well known Davies (1977, 1987) problem, and while it was addressed in the univariate case in Ghysels, Sinko, and Valkanov (2007), it remains to be studied in the MIDAS-panel literature. This partly motivates our work.

Indeed, with the exception of a few recent applications to vector autoregression models such as Ghysels (2016) and Binder and Krause (2014), the bulk of this literature remains univariate. In this paper we introduce MIDAS to panel data regressions including dynamic models suitable for analysis with GMM methods of the Anderson and Hsiao (1982) and Arellano and Bond (1991) (denoted as Arellano-Bond) form. To our knowledge, this is the first formal extension of MIDAS methods to the panel context.

Available time series procedures are not guaranteed to extend to dynamic panels due to the dual-indexing of observations, which exacerbates incidental parameter biases. In particular lags may cause various complications, including spurious seasonality as emphasized by Clements and Galvão (2008, 2009). In a GMM panel framework, the instruments are lags of the endogenous variables and the regressors, and the lag structure will thus interfere by construction. Formal specification checks are thus needed. Indeed such specification checks are rare for MIDAS, even in the univariate case (Kvedaras and Zemlys (2012); Miller (2013, 2018)). Lag and nuisance parameter specification issues provide further motivations for our work.

To address these problems, we propose an inference method that can be applied with any panel data estimator that is valid when $\boldsymbol{\theta}$ is known, and where the estimator can be paired with a specification test that varies with $\boldsymbol{\theta}$. Our identification and numerical work uses the Arellano-Bond estimator for illustrative purposes as it has been extensively studied in the dynamic panel literature. We pair this with the Sargan statistics, or alternatively, with tests for neglected dynamics, which are commonly used specification tests. Below, our methodology is presented beyond the GMM special case, and it can be summarized as follows.

For a fixed $\boldsymbol{\theta}$, the MIDAS regressor becomes an observable aggregation of the high frequency series. Estimation thus reverts to the standard context where two statistics are typically available: a criterion to test the significance of β given $\boldsymbol{\theta}$, and a diagnostic test to assess the structure, again knowing $\boldsymbol{\theta}$. We first construct a confidence set for $\boldsymbol{\theta}$, by inverting the diagnostic test. Formally, we collect the $\boldsymbol{\theta}$ values that are not rejected by this test at the desired level of significance. Next we put forth two bound tests for β: based on supremum p-value over a confidence set for $\boldsymbol{\theta}$, or over its entire parameter space.

Thus defined, our approach allows for the possibility of an empty confidence set for $\boldsymbol{\theta}$ which signals misspecification. We thus jointly address model specification and the underlying nonlinearity, as has been recently emphasized in the weak-IV literature; see Bun and Windmeijer (2010) for a panel data application, and more generally Stock and Wright (2000), Andrews and Cheng (2012), Dufour (1997), Kleibergen (2005), Ghysels and Wright (2010), and references therein.

We conduct a simulation study that illustrates size and power properties for our procedures. The simulations focus on exponential Almon lag polynomials and two different designs that allow us to quantify the extent of identification difficulties. Results show adequate size and good power for our proposed inference methods on both $\boldsymbol{\theta}$ and β.

We demonstrate the use of our methodology with an empirical application drawn from the international macroeconomics field[1]. In this literature a key variable is the exchange rate, and studies are often interested in its impact on various macroeconomic variables in the associated countries. We adapt the panel model of Obstfeld, Shambaugh, and Taylor (2010) on country reserve holdings to a panel-MIDAS setting that conforms with our proposed methods and we focus on the 'Horserace' models considered in Tables 1 and 2 of the original study. The high frequency series is the standard deviation of the monthly exchange rate, and we consider three aggregation schemes for this variable: equal-weights, MIDAS weights, and UMIDAS. Results reveal important policy-relevant differences between an equally-weighted and a MIDAS weighted volatility aggregate, both in the originally proposed cluster-corrected specification as well as in the alternative dynamic structure that we also analyze.

[1]Relevant examples from other fields such as political economy, development, exchange rate effects, and real-financial linkages; include Habib, Mileva, and Stracca (2017), Martin (2016), Li, Ma, and Xu (2015), Law and Singh (2014), Allegret, Couharde, Coulibaly, and Mignon (2014), Aghion, Bacchetta, Ranciere, and Rogoff (2009), Acemoglu, Johnson, Robinson, and Yared (2008), Beck and Levine (2004), and Bleaney and Greenaway (2001)

In Section 2 we set the framework. Our methology is presented in Section 3. Section 4 summarizes our simulation results. Our empirical analysis is discussed in Section 5 and Section 6 provides concluding comments.

The mixed-data frequency introduces challenges to the notation of the paper, since some series from the data is indexed by cross-section ($i = 1, \ldots, N$) and time ($t = 1, \ldots, T$) while other series include a third index for high-frequency time observations ($j = 1, \ldots, m$ for each t). Throughout the paper bold face lowercase variables represent vectors, and bold face uppercase variables represent matrices. To distinguish between a two-index and three-index vector (or matrix) we introduce a 'dot' subscript to indicate that all available observations of a vector (or matrix) are retained. The first example is a series of scalars with two indexes, $c_{i,t}$, which is used to construct a subvector of dimensions $T \times 1$ by stacking each cross-section by time, $\boldsymbol{c}_{i,\cdot} = (c_{i,1}, c_{i,2}, \ldots, c_{i,T})'$, then stacking the cross-sections results in a $NT \times 1$ vector, $\boldsymbol{c} = (\boldsymbol{c}'_{1,\cdot}, \boldsymbol{c}'_{2,\cdot}, \ldots, \boldsymbol{c}'_{N,\cdot})'$.

The second example is a series of scalars with three indexes, $d_{i,t,j}$, which is used to construct a vector of dimensions $(T-1) \times 1$ by stacking each cross-section and high-frequency by time, $\boldsymbol{d}_{i,\cdot,j} = (d_{i,2,j}, d_{i,3,j}, \ldots, d_{i,T,j})'$, then stacking the cross-section results in a $N(T-1) \times 1$ vector, $\boldsymbol{d}_{\cdot,\cdot,j} = (\boldsymbol{d}'_{1,\cdot,j}, \boldsymbol{d}'_{2,\cdot,j}, \ldots, \boldsymbol{d}'_{N,\cdot,j})'$, and finally concatenating these high-frequency vectors to construct the $N(T-1) \times m$ matrix, $\boldsymbol{D} = (\boldsymbol{d}_{\cdot,\cdot,1}, \boldsymbol{d}_{\cdot,\cdot,2}, \ldots, \boldsymbol{d}_{\cdot,\cdot,m})$.

Let $\mathsf{L_t}$ and $\mathsf{L_j}$ represent the lag-operators for t and j, respectively. For a given time-series vector, $\boldsymbol{d}_{i,\cdot,j}$, applying $\mathsf{L_t}$ gives $\mathsf{L_t}\boldsymbol{d}_{i,\cdot,j} = (d_{i,1,j}, d_{i,2,j}, \ldots, d_{i,T-1,j})'$. Applying $\mathsf{L_j}$ to the same vector gives,

$$\begin{aligned} \mathsf{L_j}\boldsymbol{d}_{i,\cdot,j} &= (d_{i,2,j-1}, d_{i,3,j-1}, \ldots, d_{i,T,j-1})' \text{ for } j = 2, 3, \ldots, m \\ \mathsf{L_j}\boldsymbol{d}_{i,\cdot,1} &= (d_{i,1,m}, d_{i,2,m}, \ldots, d_{i,T-1,m})' \text{ for } j = 1. \end{aligned}$$

The lag-operator raised to an integer power indicates the number of times the lag-operator is applied, for example $\mathsf{L_t}^p d_{i,t,j} = d_{i,t-p,j}$.

2 Panel framework with MIDAS regressors

Our proposed method introduces MIDAS regressors into a possibly dynamic panel framework. To set focus, we consider a prototypical single-regressor model of the form:

$$y_{i,t} = \delta y_{i,t-1} + \beta x_{i,t}(\boldsymbol{\theta}) + u_{i,t}, \tag{1}$$

$$x_{i,t}(\boldsymbol{\theta}) = \sum_{j=1}^{m} x_{i,t,j} w_j(\boldsymbol{\theta}), \tag{2}$$

$$u_{i,t} = \mu_i + \nu_{i,t}, \tag{3}$$

where $y_{i,t}$, $i = 1, ..., n$ and $t = 1, ..., T$, are observed dependent variables for cross-sectional unit i and time t, and $x_{i,t,j}$, $j = 1, ..., m$, are high frequency explanatory variables. The latter are aggregated into a covariate at frequency t using nuisance parameter dependent weights: $w_j(\boldsymbol{\theta})$ thus refers to these weighting functions whose functional form is assumed given up to the unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ denotes the relevant parameter space. Conformably, $x_{i,t}(\boldsymbol{\theta})$ denotes the latent regressor which we formally express as a function of $\boldsymbol{\theta}$.

In the above, the lag coefficient δ may be pre-set to zero (we allow for a standard panel with no dynamics). The regression disturbances conform with standard panel set-ups: with $\delta \neq 0$, $\mu_i \sim iid(0, \sigma_\mu^2)$, $\nu_{i,t} \sim iid(0, \sigma_\nu^2)$, and these two components are independent. In non-dynamic contexts, the *iid* assumption on $\nu_{i,t}$ can be relaxed in which case clustering may be considered. Additional exogenous or endogenous regressors may be included, which will of course affect estimation and inference. In fact, our general framework can also be extended beyond the above homogenous set-up and thus covers large N as well as large T motivated approaches.

In this paper, we follow Ghysels, Santa-Clara, and Valkanov (2004) and use the exponential Almon lag polynomial of length H [Almon (1965)] defined as,

$$w_j(\boldsymbol{\theta}) = \frac{\prod_{h=1}^{H} exp\{j^h \theta_h\}}{\sum_{k=1}^{m} \left(\prod_{h=1}^{H} exp\{k^h \theta_h\}\right)}, \tag{4}$$

where $\boldsymbol{\Theta} = \mathbb{R}^H$ so $\boldsymbol{\theta} \in \mathbb{R}^H$, and the weights sum to one. The equal weights assumption corresponds to $\boldsymbol{\theta} = \mathbf{0}$ that is

$$w_j(\mathbf{0}) = \frac{\prod_{h=1}^{H} exp\{j^h 0\}}{\sum_{k=1}^{m} \left(\prod_{h=1}^{H} exp\{k^h 0\}\right)} = \frac{1}{m}. \tag{5}$$

The exponential Almon lag with only two parameters can describe a variety of patterns for the weights. We thus set $H = 2$. Thus, $\boldsymbol{\theta} \in \mathbb{R}^2$ in our simulation design below, conforming with recent practice, see Appendix in Ghysels (2016) and Ghysels, Sinko, and Valkanov (2007). Other weighting functions, like the Beta distribution suggested in Ghysels et al. (2004), can also be used to construct the high-frequency weights with some minor modifications and estimation implications.

In this paper, the high frequency observations are equally spaced, but this is not a restriction. Ghysels, Sinko, and Valkanov (2007) discuss an implementation of irregularly spaced observations (e.g., tick-by-tick stock price data) in terms of the Almon lag function. McKenzie (2001), Millimet and McDonough (2017), and Sasaki and Xin (2017) find that irregular spacing means that the model will need to be respecified taking the nonuniform sampling into consideration. The latter two references correspond to GMM implementations; as such an appropriate specification test for the identification of $\boldsymbol{\theta}$ may be self-evident.

In addition to the advantages pointed out in Ghysels et al. (2006), the use of weighting functions means that the high frequency observations, $x_{i,t,j}$, will only intervene through the MIDAS aggregation. In panel frameworks, this has two important implications. First, the resulting lags of the MIDAS regressors are valid instruments in the standard panel IV setting, including the Arellano-Bond case. Second, the standard properties of the first difference transformation are preserved conditional on $\boldsymbol{\theta}$, if $\boldsymbol{\theta}$ is specified. Our approach builds on these two properties.

In particular, for any given $\boldsymbol{\theta}$, (1) can be rearranged leading to a regression in first differences (in t) where although unobserved heterogeneity, μ_i, is eliminated, $\Delta y_{i,t-1}$ remains endogenous unless δ is pre-set to zero. For further reference let

$$\Delta\boldsymbol{\nu}(\beta, \delta; \boldsymbol{\theta}) = \Delta\boldsymbol{y} - \delta\Delta\mathsf{L}_\mathsf{t}\boldsymbol{y} - \beta\Delta\boldsymbol{x}(\boldsymbol{\theta}) \tag{6}$$

refer to full set of difference error terms, where $\Delta \boldsymbol{x}(\boldsymbol{\theta}) = (\boldsymbol{X} - \mathsf{L}_{\mathsf{t}}\boldsymbol{X})\, \boldsymbol{w}(\boldsymbol{\theta})$, where

$$\begin{aligned} \boldsymbol{X} &= (\boldsymbol{x}_{\cdot,\cdot,1}, \boldsymbol{x}_{\cdot,\cdot,2}, \ldots, \boldsymbol{x}_{\cdot,\cdot,m}), \\ \boldsymbol{x}_{\cdot,\cdot,j} &= (\boldsymbol{x}'_{1,\cdot,j}, \boldsymbol{x}'_{2,\cdot,j}, \ldots, \boldsymbol{x}'_{N,\cdot,j})', \text{ and} \\ \boldsymbol{x}_{i,\cdot,j} &= (x_{i,3,j}, x_{i,4,j}, \ldots, x_{i,T,j})', \end{aligned}$$

also $\boldsymbol{X}$ is a $N(T-2) \times m$ matrix.

When β is zero, (1) is reduced to a dynamic panel model without regressors, which evacuates $\boldsymbol{\theta}$ from the regression equation. Even then, $\boldsymbol{\theta}$ may remain in the estimating equations through another channel, for example, when $x_{i,t}(\boldsymbol{\theta})$ is used as an instrument. Global identification is nevertheless compromised. This problem - which relates to the so-called Davies problem [Davies (1977, 1987)] - can be succinctly stated as the case where nuisance parameters, $\boldsymbol{\theta}$, are not present under the null, $\beta = 0$, but are present under the alternative, $\beta \neq 0$. The methodology we propose next aims to address this difficulty.

3 Methodology

Our approach involves an estimation method that is valid when $\boldsymbol{\theta}$ is fixed to a known value, say $\boldsymbol{\theta}_0$, implying that the MIDAS aggregate, denoted as $\boldsymbol{x}(\boldsymbol{\theta}_0) = \boldsymbol{X}\boldsymbol{w}(\boldsymbol{\theta}_0)$, is observable. In this context, various asymptotically valid methods exist for dynamic panel models depending on the magnitudes of N and T.

Assumption 1 *In the context of (1)-(3) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and where $\boldsymbol{\theta}_0$ is specified, the following statistics are available: (i) a test criterion, denoted $\mathsf{t}(\beta_0; \boldsymbol{\theta}_0)$, for which a p-value, denoted $p_{\mathsf{t}}(\beta_0; \boldsymbol{\theta}_0)$, can be obtained to assess*

$$H_{01} : \beta = \beta_0, \text{ for known } \beta_0 \text{ (given } \boldsymbol{\theta}_0), \tag{7}$$

where β_0 may or may not be 0, and (ii) a diagnostic criterion to assess the specification, denoted $\mathcal{J}(\boldsymbol{\theta}_0)$, such that under the H_{02} hypothesis of a correct specification for the given $\boldsymbol{\theta}_0$ a p-value can be computed that we denote as $p_{\mathcal{J}}(\boldsymbol{\theta}_0)$.

Special cases of Assumption 1 include - as with the Arellano and Bond GMM method, which we use below to set focus - a χ^2 limiting null distribution for $\mathcal{J}(\boldsymbol{\theta}_0)$ with known degrees-of-freedom, and an asymptotically standard normal $\mathsf{t}(\beta_0; \boldsymbol{\theta}_0)$. More broadly, the null distributions in question may depend on the given $\boldsymbol{\theta}_0$ as would occur with, for example, bootstrap-based approaches.

We propose to build a confidence set for the MIDAS parameters by inverting the considered $\mathcal{J}(\boldsymbol{\theta}_0)$ test. For inference on β, we propose bound-tests for H_{01}, as recommended in a general setting by Dufour (1989), by maximizing $p_{\mathsf{t}}(\beta_0; \boldsymbol{\theta}_0)$ [or minimizing $\mathsf{t}(\beta_0; \boldsymbol{\theta}_0)$ when relevant] over $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ or within the retained confidence set.[2]

To illustrate the feasibility of the proposed strategy and to concretize concepts, the Arellano-Bond framework provides a popular baseline case. So before we further formalize our general

[2]It is worth noting that Dufour (1989), which pre-dates the literature on MIDAS, did not consider dynamic panels.

inference strategy, we start by examining the identification of $\boldsymbol{\theta}$ through the Arellano-Bond moment conditions. This is key for obtaining procedures that are accurately informative on $\boldsymbol{\theta}$.

3.1 The nuisance parameter problem and inference on $\boldsymbol{\theta}$

This section focuses, as a motivational analysis, on the GMM method of Arellano-Bond when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_0$ is specified. We are interested in the informational content of the underlying moment conditions, to justify inverting a test based on the associated overidentifying restrictions test.

In this context, orthogonality conditions imply that there exists unique values of β and δ which we define $\beta(\boldsymbol{\theta}_0)$ and $\delta(\boldsymbol{\theta}_0)$ such that

$$E\left[\boldsymbol{W}(\boldsymbol{\theta}_0)'\Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_0)\right] = \mathbf{0}, \text{ for } \beta = \beta(\boldsymbol{\theta}_0),\ \delta = \delta(\boldsymbol{\theta}_0) \text{ and } \boldsymbol{\theta}_0 \text{ is specified,} \tag{8}$$

where $\boldsymbol{W}(\boldsymbol{\theta}_0)$ refers to the considered instruments; importantly, appropriate lags of the aggregated regressors can be used as instruments given $\boldsymbol{\theta}_0$.[3] The resulting (population) moment conditions identify β and δ given $\boldsymbol{\theta}_0$.

Consider two vectors, $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_A$, applied to (6) to give

$$\Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_0) = \Delta\boldsymbol{y} - \delta\Delta\mathsf{L}_\mathsf{t}\boldsymbol{y} - \beta\Delta\boldsymbol{x}(\boldsymbol{\theta}_0), \text{ and} \tag{9}$$

$$\Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_A) = \Delta\boldsymbol{y} - \delta\Delta\mathsf{L}_\mathsf{t}\boldsymbol{y} - \beta\Delta\boldsymbol{x}(\boldsymbol{\theta}_A). \tag{10}$$

Combining these equations we obtain the following,

$$\Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_0) = \Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_A) + \Delta\boldsymbol{\eta}(\beta,\boldsymbol{\theta}_A,\boldsymbol{\theta}_0), \text{ where} \tag{11}$$

$$\Delta\boldsymbol{\eta}(\beta;\boldsymbol{\theta}_A,\boldsymbol{\theta}_0) = \left(\Delta\boldsymbol{x}(\boldsymbol{\theta}_A) - \Delta\boldsymbol{x}(\boldsymbol{\theta}_0)\right)\beta = \Delta\boldsymbol{X}\left(\boldsymbol{w}(\boldsymbol{\theta}_A) - \boldsymbol{w}(\boldsymbol{\theta}_0)\right)\beta. \tag{12}$$

We describe (12) as a discrepancy between the MIDAS aggregations. Clearly, if $\boldsymbol{\theta}_A = \boldsymbol{\theta}_0$ or $\beta = 0$ then (12) evaluates to zero, and (11) simplifies to $\Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_0) = \Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_A)$. We exploit this analytical separability in the derivations below.

For either given vector for $\boldsymbol{\theta}$, the values of β and δ are defined by,

$$E\left[\boldsymbol{W}(\boldsymbol{\theta}_0)'\Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_0)\right] = \mathbf{0}, \text{ for } \beta = \beta(\boldsymbol{\theta}_0) \text{ and } \delta = \delta(\boldsymbol{\theta}_0) \text{ and} \tag{13}$$

$$E\left[\boldsymbol{W}(\boldsymbol{\theta}_A)'\Delta\boldsymbol{\nu}(\beta,\delta;\boldsymbol{\theta}_A)\right] = \mathbf{0}, \text{ for } \beta = \beta(\boldsymbol{\theta}_A) \text{ and } \delta = \delta(\boldsymbol{\theta}_A). \tag{14}$$

Here, differences between $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_0$ can be analytically tracked, for $\beta(\boldsymbol{\theta}_0) \neq \beta(\boldsymbol{\theta}_A)$ and $\delta(\boldsymbol{\theta}_0) \neq \delta(\boldsymbol{\theta}_A)$, so the discrepancy between moments comes from

$$E\left[\boldsymbol{W}(\boldsymbol{\theta}_0)'\Delta\boldsymbol{\nu}(\beta(\boldsymbol{\theta}_A),\delta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_0)\right] \neq E\left[\boldsymbol{W}(\boldsymbol{\theta}_A)'\Delta\boldsymbol{\nu}(\beta(\boldsymbol{\theta}_A),\delta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A)\right]. \tag{15}$$

Substituting (11) into the left-hand side of (15), using (14), and rearranging terms, illustrates two separable moments:

$$E\left[\left(\boldsymbol{W}(\boldsymbol{\theta}_0)' - \boldsymbol{W}(\boldsymbol{\theta}_A)'\right)\Delta\boldsymbol{\nu}(\beta(\boldsymbol{\theta}_A),\delta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A)\right] + E\left[\boldsymbol{W}(\boldsymbol{\theta}_0)'\Delta\boldsymbol{\eta}(\beta(\boldsymbol{\theta}_A),\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right] \neq \mathbf{0}. \tag{16}$$

The first moment suggests that information on the MIDAS parameter can accrue through the instruments despite the Davies problem. Indeed, if the regression coefficient is zero then $\boldsymbol{\theta}$ is not present in the $\Delta\boldsymbol{\nu}(\beta(\boldsymbol{\theta}_A),\delta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A)$ term nor in the $\Delta\boldsymbol{\eta}(\beta(\boldsymbol{\theta}_A),\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)$ term, but is present in $\boldsymbol{W}(\boldsymbol{\theta}_0)$ and $\boldsymbol{W}(\boldsymbol{\theta}_A)$. The second moment can be shown to hold information on $\boldsymbol{\theta}$ as long as the regression coefficient is not zero, which is formalized in the following Theorem.

[3] Refer to (A.10) in the Appendix for definition of $\boldsymbol{W}(\boldsymbol{\theta})$ constructed for the Arellano-Bond estimator.

Theorem 1 (MIDAS Moment Condition Under the Alternative) *Under the alternative of $\boldsymbol{\theta}_A \neq \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ does not correspond to equal weights, the MIDAS weights for $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_0$ must differ for at least one weight, formally $w_l(\boldsymbol{\theta}_A) \neq w_l(\boldsymbol{\theta}_0)$ for $l \in (1, \ldots, m)$, then*

$$w_j(\boldsymbol{\theta}_A)\left(w_l(\boldsymbol{\theta}_A) - w_l(\boldsymbol{\theta}_0)\right) \neq 0 \text{ or } w_j(\boldsymbol{\theta}_0)\left(w_l(\boldsymbol{\theta}_A) - w_l(\boldsymbol{\theta}_0)\right) \neq 0,$$

for at least one j and l combination. Thus unless $\beta(\boldsymbol{\theta}_A) = 0$, (12) entails

$$E\left[\boldsymbol{W}(\boldsymbol{\theta}_0)' \Delta\boldsymbol{\eta}(\beta(\boldsymbol{\theta}_A), \boldsymbol{\theta}_A, \boldsymbol{\theta}_0)\right] \neq \mathbf{0}. \tag{17}$$

In the Appendix, we develop the moment considered in this Theorem for the specific process we examine in our simulation study. Overall, this analysis implies that a given $\boldsymbol{\theta}_0$ can be tested using the GMM overidentifying restrictions test that specifies $\boldsymbol{\theta}_0$. The power of a test for a point that is not the true value is obviously affected by the Davies problem, yet when the model is correctly specified for the given $\boldsymbol{\theta}_0$, the test will reject with level-correct asymptotic probability.

3.2 Identification-robust inference

Our discussion so far underscores an inevitable identification difficulty resulting from the Davies problem in the considered MIDAS panel. In consequence, we consider a confidence set for $\boldsymbol{\theta}$ obtained by inverting a test that has adequate (asymptotic) level. The idea behind our choice of test, as illustrated in the previous section, is to assess model fit when $\boldsymbol{\theta}$ is specified.

Using the notation of Assumption 1, the confidence set in question can be defined as

$$CS(\boldsymbol{\theta}; \alpha) = \left\{\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}; p_{\mathcal{J}}(\boldsymbol{\theta}_0) > \alpha\right\}. \tag{18}$$

Moving from the joint confidence region to individual confidence sets for the components of $\boldsymbol{\theta}$ is achieved by projecting this region, *i.e.* by computing, in turn, the smallest and largest values for each parameter included in this region. Conceptually, a grid search, particle swarm, or simulated annealing could be implemented over a meaningful set of values for $\boldsymbol{\theta}$. If the generated confidence region is empty, the model can be considered rejected at the considered test level. We formalize this result as follows.

Theorem 2 *In the context of model (1)-(3) and Assumption 1, consider the confidence set defined by (18) that inverts the statistic $\mathcal{J}(\boldsymbol{\theta}_0)$ associated with H_{02}, at the α level. Then*

$$\sup_{\boldsymbol{\theta}_0 \in \Theta} p_{\mathcal{J}}(\boldsymbol{\theta}) \leq \alpha \Leftrightarrow CS(\boldsymbol{\theta}; \alpha) = \emptyset.$$

The intuition is that no value of $\boldsymbol{\theta}$ - within the considered family - will aggregate the higher frequency data into a model that passes the specification test at the considered level. This provides a built-in specification check for the fit of the MIDAS structure within the considered panel set-up.

Now consider hypotheses of form H_{01}. To test such hypotheses allowing for possible weak-identification of $\boldsymbol{\theta}$, and identification failure when $\beta_0 = 0$, we propose two different although related tests defined by maximizing the p-value of the considered significance test over $\boldsymbol{\theta}$. The following Theorem, again framed broadly in the context of Assumption 1, validates our reliance on these tests.

Theorem 3 *In the context of model (1)-(3) and using statistics as in Assumption 1, consider the maximized p-values*

$$\hat{p}_{\mathsf{tA}}(\beta_0) = \sup_{\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}} p_{\mathsf{t}}(\beta_0, \boldsymbol{\theta}_0) \text{ and } \hat{p}_{\mathsf{tC}}(\beta_0) = \sup_{\boldsymbol{\theta}_0 \in CS(\boldsymbol{\theta};\alpha_1)} p_{\mathsf{t}}(\beta_0, \boldsymbol{\theta}_0).$$

Then for inference on H_{01} as in (7) the test with critical regions $\hat{p}_{\mathsf{tA}}(\beta_0) \leq \alpha$ is asymptotically valid at the α level, whereas the test with critical region $\hat{p}_{\mathsf{tC}}(\beta_0) \leq \alpha_2$ is asymptotically valid with level not exceeding $\alpha_1 + \alpha_2$.

If $CS(\boldsymbol{\theta};\alpha_1)$ is informative, the test based on $\hat{p}_{\mathsf{tC}}(\beta_0)$ may outperform its counterpart despite the level correction. It is thus useful even if $\boldsymbol{\theta}$ is treated as a nuisance parameter to consider the confidence set approach.

Pursuing the above GMM example is useful to concretize our approach. While other specification checks may be used, H_{02} can be assessed using the usual over-identification test, which we denote for further reference as the modified-Sargan test. This test concretely assesses whether given $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ there exist values of β and δ [which we denote $\beta = \beta(\boldsymbol{\theta}_0)$ and $\delta = \delta(\boldsymbol{\theta}_0)$] such that $E\left[\boldsymbol{W}(\boldsymbol{\theta}_0)'\boldsymbol{\xi}(\beta(\boldsymbol{\theta}_0), \delta(\boldsymbol{\theta}_0); \boldsymbol{\theta}_0)\right] = 0$. In line with our general set-up, denote this statistic as $\mathcal{J}(\boldsymbol{\theta}_0)$. Then under H_{02} and correct specification

$$\mathcal{J}(\boldsymbol{\theta}_0) = (\Delta\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}_0))'\boldsymbol{W}(\boldsymbol{\theta}_0)[\tilde{\boldsymbol{V}}_N(\boldsymbol{\theta}_0)]^{-1}\boldsymbol{W}(\boldsymbol{\theta}_0)'(\Delta\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}_0)) \underset{N\to\infty}{\sim} \chi^2(\tau - K - 1), \tag{19}$$

where the 'tilde' indicates an estimate from the GMM second step, τ is the number of instruments in $\boldsymbol{W}(\boldsymbol{\theta}_0)$, $K = 2$ (counts the exogenous regressor and the lagged endogenous variable), the weighting matrix is defined as

$$\tilde{\boldsymbol{V}}_N(\boldsymbol{\theta}_0) = \sum_{i=1}^{N} \boldsymbol{W}_i'(\boldsymbol{\theta}_0)(\Delta\tilde{\boldsymbol{\nu}}_{i,\cdot}(\boldsymbol{\theta}_0))(\Delta\tilde{\boldsymbol{\nu}}_{i,\cdot}(\boldsymbol{\theta}_0))'\boldsymbol{W}_i(\boldsymbol{\theta}_0), \tag{20}$$

and $\boldsymbol{W}_i(\boldsymbol{\theta}_0)$ represent the $(T-2)\times\tau$ submatrix of instruments for the i^{th} cross-section.

Our proposed confidence set which inverts this test can thus be expressed as

$$CS(\boldsymbol{\theta};\alpha) = \{\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}; \mathcal{J}(\boldsymbol{\theta}_0) < \chi^2_\alpha(\tau - K - 1)\}, \tag{21}$$

where $\chi^2_\alpha(\tau - K - 1)$ is the critical point associated with (19). An alternative set can be derived using *e.g.* tests for missing dynamics, as considered in our empirical analysis. Now under the same conditions that validate (19), the standard Wald-type statistic is also valid for testing linear hypotheses on β if $\boldsymbol{\theta}$ was given. The above-defined test for unknown $\boldsymbol{\theta}$ can thus be defined using the following statistics:

$$\mathsf{t}_A(\beta_0) = \inf_{\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}} \mathsf{t}(\beta_0; \boldsymbol{\theta}_0), \tag{22}$$

$$\mathsf{t}_C(\beta_0;\alpha_1) = \inf_{\boldsymbol{\theta}_0 \in CS(\boldsymbol{\theta};\alpha_1)} \mathsf{t}(\beta_0; \boldsymbol{\theta}_0), \tag{23}$$

where $\mathsf{t}(\beta_0;\boldsymbol{\theta}_0)$ is the two-tailed (that is absolute value) t-statistic fixing $\boldsymbol{\theta}$ to $\boldsymbol{\theta}_0$, $\boldsymbol{\Theta}$ is the parameter space and $CS(\boldsymbol{\theta};\alpha_1)$ is the above defined confidence set for $\boldsymbol{\theta}$ with level α_1.

4 Simulation Study

The simulations are drawn based on model (1)-(3). The design we consider for the exogenous regressors is as follows,

$$\boldsymbol{X} = \boldsymbol{\psi} \otimes \boldsymbol{\iota}'_m + \rho \mathsf{L_t} \boldsymbol{X} + \boldsymbol{E}, \tag{24}$$

where $\boldsymbol{E}$ is a $N(T-2) \times m$ matrix of $\epsilon_{i,t,j} \sim N(0, \sigma^2_\epsilon)$, drawn independently from μ_i and $\nu_{i,t}$. The regressor DGP matches the high-frequency covariance structure, (D.1) and (D.2), of the special case described in Appendix D. This allows for a direct comparison between the simulation results and the derivations. The MIDAS regressor generated from (24) gives

$$\boldsymbol{x}(\boldsymbol{\theta}) = \boldsymbol{X}\boldsymbol{w}(\boldsymbol{\theta}) = \boldsymbol{\psi} \otimes \boldsymbol{\iota}'_m \boldsymbol{w}(\boldsymbol{\theta}) + \rho \mathsf{L_t} \boldsymbol{X}\boldsymbol{w}(\boldsymbol{\theta}) + \boldsymbol{E}\boldsymbol{w}(\boldsymbol{\theta}),$$

and the first difference (in t) thus yields:

$$\Delta \boldsymbol{x}(\boldsymbol{\theta}) = \rho \Delta \mathsf{L_t} \boldsymbol{X}\boldsymbol{w}(\boldsymbol{\theta}) + \Delta \boldsymbol{E}\boldsymbol{w}(\boldsymbol{\theta}). \tag{25}$$

This design preserves the high frequency autoregressive structure for the MIDAS aggregate, and allows us to remain (in t) close to the simulation designs adopted in the literature on the Arellano-Bond estimator.

With the exception of $\boldsymbol{\theta}$, the null model parameters are taken from the Arellano and Bond (1991) design, which provides a degree of comparability. So the null model parameters[4] are: $\delta = 0.5$, $\beta = 1.0$, $\psi_i = 0$, and $\rho = 0.8$. The case of $\psi_i = 0$ assumes that there are no fixed effects in the MIDAS regressor. The simulations draw from the standard normal distribution for each of the error terms, so $\mu_i \sim N(0, \sigma^2_\mu)$, $\nu_{i,t} \sim N(0, \sigma^2_\nu)$ and $\epsilon_{i,t,j} \sim N(0, \sigma^2_\epsilon)$, with variances set to: $\sigma^2_\mu = 1$, $\sigma^2_\nu = 1$, and $\sigma^2_\epsilon = 0.9$. Bun and Windmeijer (2010) show that the variance ratio $(\sigma^2_\mu / \sigma^2_\nu)$ affects the consistency of the GMM estimates, however we maintain the Arellano and Bond (1991) design to ensure comparability. As discussed in Section 3.1, our approach is robust to alternative estimators that do not suffer from the variance ratio problem. To avoid the instrument proliferation problem, we apply the correction of Roodman (2009), which reduces the number of instruments so $\tau = 2T - 1$.

This study uses a two-parameter exponential Almon weights, $\boldsymbol{\theta} = (\theta_1, \theta_2)'$, with five different assumptions.

- $\boldsymbol{\theta}^A = (0, 0)'$ - flat weights or arithmetic average.
- $\boldsymbol{\theta}^B = (0.1, -0.2)'$ - rapid decay with more weight on recent observations.
- $\boldsymbol{\theta}^C = (0.03, -0.02)'$ - slow decay with relatively more weight on recent observations.
- $\boldsymbol{\theta}^D = (-0.06, 0.01)'$ - slow increase with relatively more weight on older observations.
- $\boldsymbol{\theta}^E = (-0.04, 0.02)'$ - rapid increase with relatively more weight on older observations.

The parameter space, $\Theta = \{(|\theta_1| \leq 1), (|\theta_2| \leq 1)\}$, encompasses nearly all possible weighing schemes of a two-parameter exponential Almon lag. The extreme cases may be of importance in

[4] In their paper, alternative values of δ were examined, 0.2 and 0.8.

economic contexts, like government-issued cheques (direct deposit set for the first of the month), stock market opening and midday effects, and rental housing payment due the first of the month. The number of high frequency observations is chosen to reasonably approximate high frequency data, where $m = 20$ unless specified in the table.

Results in Tables 1-4 report the size and power of our proposed test for $H_{02} : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. Since this test is designed for inversion purposes, in addition to the natural special equal-weights case, that is $\boldsymbol{\theta}_0 = \mathbf{0}$, we consider various choices for $\boldsymbol{\theta}_0$. Table 1 shows that our procedure exhibits size control for all considered null values. Outside of the equal weight setting, the results could be considered modestly undersized, which is a consequence of implementing the Sargan statistic as the specification test. The Sargan is an asymptotic method, and, as such, can result in size distortions in finite samples.

It is important to recall that estimation uncertainty on $\boldsymbol{\theta}$ is inherent to the model, because of its definitional identification problem. This further explains why power results differ importantly with $\boldsymbol{\theta}_0$, as illustrated in Section (3.1). Also, it is broadly known in the general GMM literature, the information content of the Sargan statistics depends importantly on the sample size. The important question is whether allowing for non-equal weights eventually pays-off for inference on β acknowledging that information on $\boldsymbol{\theta}$ may be imprecise. Interestingly, our results support this conclusion, with reasonable sample sizes. In summary, our size study has shown that no spurious inference would result from allowing uncertainty around $\boldsymbol{\theta}$.

In Table 2, the tested hypothesis is $\boldsymbol{\theta}_0 = \mathbf{0}$; the DGP is drawn using various choices for $\boldsymbol{\theta}$ reported in the first and second column. Power is consistent with conventional wisdom as it increases in both N and T. On balance, this table shows that our Sargan statistic can reliably disentangle equal weights from a MIDAS alternative as N increases, except for small T. Recall that information on $\boldsymbol{\theta}$ more or less accrues via the time series dimension of the model. Our findings with $T = 10$ are thus noteworthy. These results should be assessed given our findings on testing β [reported below]. This table also demonstrates the power improvements as the weights become more extreme, so Case 2 applies. The model is able to discriminate against equal weights because $\mathcal{A}_{t,r}$ is non-zero and $\boldsymbol{w}'(\boldsymbol{\theta}_A)\boldsymbol{w}(\boldsymbol{\theta}_A)$ will approach one as weighting schemes become more extreme.

In Table 3 and 4, we study power when the tested null hypothesis is one or the other boundary of the Almon function. We aim to assess whether the Sargan test can detect extreme opposite weighting schemes. Results line up with our analysis for the equal weights null: a small T would not suffice for the test to be informative, yet good power can be achieved as T approaches 15. Power also changes asymmetrically as $\boldsymbol{\theta}$ deviates from the null. For a pair of parameter values, it is possible to contrast the power results when the null is equal weights tested against alternative values of $\boldsymbol{\theta}$, with results corresponding to the opposite design where the null is not equal weights and the alternative is equal weights. The natural null hypothesis to be considered here is that of equal weights, and our results confirm good performance in this case the converse.

Power responds more to deviations of θ_2, which is a consequence of the "j^2" part of the Almon function. More details are provided in the Appendix tables confirming the asymmetric power be-

Table 1: Rejection Frequency under the Null

Weight	DGP & Null		$N = 500$			$N = 1000$		
	θ_1	θ_2	$T = 5$	$T = 10$	$T = 15$	$T = 5$	$T = 10$	$T = 15$
older	-0.04	0.02	0.036	0.036	0.039	0.035	0.037	0.042
$\Uparrow$	-0.06	0.01	0.042	0.034	0.044	0.035	0.046	0.035
equal	0	0	0.041	0.051	0.041	0.052	0.042	0.043
$\Downarrow$	0.03	-0.02	0.042	0.049	0.040	0.046	0.053	0.050
recent	0.1	-0.2	0.028	0.051	0.049	0.020	0.046	0.042

$m = 20$, which coincides roughly with the number of trading days in a month.

Table 2: Rejection Frequency with $\boldsymbol{\theta}_0 = (0, 0)'$

Weight	DGP		$N = 500$			$N = 1000$		
	θ_1	θ_2	$T = 5$	$T = 10$	$T = 15$	$T = 5$	$T = 10$	$T = 15$
older	-0.04	0.02	0.060	0.242	0.472	0.275	0.680	0.968
$\Uparrow$	-0.06	0.01	0.043	0.077	0.091	0.077	0.109	0.225
equal	0	0						
$\Downarrow$	0.03	-0.02	0.044	0.078	0.082	0.062	0.073	0.127
recent	0.1	-0.2	0.068	0.655	0.866	0.161	0.597	0.944

$m = 20$, which coincides roughly with the number of trading days in a month.

tween θ_1 and θ_2. There is no reason to expect symmetry in power between each parameter, since the confidence set we propose is not a symmetric interval. Globally, these results confirm the difficulty of identifying $\boldsymbol{\theta}$ and the time-series roots of the MIDAS problem: the more time series data is available the more information on $\boldsymbol{\theta}$ will accrue yet perfect identification is not to be expected. Configurations that are more skewed than the true DGP model can be tested more reliably, and values of $\boldsymbol{\theta}$ that induce an important effect on the aggregated variable can be disentangled with sufficient data. Consequently, the concrete question that matters is whether possibly weak information on $\boldsymbol{\theta}$ has significant costs regarding inference on β. We thus examine this question in what follows.

The autoregressive structure can also deviate from our Special Case, and we examine this effect by implementing our approach with a high-frequency autoregressive (hfAR) model of order one, refered to as hfAR(1). The autoregressive relationship is between j and $j - 1$, which is discussed in Appendix A in regards to the model and implications of aggregation. The simulation results, shown in Table E6, do not change our findings or conclusions.

We thus now turn our attention to inference on β. Tables 5 and 6 report two sets of results: (i) tests which assume that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ we refer to as the "oracle" method (do not minimize over $\boldsymbol{\Theta}$);

Table 3: Rejection Frequency with $\boldsymbol{\theta}_0 = (1, 1)'$, $w_{20}(\boldsymbol{\theta}_0) = 1$

Weight	DGP		$N = 500$			$N = 1000$		
	θ_1	θ_2	$T = 5$	$T = 10$	$T = 15$	$T = 5$	$T = 10$	$T = 15$
older	-0.04	0.02	0.046	0.051	0.063	0.030	0.043	0.060
$\Uparrow$	-0.06	0.01	0.059	0.077	0.101	0.057	0.121	0.168
equal	0	0	0.06	0.058	0.06	0.047	0.038	0.092
$\Downarrow$	0.03	-0.02	0.049	0.067	0.091	0.067	0.096	0.247
recent	0.1	-0.2	0.065	0.144	0.259	0.122	0.361	0.755

$m = 20$, which coincides roughly with the number of trading days in a month.

Table 4: Rejection Frequency with $\boldsymbol{\theta}_0 = (-1,-1)'$, $w_1(\boldsymbol{\theta}_0) = 1$

Weight	DGP		$N = 500$			$N = 1000$		
	θ_1	θ_2	$T = 5$	$T = 10$	$T = 15$	$T = 5$	$T = 10$	$T = 15$
older	-0.04	0.02	0.062	0.192	0.738	0.031	0.510	0.987
$\Uparrow$	-0.06	0.01	0.070	0.06	0.088	0.038	0.117	0.235
equal	0	0	0.054	0.055	0.048	0.034	0.054	0.086
$\Downarrow$	0.03	-0.02	0.055	0.049	0.045	0.05	0.083	0.126
recent	0.1	-0.2	0.061	0.038	0.028	0.051	0.049	0.060

$m = 20$, which coincides roughly with the number of trading days in a month.

Table 5: Rejection Frequency for $\beta_0 = 0^{\dagger}$

Weight	$\boldsymbol{\theta}$		$\mathrm{t}(\beta_0)$ at 2.5% level						
	θ_1	θ_2	$\beta = -2$	$\beta = -1$	$\beta = -0.5$	$\beta = 0$	$\beta = 0.5$	$\beta = 1$	$\beta = 2$
older	0	0.1	0.999	0.996	0.956	0.009	0.954	0.997	1
$\Uparrow$	0	0.05	0.998	0.993	0.938	0.010	0.929	0.990	1
equal	0	0	0.513	0.523	0.559	**0.006**	0.551	0.517	0.530
$\Downarrow$	0	-0.05	1	1	0.989	0.005	0.990	1	1
recent	0	-0.1	1	1	0.989	0.005	0.991	1	1
			$\inf \mathrm{t}(\beta_0)$ at 2.5% level						
	$CS(\boldsymbol{\theta}; 2.5\%)$		0.461	0.391	0.120	**0.010**	0.114	0.382	0.444

†The simulation settings are based on more high frequency observations, $m = 40$, and the panel dimension is relatively low at $T = 5$ and $N = 500$. The final row takes into account the confidence set for $\boldsymbol{\theta}$ at a 2.5% level. The DGP is $\boldsymbol{\theta} = \mathbf{0}$ and $\beta = 0$.

(ii) the "consistent-set" counterpart, that is, the test based on $\mathtt{t}_C(\beta_0) = \inf_{\boldsymbol{\theta} \in CS(\boldsymbol{\theta};\alpha_1)} \mathtt{t}(\beta_0; \boldsymbol{\theta})$. The above described Sargan test is inverted to construct $CS(\boldsymbol{\theta}; \alpha_1)$ and we set $\alpha_1 = \alpha_2 = 2.5\%$. This experiment is implemented with $T = 5$ which is the least favorable scenario on the usefulness of the pre-test confidence set. When the DGP matches the null, the method is able to control the 2..5% level of the t-test but the small sample leads to undersized rejection frequencies when $\beta = 0$ and marginally oversized when $\beta = 2$.

Comparing Table 5 to Table 6 allows us to analyze the Davies problem. In the former, the null we test corresponds to $\beta_0 = 0$ and in the latter to $\beta_0 = 2$. The power of the oracle test varies with $\boldsymbol{\theta}$, which may also be interpreted along the following important dimension: spurious inference on β may result if we misspecify $\boldsymbol{\theta}$. In contrast, we see that the bound test almost matches the

Table 6: Rejection Frequency for $\beta_0 = 2^{\dagger}$

Weight	$\boldsymbol{\theta}$		$\mathrm{t}(\beta_0)$ at 2.5% level							
	θ_1	θ_2	$\beta = -1$	$\beta = -0.5$	$\beta = 0$	$\beta = 0.5$	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$
older	0	0.1	1	1	1	0.999	0.997	0.019	0.991	1
$\Uparrow$	0	0.05	0.998	1	1	0.997	0.988	**0.029**	0.988	1
equal	0	0	0.949	0.987	0.988	0.924	0.523	0.094	0.076	0.135
$\Downarrow$	0	-0.05	1	1	1	1	1	0.018	0.999	1
recent	0	-0.1	1	1	1	1	1	0.003	1	1
			$\inf \mathrm{t}(\beta_0)$ at 2.5% level							
	$CS(\boldsymbol{\theta}; 2.5\%)$		0.950	0.987	0.988	0.925	0.595	**0.010**	0.423	0.496

†The simulation settings are the same as Table 5, except the DGP is $\boldsymbol{\theta} = (0, 0.05)'$ and $\beta = 2$.

oracle power when $\beta_0 = 2$, and the cost of accounting for an unknown $\boldsymbol{\theta}$ is not large when $\beta_0 = 0$. This is the worst-case scenario since $T = 5$, so we should except some power discrepancy from the (infeasible) oracle test. It is important to highlight that power is symmetric in the case of $\beta_0 = 0$ and asymmetric when $\beta_0 = 2$, since the former does not depend on $\boldsymbol{\theta}$. This shows that the method holds concrete promise particularly as we weigh-in the misspecification test of falsely fixing $\boldsymbol{\theta}$. It should be noted that the oracle test is conservative

Overall, results confirm the usefulness of testable information on the equal weights and zero β null hypothesis, which would concretely produce unbounded confidence regions. The latter, in addition to empty outcomes reveal misspecification that should be flagged empirically.

5 Empirical Analysis

We demonstrate the use of our methodology with an empirical application. The example that we focus on is a study by Obstfeld, Shambaugh, and Taylor (2010) which presents a model to explain the recent rapid accumulation of international reserve holdings (notably by emerging economies). The authors argue that nowadays an important part of reserve holdings is a hedge against international financial shocks since the latter can cause important damage to the domestic financial sector and the country's currency. They support their views based on the outcomes of a panel OLS regression with annual data and error clustering. In particular, they find that the explanatory variables representing financial openness and depth are both significant, important, and they have the correct signs. On the other hand, only a few of the traditional variables are found to be significant.

The exchange rate volatility, defined as the standard deviation of equally-weighted monthly changes over the year, is among these traditional regressors. We investigate whether the aggregation manner of these monthly exchange rate observations affects the obtained results. Our examinations are conducted in the context of the general specification dubbed 'Horserace' in Tables 1 and 2 of Obstfeld et al. (2010), and which includes regressors suggested by traditional theories, as well as variables representing financial openness and financial depth.

We collect monthly data on the exchange rate for as many countries in the original study as possible.[5] This monthly data is used in conjunction with the original annual data on the remaining regressors of the model. We consider the OLS specification from Obstfeld et al. (2010) utilizing the average exchange rate volatility, and another specification that allows for MIDAS weights.[6] Specifically for MIDAS weights, we minimize the clustered-standard-error-corrected t-statistic over the full MIDAS parameter space, as in (22). The results are presented in the last two columns of Table 7 (for all countries), and Table 8 (for the emerging country subset).

From Table 7, we see that estimates and associated p-values are largely similar across the two OLS specifications except in the case of the exchange rate volatility; this coefficient is highly signifi-

[5] As these were not available for all of the countries considered in the original study, our observations are fewer than those reported in their Tables 1 and 2.

[6] For ease of exposition, we omit the dummy terms present in the original model specifications.

cant with equal-weight aggregation (p-value of zero) but insignificant with MIDAS weights (p-value of 1). The same outcome is observed when we consider only the subset of emerging countries in our sample. Table 8 shows that the significance of the exchange rate volatility coefficient is reversed when MIDAS weights are used instead of equal weights. Clearly, the aggregation scheme matters regarding the importance of the role of exchange rate movements.

Table 7: Estimates with Alternative Aggregation Schemes; All Countries

	Arellano-Bond			OLS	
Specifiation	Equal Weights	MIDAS	UMIDAS	Equal Weights	MIDAS
Lag ln(Reserves/GDP)	0.771	0.773	0.769		
	(.000)	(.000)	(.000)		
ln(Population)	-0.009	0.000	0.020	-0.037	-0.036
	(.011)	(.992)	(.000)	(.344)	(.356)
ln(trade/GDP)	0.055	0.027	–	0.683	0.677
	(.000)	(.000)		(.000)	(.000)
ln(GDP/person)	-0.059	0.001	-0.114	-0.152	-0.151
	(.000)	(.167)	(.000)	(.008)	(.007)
Exchange Rate Volatility	0.000	0.000	Included but	-0.000	0.000
	(.000)	(.988)	not reported	(.000)	(1.000)
Financial Openness	0.008	-0.000	-0.242	0.301	0.296
	(.272)	(.982)	(.000)	(.227)	(.230)
Financial Depth	0.185	-0.000	0.058	0.316	0.314
	(.000)	(.814)	(.000)	(.000)	(.000)
Observations	1588	1588	1588	1633	1633
Missing Dynamics p-value	0.679	0.732	0.669	0.000	0.000
# instruments	170	170	445		
Sargan p-value	0.000	0.000	0.311		
Sargan df	162	162	426		

Note 1: Values in parentheses are p-values. In addition to the lagged term, the Arellano-Bond GMM instruments include current and lags of the covariates. Estimates and p-values under the MIDAS heading correspond to inf t-statistics as follows. For the Arellano-Bond case, the GMM-t statistic is minimized over a confidence set for the MIDAS parameter, as in (21). The underlying confidence set inverts the GMM Sargan statistic at the 2.5% level. For the OLS case, the clustered standard error corrected t-statistic is minimized over the full MIDAS parameter space as in (20). The missing dynamics p-value corresponds to the sup p-value over the considered MIDAS confidence set in GMM, and the full parameter space for OLS.
Note 2: Results are from an unbalanced panel of 120 countries with annual data ranging from 3 to 25 years.

The original OLS setting employs clustered standard errors to account for serial correlation (Bertrand, Duflo, and Mullainathan (2004)). This approach is valid, however if the model is genuinely dynamic (lagged dependent) then clustering can be insufficient. In the context of this analysis, some countries adjust their reserve holdings based on announced criteria or rule-based programs, both of which suggest that these countries may have some form of target for their reserve holdings. This may explain why current reserve holdings react to genuine dynamics in holdings; see Domanski et al. (2016). This motivates us to consider a genuinely dynamic panel model and implement the Arellano-Bond estimator.

We move to a dynamic specification with the existing regressors and conduct an Arellano-

Bond GMM estimation. The exchange rate volatility variable is included among the regressors in three ways: aggregated with equal weights, aggregated via MIDAS weights, and the UMIDAS specification. Results are presented in the first three columns of Tables 7 and 8. The p-values of the estimates are reported in the top panel and correspond to the GMM-t statistic that is minimized over a confidence set for the MIDAS parameter, as in (23). In the lower panel we report the GMM Sargan statistic p-value, and the supremum p-value for MIDAS weights at the 2.5% level.

We observe that the coefficient on the lag of the dependent variable is significant and large in magnitude. Combined with the test for missing dynamics that is insignificant, means that this setting accounts for genuine dynamics or serial correlation. This allows us to return focus on the parameters of interest from the original OLS specification.

When considering all the countries in our sample, the UMIDAS is the only specification not rejected by the Sargan test, whereas with the emerging country subset, only the MIDAS model survives at the 5% level.[7] This implies that the aggregation method matters importantly.

Looking more closely at the results of the UMIDAS in Table 7, we see that while some of the traditional variables still play a significant role in determining reserve holdings, financial depth and financial openness are also relevant for the latter. However, although the former has the expected sign, its estimated importance is found to be much smaller compared to its OLS counterparts. On the other hand, the sign of the financial openness coefficient is opposite to that expected.[8] Interestingly, the OLS-MIDAS specifications yield insignificant outcomes for the estimated coefficient of the financial openness variable.

Turning to the MIDAS results pertaining to the subset of emerging economies (Table 8), we find some differences compared to the all-country case. In particular, while several traditional variables again play a role in the evolution of reserve holdings, we observe that the financial depth variable is no longer significant. One explanation may be that the relative solidity of the financial sector in emerging countries is not sufficiently important to adequately buffer against (important) external financial shocks. In addition, and as in Table 7, while the estimate of the financial openness variable is significant, it has a negative sign. This contrasts with the large values and positive signs obtained for this coefficient in the OLS settings.

Our method also conveys information on the estimated MIDAS aggregation weights. In Figure 1 we provide a graphical representation of least rejected values for the obtained weights; the implications of the rejected values at the 5% level are discussed further below. The figure allows a direct reading of weights for each month of the year, with the graphs representing the $w_j(\boldsymbol{\theta})$ associated with the supremum Sargan p-values, and the equal-weights case (1/12 for each month).

[7] One reason why the UMIDAS is rejected in the latter case may be that the ratio of instruments to observations is large.

[8] Generally, we would expect greater financial openness to be associated with a higher crisis vulnerability and therefore a greater demand for reserves. At the same time, a more flexible exchange rate reduces the demand for reserves given that central banks do not need large reserves to manage their exchange rate. Ultimately, the net effect has to be measured empirically. In addition, if the financial openness variable and the exchange rate volatility variable reflect largely similar information, there could be a transfer of coefficient signs; the exchange rate volatility coefficients are not reported to save space.

Table 8: Estimates with Alternative Aggregation Schemes; Emerging Countries

	Arellano-Bond			OLS	
Specification	Equal Weights	MIDAS	UMIDAS	Equal Weights	MIDAS
Lag of ln(Reserves/GDP)	0.636	0.588	0.860		
	(.000)	(.000)	(.000)		
ln(Population)	-0.127	0.000	0.494	-0.030	-0.022
	(.780)	(1.00)	(.660)	(.634)	(.724)
ln(trade/GDP)	0.216	0.072	-0.394	0.439	0.437
	(.410)	(.688)	(0.603)	(.004)	(.002)
ln(GDP/person)	0.230	0.277	0.261	0.036	0.000
	(.002)	(.002)	(.073)	(.691)	(0.719)
Exchange Rate Volatility	0.161	0.130	Included but	0.064	0.000
	(.000)	(.016)	not reported	(.024)	(0.999)
Financial Openness	-0.728	-0.388	-1.052	0.770	0.763
	(.002)	(.021)	(.192)	(.035)	(.030)
Financial Depth	0.306	0.242	-0.022	0.480	0.469
	(.100)	(.175)	(.925)	(.001)	(.000)
Observations	456	456	456	466	466
Missing Dynamics p-value	0.446	0.924	0.532	0.000	0.000
# instruments	170	170	408		
Sargan p-value	0.073	0.202	0.029		
Sargan df	162	162	389		

Note: Refer to Note 1 from Table 7. Results are from an unbalanced panel of 46 countries with annual data ranging from 9 to 25 years.

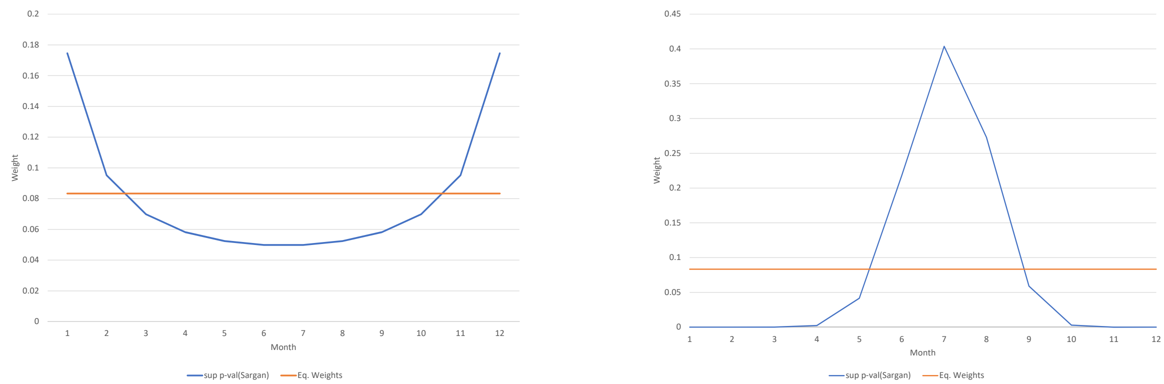
Figure 1(a) corresponds to the results in Table 7 which reports that the supremum Sargan p-value rejects the model for all values of $\boldsymbol{\theta}$ at almost any significance level. We thus see that both the (least-rejected) MIDAS weights, and the average weight scheme, are rejected. Figure 1(b) corresponds to the outcomes reported in Table 8, and where the model is not rejected for a subset of $\boldsymbol{\theta}$. In this case, the graph shows that the information that pertains to the exchange rate volatility variable, and which is relevant for explaining reserve holdings, essentially accrues over the period extending from the end of the second quarter of the year through to its third quarter.

One way to interpret the above results is that the equal weights assumption is supported at the 5% level but rejected at levels above 7.3%. Nevertheless, in view of the sample size, the MIDAS analysis should not be dismissed in favour of a hasty conclusion based on 5% cut-offs. At the 5% level, the modified-Sargan rejects the model when the weights deviate significantly from the point estimate weights. Given that the retained $\boldsymbol{\theta}$ are a subset of the $\boldsymbol{\theta}$ parameter space, this demonstrates that our power results are viable in empirical settings.

Recalling the required level adjustment, we observe that the lag regressor, per capita GDP, exchange rate volatility, and financial openness remain significant at 10% level. Again, in view of the sample size, this cut-off is reasonably cautious. Both average and MIDAS weights models are retained, although the least rejected model deviates significantly from the equal weights case (see Figure 1(b)).

As for the resulting policy implications, given that decision makers in central banks convene

Figure 1: MIDAS weights for supremum Sargan Statistic



(a) Full Sample

(b) Emerging Markets

periodically to make adjustments to their reserve holdings, such information can be useful with regard to the period of the year over which they should take account of exchange rate movements more particularly, as well as for deciding on appropriate timings of central bank interventions.

Taken collectively, our results demonstrate the relevance of considering genuinely dynamic model. Our results also show that different weighing schemes of exchange rate volatility can change drastically a major result.

6 Concluding Remarks

Our proposed method introduces MIDAS regressors into the context of dynamic panel data models. We show that a specification test can be inverted for inference on MIDAS parameters. An empty confidence set indicates a lack of fit of the imposed model. Our theoretical results are paralleled in a simulation study, which demonstrates level control, local power for alternatives with similar weighing schemes, and power against an equal weight assumption.

We apply the proposed methodology to revisit the results of Obstfeld, Shambaugh, and Taylor (2010) that try to explain the rapid accumulation since the nineties of international reserve holdings. An important variable of the model that requires aggregation is exchange rate volatility. Our findings underscore the usefullness of our proposed data driven aggregations scheme.

More broadly, this paper sets a promising template for dealing with shrinkage parameters or methods, in general, when unrestricted estimation is not desirable nor even possible. Specification checks in general hold concrete information on shrinkage parameters, that can be harvested to formally identify intervening parameters.

7 Acknowledgements

The authors would like to thank the editor, two anonymous referees, Stéphane Bonhomme, Valentina Corradi, and Stepana Lazarova. This work was partly supported by the Social Sciences and Humanities Research Council of Canada, and the Fonds FQRSC (Government of Québec).

References

Daron Acemoglu, Simon Johnson, James A Robinson, and Pierre Yared. Income and democracy. *American Economic Review*, 98(3):808–42, 2008.

Philippe Aghion, Philippe Bacchetta, Romain Ranciere, and Kenneth Rogoff. Exchange rate volatility and productivity growth: The role of financial development. *Journal of Monetary Economics*, 56(4):494–513, 2009.

Jean-Pierre Allegret, Cécile Couharde, Dramane Coulibaly, and Valérie Mignon. Current accounts and oil price fluctuations in oil-exporting countries: the role of financial development. *Journal of International Money and Finance*, 47:185–201, 2014.

Shirley Almon. The distributed lag between capital appropriations and expenditures. *Econometrica*, 33(1):178–196, 1965.

Theodore Wilbur Anderson and Cheng Hsiao. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1):47–82, 1982.

Donald WK Andrews and Xu Cheng. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80(5):2153–2211, 2012.

Manuel Arellano and Stephen Bond. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297, 1991.

Jennie Bai, Eric Ghysels, and Jonathan H Wright. State space models and MIDAS regressions. *Econometric Reviews*, 32(7):779–813, 2013.

Thorsten Beck and Ross Levine. Stock markets, banks, and growth: Panel evidence. *Journal of Banking & Finance*, 28(3):423–442, 2004.

Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275, 2004.

Michael Binder and Melanie Krause. When is inequality detrimental to growth? - perspectives from a novel cross-country approach. *Unpublished results*, 2014.

Michael Bleaney and David Greenaway. The impact of terms of trade and real exchange rate volatility on investment and growth in sub-saharan africa. *Journal of Development Economics*, 65(2):491–500, 2001.

Maurice JG Bun and Frank Windmeijer. The weak instrument problem of the system GMM estimator in dynamic panel data models. *The Econometrics Journal*, 13(1):95–126, 2010.

Michael P Clements and Ana Beatriz Galvão. Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the united states. *Journal of Business & Economic Statistics*, 26(4):546–554, 2008.

Michael P Clements and Ana Beatriz Galvão. Forecasting us output growth using leading indicators: An appraisal using MIDAS models. *Journal of Applied Econometrics*, 24(7):1187–1206, 2009.

Robert B Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2):247–254, 1977.

Robert B Davies. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1):33–43, 1987.

D. Domanski, E. Kohlscheen, and R. Moreno. foreign exchange market intervention in emes: What has changed? *BIS Quartely Review*, pages 65–79, 2016.

Jean-Marie Dufour. Nonlinear hypotheses, inequality restrictions, and non-nested hypotheses: Exact simultaneous tests in linear regressions. *Econometrica*, pages 335–355, 1989.

Jean-Marie Dufour. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65(6):1365–1387, 1997.

Claudia Foroni and Massimiliano Giuseppe Marcellino. A survey of econometric methods for mixed-frequency data. *Available at SSRN 2268912*, 2013.

Claudia Foroni, Massimiliano Marcellino, and Christian Schumacher. Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2013.

Eric Ghysels. Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, 193(2):294–314, 2016.

Eric Ghysels and Jonathan H Wright. Midas instruments. Technical report, Working Paper, 2010.

Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov. The MIDAS touch: Mixed data sampling regression models. *Finance*, 2004.

Eric Ghysels, Pedro Santa-Clara, and Rossen Valkanov. Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1):59–95, 2006.

Eric Ghysels, Arthur Sinko, and Rossen Valkanov. Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90, 2007.

Thomas B Götz, Alain Hecq, and Stephan Smeekes. Testing for granger causality in large mixed-frequency vars. *Journal of Econometrics*, 2016.

Pierre Guérin and Massimiliano Marcellino. Markov-switching MIDAS models. *Journal of Business & Economic Statistics*, 31(1):45–56, 2013.

Maurizio Michael Habib, Elitza Mileva, and Livio Stracca. The real exchange rate and economic growth: Revisiting the case using external instruments. *Journal of International Money and Finance*, 73:386–398, 2017.

Frank Kleibergen. Testing parameters in GMM without assuming that they are identified. *Econometrica*, 73(4):1103–1123, 2005.

Vladimir Kuzin, Massimiliano Marcellino, and Christian Schumacher. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27(2):529–542, 2011.

Virmantas Kvedaras and Vaidotas Zemlys. Testing the functional constraints on parameters in regressions with variables of different frequency. *Economics Letters*, 116(2):250–254, 2012.

Siong Hook Law and Nirvikar Singh. Does too much finance harm economic growth? *Journal of Banking & Finance*, 41:36–44, 2014.

Hongbin Li, Hong Ma, and Yuan Xu. How do exchange rate movements affect chinese exports?—a firm-level investigation. *Journal of International Economics*, 97(1):148–161, 2015.

Fernando Eguren Martin. Exchange rate regimes and current account adjustment: An empirical investigation. *Journal of International Money and Finance*, 65:69–93, 2016.

David J McKenzie. Estimation of ar (1) models with unequally spaced pseudo-panels. *The Econometrics Journal*, 4(1):89–108, 2001.

J Isaac Miller. Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *Journal of Financial Econometrics*, 12(3):584–614, 2013.

J Isaac Miller. Simple robust tests for the specification of high-frequency predictors of a low-frequency series. *Econometrics and Statistics*, 5:45–66, 2018.

Daniel L Millimet and Ian K McDonough. Dynamic panel data models with irregular spacing: With an application to early childhood development. *Journal of Applied Econometrics*, 32(4):725–743, 2017.

Maurice Obstfeld, Jay C Shambaugh, and Alan M Taylor. Financial stability, the trilemma, and international reserves. *American Economic Journal: Macroeconomics*, 2(2):57–94, 2010.

David Roodman. How to do xtabond2: An introduction to difference and system gmm in stata. *Stata Journal*, 9(1):86–136, 2009.

Yuya Sasaki and Yi Xin. Unequal spacing in dynamic panel data: Identification and estimation. *Journal of Econometrics*, 196(2):320–330, 2017.

James H Stock and Jonathan H Wright. GMM with weak identification. *Econometrica*, 68(5): 1055–1096, 2000.

A Recursion and Instruments

The model dynamic panel model with MIDAS covariates, defined by (1) – (3), can be written as,

$$y_{i,t} = \delta y_{i,t-1} + x_{i,t}(\boldsymbol{\theta})\beta + v_{i,t} + \mu_i, \tag{A.1}$$

and via recursion to obtain:

$$y_{i,t} = \delta^t y_{i,0} + \left(\sum_{k=0}^{t-1} \delta^k \mathsf{L_t}^k x_{i,t}(\boldsymbol{\theta})\right)\beta + \sum_{k=0}^{t-1} \delta^k \mathsf{L_t}^k \nu_{i,t} + \mu_i \sum_{k=0}^{t-1} \delta^k. \tag{A.2}$$

Assuming the model initiated in the distant past, then the culmination of this past can be represented by the unobserved value, $y_{i,0}$. We assume for all high frequency observations, $j, k = 1, \ldots, m$, the following:

$$\begin{aligned} E[y_{i,0}x_{i,t,j}] &= 0 \ \forall\ t, && \text{(A.3)}\\ E[\nu_{i,t}x_{i,s,j}] &= 0 \ \forall\ t, s \text{ and} && \text{(A.4)}\\ E[\mu_i x_{i,t,j}] &= E[\mu_i x_{i,s,k}] \ \forall\ t, s, && \text{(A.5)} \end{aligned}$$

where the first two assumptions assume strict exogeneity, the latter assumption allows for correlation between the regressor and the unobserved heterogeneity.

The instrument matrix is constructed from y's and $x(\boldsymbol{\theta})$'s in a block diagonal matrix form, where each block is indexed by $s = 3, \ldots, T$, so let

$$\begin{aligned} \boldsymbol{z}_{i,s} := \boldsymbol{z}_{i,s}(y, x; \boldsymbol{\theta}) &= (\boldsymbol{z}_{i,s}(y), \boldsymbol{z}_{i,s}(x; \boldsymbol{\theta})) \text{ where} && \text{(A.6)}\\ \boldsymbol{z}_{i,s}(y) &= (y_{i,1}, \ldots, y_{i,s-2}) \text{ and} && \text{(A.7)}\\ \boldsymbol{z}_{i,s}(x; \boldsymbol{\theta}) &= (x_{i,1}(\boldsymbol{\theta}), x_{i,2}(\boldsymbol{\theta}), \ldots, x_{i,s}(\boldsymbol{\theta})). && \text{(A.8)} \end{aligned}$$

The instruments in each block are functions of the observed data and $\boldsymbol{\theta}$, so we denote the instrument matrix as $\boldsymbol{W}(\boldsymbol{\theta})$. The matrix of instruments for each i is denoted $\boldsymbol{W}_i(\boldsymbol{\theta})$ takes the following form,

$$\boldsymbol{W}_i := \boldsymbol{W}_i(\boldsymbol{\theta}) = (\boldsymbol{W}_i(y),\ \boldsymbol{W}_i(x, \boldsymbol{\theta})) \ \text{ where} \tag{A.9}$$

$$\boldsymbol{W}_i(y) = \begin{bmatrix} \boldsymbol{z}_{i,3}(y) & 0 & 0 & 0 \\ 0 & \boldsymbol{z}_{i,4}(y) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boldsymbol{z}_{i,T}(y) \end{bmatrix} \text{ and}$$

$$\boldsymbol{W}_i(x, \boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{z}_{i,3}(x; \boldsymbol{\theta}) & 0 & 0 & 0 \\ 0 & \boldsymbol{z}_{i,4}(x; \boldsymbol{\theta}) & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boldsymbol{z}_{i,T}(x; \boldsymbol{\theta}) \end{bmatrix}.$$

$\boldsymbol{W}_i$ is a $(T-2) \times \tau$ matrix, where τ is a function of T and the number of exogenous regressors (K). The full matrix of instruments, for all i, is

$$\begin{aligned} \boldsymbol{W}(\boldsymbol{\theta}) &= (\boldsymbol{W}_1'(\boldsymbol{\theta}), \ldots, \boldsymbol{W}_N'(\boldsymbol{\theta}))' = (\boldsymbol{W}(y),\ \boldsymbol{W}(x, \boldsymbol{\theta})) \ \text{ where} && \text{(A.10)}\\ \boldsymbol{W}(\boldsymbol{y}) &= (\boldsymbol{W}_1'(\boldsymbol{y}), \ldots, \boldsymbol{W}_N'(\boldsymbol{y}))' \text{ and} && \text{(A.11)}\\ \boldsymbol{W}(x, \boldsymbol{\theta}) &= (\boldsymbol{W}_1'(x, \boldsymbol{\theta}), \ldots, \boldsymbol{W}_N'(x, \boldsymbol{\theta}))'. && \text{(A.12)} \end{aligned}$$

$\boldsymbol{W}(\boldsymbol{\theta})$ is a matrix with dimensions $N(T-2) \times \tau$. The instrument matrix can be constructed in this manner for any given value of $\boldsymbol{\theta}$, since the MIDAS aggregate is then observable so $\boldsymbol{\theta}_0$ will yield $\boldsymbol{W}(\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}_A$ will yield $\boldsymbol{W}(\boldsymbol{\theta}_A)$.

B Proofs

Proof of Theorem 1. The population moment (17) used to indicate power in terms of $\boldsymbol{\theta}$ can be defined as

$$E\left[\boldsymbol{W}(\boldsymbol{\theta}_0)'\Delta\boldsymbol{\eta}(\beta(\boldsymbol{\theta}_A),\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right] = \begin{bmatrix} \mathcal{A} \\ \mathcal{B} \end{bmatrix} \text{ where} \tag{B.1}$$

$$\mathcal{A} = E\left[\boldsymbol{W}(y)'\Delta\boldsymbol{\eta}(\beta(\boldsymbol{\theta}_A),\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right] \text{ and} \tag{B.2}$$

$$\mathcal{B} = E\left[\boldsymbol{W}(x,\boldsymbol{\theta}_0)'\Delta\boldsymbol{\eta}(\beta(\boldsymbol{\theta}_A),\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right]. \tag{B.3}$$

This exploits the separability of $\boldsymbol{W}(\boldsymbol{\theta})$ into a set of instruments from lags of the dependant variable and a set from the aggregated regressor. From these we can examine each row of $\mathcal{A}$ and $\mathcal{B}$, by formally considering the instruments separately:

$$\mathcal{A}_{t,r} := E\left[\mathsf{L_t}^r\boldsymbol{y}'_{\cdot,t}\Delta\boldsymbol{\eta}(\beta(\boldsymbol{\theta}_A),\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right], \text{ and} \tag{B.4}$$

$$\mathcal{B}_{t,q} := E\left[\mathsf{L_t}^q\boldsymbol{x}'_{\cdot,t}(\boldsymbol{\theta}_0)\Delta\boldsymbol{\eta}(\beta(\boldsymbol{\theta}_A),\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right], \tag{B.5}$$

where $r = 2,\ldots,t-1$ and $q = 0,\ldots,t-1$. For reference below, we define these the MIDAS moments, which captures the information coming from to deviations between the MIDAS regressors based on $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_A$.

To simplify the following derivations, the covariance of high frequency exogenous observations is defined as,

$$\Gamma(q,p) = E\left[\boldsymbol{x}'_{\cdot,t,j}(\mathsf{L_t}^q\mathsf{L_j}^p\boldsymbol{x}_{\cdot,t,j})\right], \tag{B.6}$$

where q and p are the differences between the low-frequency and high-frequency time indexes, respectively.[9]

Substitute in (A.2) into $\mathcal{A}_{t,r}$ to obtain,

$$\begin{aligned}
\mathcal{A}_{t,r} &= E\left[\sum_{i=1}^{N}\mathsf{L_t}^r y_{i,t}\Delta\eta_{i,t}(\beta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right] \\
&= E\left[\sum_{i=1}^{N}\delta^{t-r}y_{i,0}\Delta\eta_{i,t}(\beta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right] \\
&\quad +E\left[\sum_{i=1}^{N}\left(\sum_{k=0}^{t-r-1}\delta^k\mathsf{L_t}^{r+k}x_{i,t}(\boldsymbol{\theta}_A)\right)\beta(\boldsymbol{\theta}_A)\Delta\eta_{i,t}(\beta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right] \\
&\quad +E\left[\sum_{i=1}^{N}\left(\sum_{k=0}^{t-r-1}\delta^k\mathsf{L_t}^{r+k}\nu_{i,t}\right)\Delta\eta_{i,t}(\beta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right] \\
&\quad +E\left[\sum_{i=1}^{N}\left(\mu_i\sum_{k=0}^{t-r-1}\delta^k\right)\Delta\eta_{i,t}(\beta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)\right],
\end{aligned} \tag{B.7}$$

[9] The variance at the high and low frequency are given by $\Gamma(q,p) = \Gamma(0,0)$ for covariance stationary series.

where under assumptions (A.3), (A.4) and (A.5) only the second term is nonzero. Substituting the definition of $\Delta\eta_{i,t}(\beta(\boldsymbol{\theta}_A);\boldsymbol{\theta}_A,\boldsymbol{\theta}_0)$ into the second term, and rearrange terms to obtain,

$$\begin{aligned} \mathcal{A}_{t,r} &= E\left[\sum_{i=1}^{N}\left(\sum_{k=0}^{t-r-1}\delta^k \mathsf{L_t}^{r+k}x_{i,t}(\boldsymbol{\theta}_A)\right)(x_{i,t}(\boldsymbol{\theta}_A)-x_{i,t}(\boldsymbol{\theta}_0))\right]\beta(\boldsymbol{\theta}_A)^2 \\ &\quad -E\left[\sum_{i=1}^{N}\left(\sum_{k=0}^{t-r-1}\delta^k \mathsf{L_t}^{r+k}x_{i,t}(\boldsymbol{\theta}_A)\right)(x_{i,t-1}(\boldsymbol{\theta}_A)-x_{i,t-1}(\boldsymbol{\theta}_0))\right]\beta(\boldsymbol{\theta}_A)^2. \end{aligned} \tag{B.8}$$

Applying (B.6) and rearranging terms leads to

$$\begin{aligned} \mathcal{A}_{t,r} &= \left(\sum_{k=0}^{t-r-1}\delta^k\sum_{j=1}^{m}\sum_{l=1}^{m}\Gamma(r+k,j-l)\left(w_j(\boldsymbol{\theta}_A)w_l(\boldsymbol{\theta}_A)-w_j(\boldsymbol{\theta}_A)w_l(\boldsymbol{\theta}_0)\right)\right)\beta(\boldsymbol{\theta}_A)^2 \\ &\quad -\left(\sum_{k=0}^{t-r-1}\delta^k\sum_{j=1}^{m}\sum_{l=1}^{m}\Gamma(r+k-1,j-l)\left(w_j(\boldsymbol{\theta}_A)w_l(\boldsymbol{\theta}_A)-w_j(\boldsymbol{\theta}_A)w_l(\boldsymbol{\theta}_0)\right)\right)\beta(\boldsymbol{\theta}_A)^2, \end{aligned} \tag{B.9}$$

which further simplifies to,

$$\begin{aligned} \mathcal{A}_{t,r} &= \left(\sum_{j=1}^{m}\sum_{l=1}^{m}\Gamma(r-1,j-l)w_j(\boldsymbol{\theta}_A)\left(w_l(\boldsymbol{\theta}_A)-w_l(\boldsymbol{\theta}_0)\right)\right)\beta(\boldsymbol{\theta}_A)^2 \\ &\quad + (1-\delta)\left(\sum_{k=0}^{t-r-1}\delta^k\sum_{j=1}^{m}\sum_{l=1}^{m}\Gamma(r+k,j-l)w_j(\boldsymbol{\theta}_A)\left(w_l(\boldsymbol{\theta}_A)-w_l(\boldsymbol{\theta}_0)\right)\right)\beta(\boldsymbol{\theta}_A)^2. \end{aligned} \tag{B.10}$$

Similar substitutions, rearrangements, and assumptions when applied to (B.5) will give,

$$\mathcal{B}_{t,q}=\left(\sum_{j=1}^{m}\sum_{l=1}^{m}\left(\Gamma(q,j-l)-\Gamma(q+1,j-l)\right)w_j(\boldsymbol{\theta}_0)\left(w_l(\boldsymbol{\theta}_A)-w_l(\boldsymbol{\theta}_0)\right)\right)\beta(\boldsymbol{\theta}_A). \tag{B.11}$$

The above shows that Theorem (1) can be equivalently expressed in terms of (B.10) and (B.11). So proving Theorem (1) involves demonstrating,

$$\mathcal{A}_{t,r}\neq 0 \quad \text{or} \quad \mathcal{B}_{t,q}\neq 0. \tag{B.12}$$

The weights from either $\boldsymbol{\theta}$ must sum to one, so a single deviation in a representative weight must result in a deviation in at least one other weight. It is clear from $w_l(\boldsymbol{\theta}_A)\neq w_l(\boldsymbol{\theta}_0)$ that $(w_l(\boldsymbol{\theta}_A)-w_l(\boldsymbol{\theta}_0))\neq 0$ and either $w_j(\boldsymbol{\theta}_A)\neq 0$ or $w_j(\boldsymbol{\theta}_0)\neq 0$ so the product must be non-negative for at least one j and l combination. When $\boldsymbol{\theta}_0$ corresponds to equal weights, then the double summation of $w_j(\boldsymbol{\theta}_0)\left(w_l(\boldsymbol{\theta}_A)-w_l(\boldsymbol{\theta}_0)\right)$, in (B.11), evaluates to zero independent of the $\boldsymbol{\theta}_A$-based weights. The above proves that (B.12) holds under these assumptions. So the trivial case of $\Gamma(0,0)\neq 0$ ensures $\mathcal{B}_{t,q}\neq 0$ when $q=0$. ■

Proof of Theorem 2. Because the cut-off point χ^2_α does not vary with $\boldsymbol{\theta}$, we have:

$$\inf_{\boldsymbol{\theta_0}}\mathcal{J}(\boldsymbol{\theta_0}) > \chi^2_\alpha \Longleftrightarrow \mathcal{J}(\boldsymbol{\theta})>\chi^2_\alpha, \forall\boldsymbol{\theta}\in\Theta \Longleftrightarrow CS(\boldsymbol{\theta};\alpha)=\emptyset. \tag{B.13}$$

■

Proof of Theorem 3. Under assumption (1), the t-test that fixes $\boldsymbol{\theta}$ is asymptotically valid in the sense that the underlying estimator is asymptotically normal and its variance covariance is consistently estimated given $\boldsymbol{\theta}$ for large N. If follows that

$$\inf_{\boldsymbol{\theta_0}\in\Theta} \mathsf{t}(\beta_0;\boldsymbol{\theta_0}) > \mathsf{t}_c(\alpha) \Leftrightarrow \mathsf{t}(\beta_0;\boldsymbol{\theta}) > \mathsf{t}_c(\alpha) \text{ for any } \boldsymbol{\theta}\in\Theta. \tag{B.14}$$

In the same vein

$$\inf_{\boldsymbol{\theta_0}\in CS(\boldsymbol{\theta_0};\alpha_1)} \mathsf{t}(\beta_0;\boldsymbol{\theta_0}) > \mathsf{t}_c(\alpha_2) \Leftrightarrow \mathsf{t}(\beta_0;\boldsymbol{\theta}) > \mathsf{t}_c(\alpha_2) \text{ for any } \boldsymbol{\theta}\in CS(\boldsymbol{\theta};\alpha_1).$$

■

C MIDAS Data Generating Process

Taking the first difference of a MIDAS regressor implies a specific structural form of the underlying data generation process. The first formulation of the high frequency observations defines a relationship between the j^{th} observation in period t and the j^{th} observation in period $t-1$, we repeat here (24) for a given i, t and j: $x_{i,t,j} = \psi_i + \rho\mathsf{L_t}x_{i,t,j} + \epsilon_{i,t,j}$.

The MIDAS regressor generated gives:

$$\begin{aligned} x_{i,t}(\boldsymbol{\theta}) &= \sum_{j=1}^{m} w_j(\boldsymbol{\theta})x_{i,t,j} && \text{(C.1)}\\ &= \psi_i + \rho\sum_{j=1}^{m} w_j(\boldsymbol{\theta})\mathsf{L_t}x_{i,t,j} + \sum_{j=1}^{m} w_j(\boldsymbol{\theta})\epsilon_{i,t,j}. && \text{(C.2)}\end{aligned}$$

The first difference (in t) of this MIDAS regressor from (24) gives (25), repeated here:

$$\Delta x_{i,t}(\boldsymbol{\theta}) = \rho\sum_{j=1}^{m} w_j(\boldsymbol{\theta})\Delta\mathsf{L_t}x_{i,t,j} + \sum_{j=1}^{m} w_j(\boldsymbol{\theta})\Delta\epsilon_{i,t,j}.$$

The second formulation is a predetermined relationship between the j and $j-1$ observations in period t, given by:

$$x_{i,t,j} = \psi_i + \gamma\mathsf{L_j}x_{i,t,j} + \epsilon_{i,t,j}. \tag{C.3}$$

To obtain the first difference in t of this formulation of the high frequency observations, we begin with recursion of the series to obtain,

$$x_{i,t,j} = \psi_i\sum_{k=0}^{m-1}\gamma^k + \gamma^m\mathsf{L_t}x_{i,t,j} + \left[\sum_{k=0}^{m-1}\gamma^k\mathsf{L_j}^k\epsilon_{i,t,j}\right], \tag{C.4}$$

where the bracketed part is a moving average error term, denoted $e_{i,t,j}(\gamma)$.

The MIDAS aggregation from (C.4) gives:

$$\begin{aligned} x_{i,t}(\boldsymbol{\theta}) &= \sum_{j=1}^{m} w_j(\boldsymbol{\theta})x_{i,t,j} && \text{(C.5)}\\ &= \psi_i\sum_{k=0}^{m-1}\gamma^k + \gamma^m\sum_{j=1}^{m} w_j(\boldsymbol{\theta})\mathsf{L_t}x_{i,t,j} + \sum_{j=1}^{m} w_j(\boldsymbol{\theta})e_{i,t,j}(\gamma). && \text{(C.6)}\end{aligned}$$

Taking the first difference in t of this second MIDAS formulation results in:

$$\Delta x_{i,t}(\boldsymbol{\theta}) = \gamma^m\sum_{j=1}^{m} w_j(\boldsymbol{\theta})\Delta\mathsf{L_t}x_{i,t,j} + \sum_{j=1}^{m} w_j(\boldsymbol{\theta})\Delta e_{i,t,j}(\gamma). \tag{C.7}$$

Equations (25) and (C.7) both satisfy the condition that the MIDAS aggregate is uncorrelated with $\nu_{i,s}$ from (3) for all s in t. It is apparent that the moving average component in (C.7) leads to correlation at the high frequency observations.

For the simulation study, the original Arellano and Bond assumption that $\rho = 0.8$ at the low frequency is maintained. It is clear that (25) is more managable, since the value can be set. For (C.7), the relationship between ρ and γ can be derived from the recursion of the series, $\gamma = \rho^{1/m}$, so for a stationary low frequency series the value of γ approaches unity as m increases, and is further complicated by the moving average component of the errors. To avoid complications of the moving average terms and the approach of the high frequency observations to unity, our simulation study employs the first representation of the high frequency observations which leads to correlation at the low frequency only.

D A Special Case under the Alternative

In the simulation study, we examine a special case that provides guidance on instrument importance under some common weight combinations: either $\boldsymbol{\theta}_A$ or $\boldsymbol{\theta}_0$ leading to equal weights, and when the weighting schemes are extreme opposites. To better illustrate implications of these settings, we introduce a special case for the data generation process of the MIDAS regressor. This case allows us to decompose the information content of each MIDAS population moment under these various alternative hypotheses. Clearly, under the null the moment conditions hold. Our special case of the data generating process of the high frequency observation is defined as:

$$\Gamma(q,p) = 0 \text{ for all } p \neq 0, \text{ and} \tag{D.1}$$

$$\Gamma(q,0) \neq 0 \text{ for all } q \geq 0, \tag{D.2}$$

which means that the covariance between high frequency observations are zero except when the pair are a high frequency match (e.g., $x_{i,t,j}$ and $\mathsf{L_t}^q x_{i,t,j}$). This data structure ensures that any aggregation of the high frequency observations will be autoregressive, with a know parameter. It is well known that aggregating an autoregressive high frequency series will result in an autoregressive moving average low frequency series, and the parameters of this ARMA series are sensitive to the aggregation weights.[10] This special case is implemented in the simulation study, and we also examine the case when high frequency observations follow an autoregressive process.

Under this special case, the population moments defined by (B.10) are,

$$\mathcal{A}_{t,r} = \left(\Gamma(r-1,0) + (1-\delta)\sum_{k=0}^{t-r-1}\delta^k\Gamma(r+k,0)\right)\boldsymbol{w}'(\boldsymbol{\theta}_A)\left(\boldsymbol{w}(\boldsymbol{\theta}_A) - \boldsymbol{w}(\boldsymbol{\theta}_0)\right)\beta(\boldsymbol{\theta}_A)^2 \tag{D.3}$$

and (B.11) are,

$$\mathcal{B}_{t,q} = \big((\Gamma(q,0) - \Gamma(q+1,0))\,\boldsymbol{w}'(\boldsymbol{\theta}_0)\left(\boldsymbol{w}(\boldsymbol{\theta}_A) - \boldsymbol{w}(\boldsymbol{\theta}_0)\right)\big)\,\beta(\boldsymbol{\theta}_A), \tag{D.4}$$

where $q \geq 0$ and where $r \geq 2$.

Case 1 (Equal MIDAS Weights) *Consider the case where $\boldsymbol{\theta}_A$ generates **equal** MIDAS weights*

[10]Detailed derivations and discussion provided in Appendix C.

on each high frequency observation, $w_j(\boldsymbol{\theta}_A) = 1/m$, where the null hypothesis imposes ***unequal*** *weights, so*

$$H_0 : w_j(\boldsymbol{\theta}_0) \neq \frac{1}{m} \text{ for at least one } j. \tag{D.5}$$

The MIDAS moments simplify to

$$\begin{aligned} \mathcal{A}_{t,r} &= 0 \text{ and} & \text{(D.6)} \\ \mathcal{B}_{t,q} &= (\Gamma(q,0) - \Gamma(q+1,0)) \left(\frac{1}{m} - \boldsymbol{w}'(\boldsymbol{\theta}_0)\boldsymbol{w}(\boldsymbol{\theta}_0)\right) \beta(\boldsymbol{\theta}_A). & \text{(D.7)} \end{aligned}$$

The result for $\mathcal{A}_{t,r}$ is evident since $w_j(\boldsymbol{\theta}_A)$ does not vary with j, while maintaining the sum to unity. By theorem (1), the first moment evaluates to zero for all values of $\boldsymbol{\theta}_0$, which means the information content on $\boldsymbol{\theta}$ is concentrated in $\mathcal{B}_{t,q}$.

Case 2 (Null of Equal MIDAS Weights) *Consider the case where $\boldsymbol{\theta}_A$ generates* ***unequal*** *MIDAS weights on each high frequency observation ($w_j(\boldsymbol{\theta}_A) \neq 1/m$), where the null hypothesis imposes* ***equal*** *weights, so*

$$H_0 : w_j(\boldsymbol{\theta}_0) = \frac{1}{m} \text{ for all } j. \tag{D.8}$$

The MIDAS moments are

$$\mathcal{A}_{t,r} = \left(\Gamma(r-1,0) + (1-\delta)\sum_{k=0}^{t-r-1} \delta^k \Gamma(r+k,0)\right)\left(-\frac{1}{m} + \boldsymbol{w}'(\boldsymbol{\theta}_A)\boldsymbol{w}(\boldsymbol{\theta}_A)\right)\beta(\boldsymbol{\theta}_A)^2 \tag{D.9}$$

and

$$\mathcal{B}_{t,q} = 0. \tag{D.10}$$

By Case 2, the second moment evaluates to zero for all values of $\boldsymbol{\theta}_A$, since $w_j(\boldsymbol{\theta}_0)$ does not depend on j, so the information content on $\boldsymbol{\theta}$ is concentrated in $\mathcal{A}_{t,r}$.

Case 3 (Extreme Opposite MIDAS Weights) *In this case, both $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_0$ generate* ***unequal*** *MIDAS weights on each high frequency observation, where the j-th weight for $\boldsymbol{\theta}_A$ is zero if the corresponding j-th weight for $\boldsymbol{\theta}_0$ is non-zero, and vice versa, formally this means $w_j(\boldsymbol{\theta}_A)w_j(\boldsymbol{\theta}_0) = 0$ for all j. The MIDAS moment conditions are*

$$\begin{aligned} \mathcal{A}_{t,r} &= \left(\Gamma(r-1,0) + (1-\delta)\sum_{k=0}^{t-r-1} \delta^j \Gamma(r+k,0)\right)\boldsymbol{w}'(\boldsymbol{\theta}_A)\boldsymbol{w}(\boldsymbol{\theta}_A)\beta(\boldsymbol{\theta}_A)^2 \text{ and} & \text{(D.11)} \\ \mathcal{B}_{t,q} &= -\left(\Gamma(q,0) - \Gamma(q+1,0)\right)\boldsymbol{w}'(\boldsymbol{\theta}_0)\boldsymbol{w}(\boldsymbol{\theta}_0)\beta(\boldsymbol{\theta}_A). & \text{(D.12)} \end{aligned}$$

Case 3 shows that the information content of the MIDAS population moments are separable when $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_0$ are extreme opposites.

These three cases demonstrate that identification of $\boldsymbol{\theta}$ requires both y and x-based instruments. Furthermore, these cases demonstrate that power to discriminate between alternatives will be asymmetric, since the information content of the moments is not symmetric over $\boldsymbol{\theta}$. These cases also show that excluding moments can result in a loss of power, especially considering that some moment results evaluate to exactly zero. The Davies problem is reinforced by these cases, since the identification of $\boldsymbol{\theta}$ requires that $\beta(\boldsymbol{\theta}_A) \neq 0$.

E Tables

Table E1: Rejection Frequency for Sargan Test: Equal Weights†

		$N = 500$			$N = 1000$		
θ_1^A	θ_2^A	$T = 5$	$T = 10$	$T = 15$	$T = 5$	$T = 10$	$T = 15$
Size							
0	0	0.041	0.051	0.041	0.052	0.042	0.043
Power							
-1	-1	0.054	0.055	0.048	0.034	0.054	0.086
	-0.5	0.054	0.049	0.051	0.048	0.061	0.077
	0	0.046	0.042	0.041	0.057	0.071	0.063
	0.5	0.053	0.05	0.053	0.042	0.055	0.091
	1	0.051	0.048	0.058	0.051	0.059	0.099
-0.5	-1	0.059	0.045	0.054	0.035	0.067	0.08
	-0.5	0.053	0.041	0.056	0.042	0.063	0.066
	0	0.055	0.042	0.06	0.041	0.05	0.06
	0.5	0.042	0.05	0.064	0.049	0.054	0.087
	1	0.054	0.042	0.057	0.06	0.065	0.086
0	-1	0.056	0.04	0.035	0.047	0.063	0.077
	-0.5	0.05	0.047	0.063	0.05	0.053	0.077
	0	-	-	-	-	-	-
	0.5	0.052	0.051	0.057	0.049	0.052	0.086
	1	0.052	0.047	0.051	0.045	0.064	0.079
0.5	-1	0.045	0.05	0.053	0.052	0.055	0.077
	-0.5	0.062	0.042	0.053	0.05	0.061	0.067
	0	0.051	0.045	0.06	0.039	0.057	0.068
	0.5	0.05	0.058	0.052	0.054	0.055	0.112
	1	0.038	0.063	0.058	0.053	0.038	0.083
1	-1	0.046	0.046	0.059	0.047	0.054	0.081
	-0.5	0.04	0.036	0.056	0.04	0.055	0.073
	0	0.034	0.036	0.075	0.064	0.062	0.08
	0.5	0.046	0.046	0.046	0.038	0.055	0.081
	1	0.05	0.058	0.06	0.047	0.038	0.092

†MIDAS weights are equal for all high frequency observation ($\boldsymbol{\theta}^A$) and a significance level of 5%.

Table E2: Rejection Frequency for Sargan Test: Large Weights on Recent Observations†

		$N = 500$			$N = 1000$		
θ_1^B	θ_2^B	$T = 5$	$T = 10$	$T = 15$	$T = 5$	$T = 10$	$T = 15$
Size							
0.1	-0.2	0.028	0.051	0.049	0.020	0.046	0.042
Power							
-1	-1	0.062	0.192	0.738	0.031	0.510	0.987
	-0.5	0.055	0.134	0.541	0.037	0.379	0.909
	0	0.051	0.039	0.070	0.018	0.041	0.072
	0.5	0.038	0.047	0.046	0.014	0.043	0.066
	1	0.046	0.060	0.055	0.023	0.030	0.066
-0.5	-1	0.064	0.153	0.700	0.041	0.512	0.979
	-0.5	0.046	0.104	0.414	0.037	0.279	0.813
	0	0.038	0.047	0.059	0.017	0.061	0.064
	0.5	0.042	0.049	0.047	0.023	0.045	0.063
	1	0.031	0.058	0.061	0.021	0.034	0.059
0	-1	0.056	0.147	0.613	0.050	0.415	0.965
	-0.5	0.042	0.078	0.218	0.035	0.157	0.507
	0	0.032	0.655	0.597	0.026	0.866	0.944
	0.5	0.040	0.045	0.040	0.021	0.049	0.062
	1	0.050	0.042	0.054	0.020	0.037	0.058
0.5	-1	0.054	0.138	0.564	0.025	0.365	0.925
	-0.5	0.041	0.046	0.101	0.020	0.068	0.117
	0	0.050	0.090	0.042	0.029	0.042	0.026
	0.5	0.036	0.062	0.047	0.033	0.041	0.067
	1	0.040	0.056	0.052	0.019	0.044	0.061
1	-1	0.048	0.101	0.435	0.024	0.308	0.819
	-0.5	0.030	0.041	0.048	0.016	0.053	0.046
	0	0.055	0.054	0.035	0.029	0.034	0.046
	0.5	0.039	0.044	0.045	0.018	0.027	0.055
	1	0.046	0.051	0.063	0.030	0.043	0.060

†MIDAS weights are much larger for more recent high frequency observation ($\boldsymbol{\theta}^B$) and a significance level of 5%.

Table E3: Rejection Frequency for Sargan Test: Recent Observations have more Weight[†]

		$N = 500$			$N = 1000$		
θ_1^C	θ_2^C	$T = 5$	$T = 10$	$T = 15$	$T = 5$	$T = 10$	$T = 15$
Size							
0.03	-0.02	0.042	0.049	0.04	0.046	0.053	0.05
Power							
-1	-1	0.070	0.060	0.088	0.038	0.117	0.235
	-0.5	0.058	0.051	0.074	0.042	0.112	0.211
	0	0.052	0.043	0.061	0.055	0.109	0.158
	0.5	0.070	0.058	0.109	0.047	0.102	0.162
	1	0.068	0.078	0.097	0.061	0.112	0.163
-0.5	-1	0.064	0.044	0.110	0.045	0.108	0.223
	-0.5	0.065	0.048	0.094	0.043	0.125	0.202
	0	0.057	0.046	0.061	0.036	0.051	0.083
	0.5	0.054	0.072	0.098	0.061	0.110	0.159
	1	0.058	0.068	0.101	0.061	0.096	0.165
0	-1	0.058	0.036	0.068	0.045	0.115	0.210
	-0.5	0.061	0.046	0.101	0.058	0.113	0.203
	0	0.044	0.078	0.082	0.062	0.073	0.127
	0.5	0.066	0.067	0.110	0.049	0.102	0.160
	1	0.062	0.070	0.090	0.052	0.112	0.152
0.5	-1	0.066	0.052	0.108	0.051	0.109	0.218
	-0.5	0.056	0.048	0.082	0.051	0.109	0.175
	0	0.055	0.059	0.100	0.041	0.094	0.136
	0.5	0.070	0.076	0.104	0.067	0.123	0.183
	1	0.058	0.071	0.095	0.054	0.119	0.156
1	-1	0.044	0.050	0.107	0.060	0.121	0.225
	-0.5	0.041	0.048	0.074	0.045	0.114	0.165
	0	0.043	0.049	0.115	0.058	0.110	0.168
	0.5	0.063	0.060	0.100	0.040	0.088	0.157
	1	0.059	0.077	0.101	0.057	0.121	0.168

†MIDAS weights are modestly larger for more recent high frequency observation ($\boldsymbol{\theta}^C$) and a significance level of 5%.

Table E4: Rejection Frequency for Sargan Test: Older Observations have more Weight†

		$N = 500$			$N = 1000$		
θ_1^D	θ_2^D	$T = 5$	$T = 10$	$T = 15$	$T = 5$	$T = 10$	$T = 15$
Size							
-0.06	0.01	0.042	0.034	0.044	0.035	0.046	0.035
Power							
-1	-1	0.055	0.049	0.045	0.050	0.083	0.126
	-0.5	0.048	0.044	0.052	0.062	0.078	0.114
	0	0.050	0.034	0.049	0.068	0.093	0.101
	0.5	0.047	0.070	0.092	0.053	0.096	0.229
	1	0.042	0.067	0.084	0.062	0.094	0.238
-0.5	-1	0.056	0.050	0.052	0.058	0.085	0.119
	-0.5	0.040	0.042	0.050	0.056	0.084	0.100
	0	0.052	0.045	0.063	0.053	0.083	0.093
	0.5	0.036	0.062	0.097	0.066	0.099	0.249
	1	0.056	0.060	0.101	0.074	0.096	0.235
0	-1	0.050	0.044	0.038	0.062	0.081	0.112
	-0.5	0.041	0.047	0.059	0.058	0.075	0.117
	0	0.043	0.077	0.091	0.077	0.109	0.225
	0.5	0.051	0.057	0.099	0.052	0.096	0.225
	1	0.046	0.056	0.070	0.061	0.094	0.231
0.5	-1	0.052	0.046	0.042	0.065	0.082	0.112
	-0.5	0.062	0.039	0.063	0.053	0.080	0.102
	0	0.058	0.043	0.046	0.043	0.046	0.038
	0.5	0.050	0.061	0.093	0.066	0.096	0.230
	1	0.033	0.073	0.103	0.062	0.094	0.220
1	-1	0.045	0.050	0.061	0.064	0.080	0.113
	-0.5	0.045	0.033	0.055	0.052	0.079	0.108
	0	0.035	0.048	0.087	0.062	0.066	0.094
	0.5	0.043	0.068	0.084	0.055	0.099	0.204
	1	0.049	0.067	0.091	0.067	0.096	0.247

†MIDAS weights are modestly larger for more older high frequency observation ($\boldsymbol{\theta}^D$) and a significance level of 5%.

Table E5: Rejection Frequency for Sargan Test: Large Weights on Older Observations†

θ_1^E	θ_2^E	$N=500$			$N=1000$		
		$T=5$	$T=10$	$T=15$	$T=5$	$T=10$	$T=15$
Size							
-0.04	0.02	0.036	0.036	0.039	0.035	0.037	0.042
Power							
-1	-1	0.061	0.038	0.028	0.051	0.049	0.060
	-0.5	0.051	0.030	0.024	0.054	0.039	0.040
	0	0.043	0.023	0.024	0.054	0.054	0.036
	0.5	0.077	0.149	0.273	0.109	0.367	0.760
	1	0.067	0.153	0.264	0.103	0.361	0.746
-0.5	-1	0.047	0.030	0.026	0.050	0.051	0.054
	-0.5	0.041	0.026	0.031	0.050	0.058	0.050
	0	0.046	0.032	0.037	0.040	0.055	0.048
	0.5	0.057	0.151	0.269	0.127	0.365	0.738
	1	0.075	0.135	0.270	0.119	0.345	0.751
0	-1	0.051	0.029	0.021	0.055	0.051	0.043
	-0.5	0.039	0.034	0.035	0.052	0.043	0.053
	0	0.060	0.242	0.472	0.275	0.680	0.968
	0.5	0.080	0.145	0.276	0.107	0.329	0.728
	1	0.069	0.145	0.261	0.110	0.367	0.741
0.5	-1	0.046	0.031	0.027	0.050	0.044	0.051
	-0.5	0.058	0.024	0.036	0.044	0.042	0.047
	0	0.049	0.042	0.051	0.050	0.053	0.054
	0.5	0.071	0.155	0.283	0.120	0.367	0.718
	1	0.045	0.147	0.285	0.117	0.361	0.753
1	-1	0.041	0.037	0.023	0.054	0.047	0.053
	-0.5	0.049	0.025	0.025	0.049	0.045	0.047
	0	0.035	0.041	0.053	0.055	0.055	0.043
	0.5	0.072	0.134	0.274	0.116	0.367	0.710
	1	0.065	0.144	0.259	0.122	0.361	0.755

†MIDAS weights are much larger for more older high frequency observation ($\boldsymbol{\theta}^E$) and a significance level of 5%.

Table E6: Rejection Frequency for Sargan Test: AR(1) High Frequency Series†

		$\boldsymbol{\theta}_0$				
θ_1	θ_2	$(-.06, .01)$	$(-.04, .02)$	$(0, 0)$	$(.03, -.02)$	$(.1, -.2)$
Size						
-0.06	0.01	0.017				
-0.04	0.02		0.027			
0	0			0.031		
0.03	-0.02				0.026	
0.1	-0.2					0.029
Power						
-1	-1	0.021	0.032	0.03	0.02	0.022
-1	-0.5	0.024	0.031	0.029	0.024	0.023
-1	0	0.027	0.029	0.035	0.025	0.024
-1	0.5	0.854	0.037	1	1	0.998
-1	1	0.854	0.037	1	1	0.998
-0.5	-1	0.021	0.032	0.03	0.019	0.022
-0.5	-0.5	0.026	0.031	0.03	0.025	0.024
-0.5	0	0.039	0.036	0.083	0.026	0.028
-0.5	0.5	0.854	0.037	1	1	0.998
-0.5	1	0.854	0.037	1	1	0.998
0	-1	0.021	0.032	0.028	0.021	0.022
0	-0.5	0.024	0.032	0.031	0.024	0.026
0	0	0.918	0.587	–	1	1
0	0.5	0.854	0.037	1	1	0.998
0	1	0.854	0.037	1	1	0.998
0.5	-1	0.024	0.031	0.028	0.024	0.023
0.5	-0.5	0.027	0.032	0.031	0.023	0.03
0.5	0	0.998	0.027	1	1	0.997
0.5	0.5	0.854	0.037	1	1	0.998
0.5	1	0.854	0.037	1	1	0.998
1	-1	0.025	0.031	0.03	0.025	0.024
1	-0.5	0.027	0.028	0.031	0.024	0.025
1	0	0.963	0.044	1	1	0.998
1	0.5	0.854	0.037	1	1	0.998
1	1	0.854	0.037	1	1	0.998

†Series based on (C.3) with $\gamma = 0.8$.