Yet another look at MIDAS regression

Philip Hans Franses

Econometric Institute, Erasmus School of Economics

Econometric Institute Report 2016-32

Abstract

A MIDAS regression involves a dependent variable observed at a low frequency and independent variables observed at a higher frequency. This paper relates a true high frequency data generating process, where also the dependent variable is observed (hypothetically) at the high frequency, with a MIDAS regression. It is shown that a correctly specified MIDAS regression usually includes lagged dependent variables, a substantial number of explanatory variables (observable at the low frequency) and a moving average term. Next, the parameters of the explanatory variables unlikely obey certain convenient patterns, and hence imposing such restrictions in practice is not recommended.

JEL Code: C32

Keywords: High frequency; low frequency; MIDAS regression

This version: August 24 2016

Author's notes: I thank Eva Janssens for excellent research assistance and Peter Boswijk and Michael Clements for helpful suggestions. Correspondence to PH Franses, Econometric Institute, Erasmus School of Economics, POB 1738, 3000 DR Rotterdam, The Netherlands, franses@ese.eur.nl

Introduction

MIDAS regressions typically involve a dependent variable observed at a low frequency and independent variables observed at a higher frequency. A MIDAS regression often also includes lags of the low frequency dependent variable on the right hand side. An example is quarterly growth in Gross Domestic Product as the dependent variable and the monthly observed growth in industrial production as an explanatory variable. Another example is category sales observed per 13 weeks, possibly being explained by weekly observed sales levels of a prominent product within that category. The first mention of a MIDAS regression is Ghysels et al. (2004), and since then a variety of studies have implemented and extended the initial MIDAS regression, see for example Clements and Galvao (2008), Andreou et al. (2010) and Ghysels et al. (2007). The focus is often on forecasting and it is frequently found that a MIDAS regression delivers accurate forecasts, usually more accurate that models where all variables are aggregated at the low frequency.

Usually, the MIDAS regression is considered in practice with the assumption that the underlying data generating process (DGP) also is a MIDAS model, that is, the variables are in reality only observable at different frequencies. In this paper, I provide an alternative look at MIDAS regressions by assuming that the underlying DGP is a model at the high frequency for both the dependent and the independent variables, but that somehow only low frequency observations of the dependent variable are available. This adds to the literature where the consequences of (temporal) aggregation are discussed, where usually the focus is on the aggregation of both the dependent and the independent variables, see Lütkepohl (1986) and Marcellino (1999). As a running example, also to save notation, I will consider quarters as the low frequency and months as the high frequency, but the main results qualitatively apply to any other mix of frequencies. The simulation experiments will consider years and half years, also to reduce the number of tables. Again, the findings from the simulations will carry on through to other mixes of frequencies.

With a DGP at the high frequency, it can be derived how the appropriate MIDAS regression looks like when there are two different frequencies. In the example of an autoregressive distributed lag (ADL) model of orders (3,3), which is a rather parameter-rich model at the high frequency, I can derive the exact expression of the appropriate corresponding MIDAS regression. This illustration leads to three key insights, which are all generalizable to other DGPs. The first is that the numbers of lags in a MIDAS regression, both for the dependent low frequency variable and for the independent high frequency variables, are not necessarily high, which seems in contrast to what one usually sees in empirical studies. The second insight is that currently fashionable restrictions on the parameters of the independent variables, like the Almon structure or the logit expressions, potentially have no particular meaning. The parameters in the high frequency DGP can lead to a wealth of different parameter configurations in the MIDAS regressions, and particular systematic patterns might be rare. The third insight is that, in some cases, the parameters in the MIDAS regression are informative about the true parameters in the high frequency DGP. This can lead to proper estimates of carryover effects, short run effects, and cumulative effects in the true high-frequency ADL process, should one be interested in that exercise.

The paper proceeds as follows. In Section 2, I present a high frequency DGP at the monthly level, and derive the features of the associated MIDAS regression where the dependent variable is observed only quarterly. I rely on a notation that is different from the one currently in use in the literature to effectively indicate the transition from DGP to a MIDAS regression. This transformation might be called a high frequency/low frequency transformation, or in brief, a HILO transformation. Our notation resembles the one that is typically used for periodic time series models for seasonal data, see Franses (1994) and Franses and Paap (2004). The main insights are also summarized in this Section 2. In Section 3, I report on simulation experiments to examine the properties of the least squares estimators in the MIDAS regression. It is possible to quantify the loss of forecasting power in case a MIDAS regression is used instead of a model that perfectly matches with the true high-frequency DGP. Section 4 includes an illustration for total vehicle sales per quarter and monthly observed sales of new passenger cars. Section 5 concludes with a summary of the main findings and a few suggestions for future research.

2. Quarterly (dependent) and monthly (explanatory) variables

The idea of this paper is to derive the properties of a MIDAS regression when there is a true data generating process at a high frequency, for which the time indicator is t and where the dependent variable turns out to be observable only at a low frequency, with indicator T. The difference between the two frequencies is denoted as S. So, t can refer to quarterly observations and T to annual observations, which makes S equal to 4. All derivations in this paper concern flow variables, that is, the low-frequency variables associate with the sum or the average of the high frequency variables. To save notation, but at the same time, to be able to illustrate as much as possible, I will use $S = 3$ in this section.

Assume that the DGP at the high frequency level is an autoregressive distributed lag model with lags (3,3) (ADL(3,3)), that is,

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \varepsilon_t \qquad (1)$$

where ε_t is a standard white noise process with mean 0 and variance σ^2. This is a rather parameter-rich model, which is perhaps rarely fitted to actual data in practice, and below it will be demonstrated what the consequences are of parameter restrictions. The key idea of this paper is first to write (1) in a low frequency format, that is, to use a HILO notation transformation. When S = 3, the equivalent of (1) is then

$$\begin{pmatrix} 1 & 0 & 0 \\ -\alpha_1 & 1 & 0 \\ -\alpha_2 & -\alpha_1 & 1 \end{pmatrix} \begin{pmatrix} Y_{1,T} \\ Y_{2,T} \\ Y_{3,T} \end{pmatrix} = \begin{pmatrix} \alpha_3 & \alpha_2 & \alpha_1 \\ 0 & \alpha_3 & \alpha_2 \\ 0 & 0 & \alpha_3 \end{pmatrix} \begin{pmatrix} Y_{1,T-1} \\ Y_{2,T-1} \\ Y_{3,T-1} \end{pmatrix} + \begin{pmatrix} \beta_0 & 0 & 0 \\ \beta_1 & \beta_0 & 0 \\ \beta_2 & \beta_1 & \beta_0 \end{pmatrix} \begin{pmatrix} X_{1,T} \\ X_{2,T} \\ X_{3,T} \end{pmatrix} +$$
$$\begin{pmatrix} \beta_3 & \beta_2 & \beta_1 \\ 0 & \beta_3 & \beta_2 \\ 0 & 0 & \beta_3 \end{pmatrix} \begin{pmatrix} X_{1,T-1} \\ X_{2,T-1} \\ X_{3,T-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,T} \\ \varepsilon_{2,T} \\ \varepsilon_{3,T} \end{pmatrix} \qquad (2)$$

Hence, the letters *y* and *x* are used for the high frequency and *Y* and *X* for the low frequency. The inverse of the left-hand side matrix in (3), say A_0, is

$$A_0^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \alpha_1 & 1 & 0 \\ \alpha_1^2 + \alpha_2 & \alpha_1 & 1 \end{pmatrix}$$

Multiplying both sides of (2) with this inverse results in

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_{1,T} \\ Y_{2,T} \\ Y_{3,T} \end{pmatrix} = A_0^{-1} A_1 \begin{pmatrix} Y_{1,T-1} \\ Y_{2,T-1} \\ Y_{3,T-1} \end{pmatrix} + A_0^{-1} B_0 \begin{pmatrix} X_{1,T} \\ X_{2,T} \\ X_{3,T} \end{pmatrix} +$$
$$A_0^{-1} B_1 \begin{pmatrix} X_{1,T-1} \\ X_{2,T-1} \\ X_{3,T-1} \end{pmatrix} + A_0^{-1} \begin{pmatrix} \varepsilon_{1,T} \\ \varepsilon_{2,T} \\ \varepsilon_{3,T} \end{pmatrix} \qquad (3)$$

with

$$A_0^{-1}A_1 = \begin{pmatrix} \alpha_3 & \alpha_2 & \alpha_1 \\ \alpha_1\alpha_3 & \alpha_1\alpha_2 + \alpha_3 & \alpha_1^2 + \alpha_2 \\ \alpha_1^2\alpha_3 + \alpha_2\alpha_3 & \alpha_1^2\alpha_2 + \alpha_2^2 + \alpha_1\alpha_3 & \alpha_1^3 + 2\alpha_1\alpha_2 + \alpha_3 \end{pmatrix}$$

where the matrices $A_0^{-1}B_0$ and $A_0^{-1}B_1$ have similarly-looking expressions now also involving the β parameters. A final part of the HILO notation transformation is to multiply both sides of (3) with the vector $(1 \quad 1 \quad 1)$, which results in

$$\begin{gathered} Y_T = Y_{1,T} + Y_{2,T} + Y_{3,T} = \alpha_1^* Y_{1,T-1} + \alpha_2^* Y_{2,T-1} + \alpha_3^* Y_{3,T-1} \\ +\beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \varepsilon_T \end{gathered} \tag{4}$$

where ε_T is again a white noise process now at the low frequency with a variance that is a function of the α parameters and σ^2. The parameters in (4) are functions of the original parameters in (1).

There are two key features to be observed from (4). The first is that the right-hand side includes three (and in general *S*) unobserved variables, that is, $Y_{1,T-1}$, $Y_{2,T-1}$ and $Y_{3,T-1}$, and these are unobserved because Y_{T-1} is assumed to be only observable at the low frequency. The second feature is that a pattern in the values of the β_0^* to β_5^* parameters depends on the (unknown) values of the underlying parameters in the high frequency DGP.

Table 1 presents two sets of parameter patterns for two specific versions of (1). Evidently, any pattern on the parameters can emerge, depending on the parameters in the high frequency DGP. This suggests that imposing structures on the parameters in a MIDAS regression using for example Almon lags or a logistic function does not seem to make sense, a priori. Next, if one were interested in the parameters in the high frequency DGP, one may impose and test parameter restrictions in the MIDAS regression using expressions like those as in Table 1.

How to deal with the *S* unobservable variables?

There are various strategies to deal with the $S = 3$ unobserved variables $Y_{1,T-1}$, $Y_{2,T-1}$ and $Y_{3,T-1}$, and later on I will rely on some simulation experiments to examine the consequences of each strategy.

A first, but unwise, strategy is to consider the MIDAS regression

$$Y_T = \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \varepsilon_T \qquad (5)$$

which simply ignores lagged information on the dependent variable. As (1) and (4) show, the omitted variables then are correlated with the included explanatory variables, and this will imply biased estimates.

A second strategy is to rely on the approximation that $Y_{1,T-1} = \gamma Y_{T-1}$, $Y_{2,T-1} = \gamma Y_{T-1}$ and $Y_{3,T-1} = \gamma Y_{T-1}$, indeed assuming that γ is the same each time. In that case, (4) becomes

$$\begin{aligned} Y_T &= (\alpha_1^* + \alpha_2^* + \alpha_3^*)\gamma Y_{T-1} \\ &+ \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \varepsilon_T \qquad (6) \end{aligned}$$

which now includes only observable variables on the right hand side. It is however not certain that γ is constant over time, so the inclusion of Y_{T-1} potentially implies the inclusion of a variable with measurement error.

A third strategy involves looking again at (1), and to recognize that a model for $Y_{1,T-1}$ includes $X_{1,T-1}, X_{3,T-2}, X_{2,T-2}, X_{1,T-2}$ as well as $\varepsilon_{1,T-1}$. For $Y_{2,T-1}$ this entails $X_{2,T-1}, X_{1,T-1}, X_{3,T-2}$, $X_{1,T-2}$ and $\varepsilon_{2,T-1}$, while for $Y_{3,T-1}$ this entails $X_{3,T-1}, X_{2,T-1}, X_{1,T-1}$, $X_{3,T-2}$ as well as $\varepsilon_{3,T-1}$. At the same time it holds that $Y_{1,T-1} = Y_{T-1} - (Y_{2,T-1} + Y_{3,T-1})$, that $Y_{2,T-1} = Y_{T-1} - (Y_{1,T-1} + Y_{3,T-1})$, and that $Y_{3,T-1} = Y_{T-1} - (Y_{1,T-1} + Y_{2,T-1})$. This suggests that a general MIDAS regression, which replaces the S = 3 unobservable variables by observables, looks like

$$\begin{aligned} Y_T &= \rho Y_{T-1} + \beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + \\ &\quad \beta_6^* X_{3,T-2} + \beta_7^* X_{2,T-2} + \beta_8^* X_{1,T-2} + v_T + \theta v_{T-1} \qquad (7) \end{aligned}$$

This MIDAS regression model in (7) has the correct dynamics given the high-frequency DGP in (1). Again, no particular structure on the β_j^* parameters is to be expected. Note that (7) involves a moving average term of order 1 (MA(1)), which is due to the inclusion of $\varepsilon_{1,T-1}$, of $\varepsilon_{2,T-1}$ and of $\varepsilon_{3,T-1}$. Exclusion of this MA term, as is apparently usually done in empirical applications, amounts to misspecification. However, when S increases, one may expect that θ approaches 0, as will also be clear from the expressions in the next section when some simulations will be performed. The parameter ρ is a function of the parameters α_1 and α_2.

The exercise in this section leads to three general conclusions. A correctly specified MIDAS regression usually includes lagged dependent variables, a substantial number of explanatory variables and a moving average term. The parameters of the explanatory variables cannot be expected to obey certain convenient patterns, and it is therefore not recommended to impose such restrictions. When S is large, the impact of the ignorance of the moving average term shall become small, but that will also depend on the size of the parameters for the lagged dependent variables in the high frequency DGP. In specific cases, one may wish to impose and to test restrictions on the parameters in the MIDAS regression to learn about the underlying parameters in the high frequency DGP. When the high frequency DGP includes a small number of lags either of the dependent or the independent variables, then the size of the MIDAS regression quickly reduces.

3. Simulations for S = 2

In this section, I present various simulation experiments to illustrate the practical consequences of the HILO transformation and the various expressions for MIDAS regressions.

Suppose now that there is a DGP at a "half-yearly frequency", that is,

$$y_t = \alpha y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t \tag{8}$$

Assuming S = 2, this DGP after HILO notation transformation is

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

Pre-multiplying both sides of (8) with the inverse of $\begin{pmatrix} 1 & 0 \\ -\alpha & 1 \end{pmatrix}$, one obtains

$$\begin{aligned} \begin{pmatrix} Y_{1,T} \\ Y_{2,T} \end{pmatrix} = \begin{pmatrix} 0 & \alpha \\ 0 & \alpha^2 \end{pmatrix} \begin{pmatrix} Y_{1,T-1} \\ Y_{2,T-1} \end{pmatrix} + \begin{pmatrix} \beta_0 & 0 \\ \alpha\beta_0 + \beta_1 & \beta_0 \end{pmatrix} \begin{pmatrix} X_{1,T} \\ X_{2,T} \end{pmatrix} \\ + \begin{pmatrix} 0 & \beta_1 \\ 0 & \alpha\beta_1 \end{pmatrix} \begin{pmatrix} X_{1,T-1} \\ X_{2,T-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,T} \\ \alpha\varepsilon_{1,T} + \varepsilon_{2,T} \end{pmatrix} \end{aligned} \tag{9}$$

Pre-multiplying (9) with the vector (1 1) gives

$$Y_T = Y_{1,T} + Y_{2,T} = (\alpha + \alpha^2)Y_{2,T-1} + (\beta_0 + \alpha\beta_0 + \beta_1)X_{1,T} + \beta_0 X_{2,T} + (\beta_1 + \alpha\beta_1)X_{2,T-1} + (\alpha + 1)\varepsilon_{1,T} + \varepsilon_{2,T} \quad (10)$$

where the variable $Y_{2,T-1}$ is unobserved.

In the simulations, I will consider three versions of a MIDAS regression relevant to (10). The first is

$$Y_T = \mu + \beta_0^* X_{2,T} + \beta_1^* X_{1,T} + \beta_2^* X_{2,T-1} + u_T$$

where it is known that the true parameters are $\beta_0^* = \beta_0$, $\beta_1^* = (1+\alpha)\beta_0 + \beta_1$, and $\beta_2^* = (1+\alpha)\beta_1 + \alpha\beta_0$. This model lacks three components, that is, the autoregressive term, the variable $X_{1,T-1}$ and the MA term.

The second MIDAS regression is

$$Y_T = \mu + \rho Y_{T-1} + \beta_0^* X_{2,T} + \beta_1^* X_{1,T} + \beta_2^* X_{2,T-1} + u_T$$

where now the variable $X_{1,T-1}$ and the MA term are missing. The true parameters are as above plus $\rho = \alpha^2$.

The third and final MIDAS regression replaces $Y_{2,T-1}$, as is done in (7), and it reads as

$$Y_T = \mu + \rho Y_{T-1} + \beta_0^* X_{2,T} + \beta_1^* X_{1,T} + \beta_2^* X_{2,T-1} + \beta_3^* X_{1,T-1} + u_T$$

where now only the MA term is ignored. Here, the true parameters are as before and additionally, $\beta_3^* = \alpha\beta_1$. All parameters in these three models are estimated using ordinary least squares (OLS).

Table 2 reports the simulation results for N = 40, 400 and 40000, and for α is 0.5, 0.8, 0.9 and 0.95, for the first mis-specified MIDAS regression. Tables 3 and 4 refer to the other two MIDAS regressions.

Table 2 shows that the larger is α, the more bias there is for δ_0 and δ_1, particularly for small samples, whereas bias for δ_2 persists. Table 3 shows that the inclusion of Y_{T-1} alleviates the bias for δ_0 and δ_1, and substantially reduces the bias for δ_2, while the parameter for Y_{T-1} itself is estimated with bias (due to still two missing terms). Table 4 shows that when only the

MA term is omitted, this is not very harmful, except for the parameter for Y_{T-1} and δ_3 when the true α is small.

Table 5 reports on the out-of-sample forecast accuracy, when measured with the root mean squared prediction error (RMSPE) for the three models. Clearly, only omitting the MA terms gives the smallest deterioration, while omitting Y_{T-1} is disastrous.

4. An illustration

In this section, I provide an empirical illustration using data on the registrations of new motor vehicles in the Netherlands. Figure 1 presents the total amount of new motor vehicles of all types, while Figure 2 additionally presents the new registrations for passenger cars only, for each of the three months within a quarter.

The first MIDAS regression for these data, explaining total vehicles by lags and the monthly passenger cars is (7). The p value of the Wald test for the restrictions $\beta_6^* = \beta_7^* = \beta_8^* = 0$, $\beta_7^* = \beta_8^* = 0$, and $\beta_8^* = 0$ are 0.061, 0.516 and 0.137, respectively, while the t value for $\beta_5^* = 0$ is -7.281, and hence the final main MIDAS model is

$$Y_T = \rho Y_{T-1} +$$
$$\beta_0^* X_{3,T} + \beta_1^* X_{2,T} + \beta_2^* X_{1,T} + \beta_3^* X_{3,T-1} + \beta_4^* X_{2,T-1} + \beta_5^* X_{1,T-1} + v_T + \theta v_{T-1}$$

The estimated parameters are presented in Table 6. The parameter for the MA term is not significant, and hence panel 2 of Table 6 presents the estimation results in case this term is ignored. Clearly, the change in the in-sample RMSPE is small, when this term is ignored. And, the estimated parameters for the explanatory variable do not show any systematic pattern. The first-order autoregressive parameter is significantly different from 0 and also from 1. Excluding the first-order autoregressive term leads to biased estimates, and also to a significant drop in in-sample forecast accuracy, as is evident from the final panel of Table 6.

5. Conclusion

This paper proposed to relate explicitly a true high frequency data generating process, where also the dependent variable is observed (hypothetically) at the high frequency, with a MIDAS regression. It was shown that a corresponding correctly specified MIDAS regression includes lagged dependent variables, a substantial number of explanatory variables (observable at the low frequency) and a moving average term. Next, the parameters of the explanatory variables unlikely obey certain convenient patterns, and hence imposing such restrictions in practice is not recommended.

The analysis reveals a few practical guidelines when MIDAS regressions are considered for actual data. First, it seems helpful to first speculate about the form of a potentially underlying DGP at a high frequency. This may help to decide on the number of lags to include in the MIDAS regression. Second, it is recommended always to include a lagged dependent variable and a moving average term. Third, it is not recommended to impose parameter restrictions on the explanatory variables, but simply to estimate the parameters without restrictions. Even in small samples, when the dynamic specification is (approximately) correct, it should be possible to estimate the parameters without (much) bias. Fourth, for practical purposes, it helps to read the data in an explicit low-frequency format, making use of the insights from the literature on periodic time series models.

Further work in this area concerns extensions to MIDAS regressions with more than a single explanatory variable and to MIDAS regressions involving more than a single dependent variable. A particular challenge shall be MIDAS regressions with cointegration.

Table 1: Parameter values in the low frequency DGP as in (4) for an ADL(3,3) model as in (1)

Variable	$\alpha_2 = \alpha_3 = 0$	$\alpha_1 = \alpha_2 = \alpha_3 = 0$
$X_{3,T}$	β_0	β_0
$X_{2,T}$	$\beta_0(1+\alpha_1)+\beta_1$	$\beta_0+\beta_1$
$X_{1,T}$	$\beta_0(1+\alpha_1+\alpha_1^2)+\beta_1(1+\alpha_1)$	$\beta_0+\beta_1$
$X_{3,T-1}$	$\beta_1(1+\alpha_1+\alpha_1^2)+\beta_2(1+\alpha_1)+\beta_3$	$\beta_1+\beta_2+\beta_3$
$X_{2,T-1}$	$\beta_2(1+\alpha_1+\alpha_1^2)+\beta_3(1+\alpha_1)$	$\beta_2+\beta_3$
$X_{1,T-1}$	$\beta_3(1+\alpha_1+\alpha_1^2)$	β_3

Table 2: Simulation results based on a sample of N “yearly” observations, when a “half-yearly” DGP is the true process with $x_t \sim N(1,1)$, $\varepsilon_t \sim N(0,1)$, $y_0 \sim N(0,1)$, DGP is $y_t = \alpha y_{t-1} + x_t + 2x_{t-1} + \varepsilon_t$, 1000 replications. The MIDAS regression is

$$Y_T = \mu + \delta_0 X_{2,T} + \delta_1 X_{1,T} + \delta_2 X_{2,T-1} + u_T$$

		δ_0		δ_1		δ_2	
α	N	mean	st.d.	mean	st.d.	mean	st.d.
True		1		3.5		3.5	
0.5	40	0.937	0.505	3.447	0.498	3.732	0.522
	400	0.996	0.145	3.493	0.154	3.745	0.152
	40000	1.001	0.014	3.500	0.015	3.750	0.015
True		1		3.8		4.4	
0.8	40	0.690	1.293	3.530	1.313	4.854	1.342
	400	0.957	0.371	3.754	0.381	5.006	0.386
	40000	1.000	0.036	3.799	0.038	5.040	0.036
True		1		3.9		4.7	
0.9	40	0.368	2.606	3.250	2.618	5.047	2.615
	400	0.898	0.634	3.803	0.639	5.421	0.647
	40000	1.000	0.059	3.898	0.062	5.509	0.059
True		1		3.95		4.85	
0.95	40	-0.069	5.554	2.665	5.463	4.887	5.480
	400	0.808	1.095	3.773	1.080	5.574	1.104
	40000	1.001	0.089	3.947	0.095	5.752	0.090

Table 3: Simulation results based on a sample of N “yearly” observations, when a “half-yearly” DGP is the true process with $x_t \sim N(1,1)$, $\varepsilon_t \sim N(0,1)$, $y_0 \sim N(0,1)$, DGP is $y_t = \alpha y_{t-1} + x_t + 2x_{t-1} + \varepsilon_t$, 1000 replications. The MIDAS regression is

$$Y_T = \mu + \gamma Y_{T-1} + \delta_0 X_{2,T} + \delta_1 X_{1,T} + \delta_2 X_{2,T-1} + u_T$$

		γ		δ_0		δ_1		δ_2	
α	N	mean	std	mean	std	mean	std	mean	std
True		0.25		1		3.5		3.5	
0.5	40	0.355	0.056	1.024	0.361	3.495	0.356	3.424	0.367
	400	0.363	0.016	1.007	0.099	3.502	0.104	3.389	0.105
	40000	0.364	0.002	1.001	0.010	3.500	0.010	3.386	0.010
True		0.64		1		3.8		4.4	
0.8	40	0.691	0.043	1.031	0.474	3.772	0.470	4.395	0.476
	400	0.710	0.012	1.011	0.129	3.801	0.137	4.335	0.136
	40000	0.712	0.001	1.001	0.013	3.800	0.013	4.328	0.013
True		0.81		1		3.9		4.7	
0.9	40	0.824	0.030	1.018	0.519	3.842	0.513	4.723	0.523
	400	0.846	0.009	1.011	0.143	3.900	0.151	4.669	0.149
	40000	0.851	0.001	1.001	0.014	3.900	0.015	4.660	0.015
True		0.9025		1		3.95		4.85	
0.95	40	0.904	0.017	1.003	0.538	3.864	0.530	4.875	0.545
	400	0.917	0.006	1.010	0.150	3.947	0.158	4.840	0.156
	40000	0.924	0.001	1.001	0.015	3.950	0.015	4.830	0.015

Table 4: Simulation results based on a sample of N "yearly" observations, when a "half-yearly" DGP is the true process with $x_t \sim N(1,1)$, $\varepsilon_t \sim N(0,1)$, $y_0 \sim N(0,1)$, DGP is $y_t = \alpha y_{t-1} + x_t + 2x_{t-1} + \varepsilon_t$, 1000 replications. The MIDAS regression is

$$Y_T = \mu + \gamma Y_{T-1} + \delta_0 X_{2,T} + \delta_1 X_{1,T} + \delta_2 X_{2,T-1} + \delta_3 X_{1,T-1} + u_T$$

		γ		δ_0		δ_1		δ_2		δ_3	
α	N	mean	std	mean	std	mean	std	mean	std	mean	std
True		0.25		1		3.5		3.5		1	
0.5	40	0.263	0.070	1.010	0.345	3.512	0.329	3.500	0.346	0.966	0.418
	400	0.270	0.020	1.005	0.093	3.502	0.096	3.480	0.095	0.933	0.116
	40000	0.272	0.002	1.001	0.009	3.500	0.010	3.478	0.009	0.923	0.011
True		0.64		1		3.8		4.4		1.6	
0.8	40	0.641	0.045	1.008	0.409	3.813	0.387	4.417	0.403	1.611	0.423
	400	0.648	0.013	1.007	0.110	3.802	0.114	4.392	0.112	1.572	0.122
	40000	0.650	0.001	1.001	0.011	3.800	0.011	4.390	0.011	1.562	0.012
True		0.81		1		3.9		4.7		1.8	
0.9	40	0.808	0.028	1.004	0.428	3.905	0.407	4.720	0.423	1.819	0.425
	400	0.814	0.010	1.008	0.116	3.902	0.120	4.697	0.118	1.791	0.125
	40000	0.815	0.001	1.001	0.012	3.900	0.012	4.696	0.012	1.780	0.012
True		0.9025		1		3.95		4.85		1.9	
0.95	40	0.901	0.014	1.002	0.435	3.950	0.419	4.871	0.433	1.913	0.429
	400	0.904	0.006	1.008	0.119	3.952	0.123	4.849	0.122	1.900	0.125
	40000	0.905	0.001	1.001	0.012	3.950	0.012	4.848	0.012	1.890	0.012

Table 5: Root mean squared prediction errors for 10, 100 and 10000 observations out-of-sample and forecast accuracy of the correctly specified high frequency DGP, that is, including the true $Y_{2,T-1}$.

		MIDAS regressions in			True DGP
		Table 2	Table 3	Table 4	
α	N				
0.5	40	1.101	0.573	0.605	0.519
	400	0.346	0.170	0.184	0.162
	40000	0.036	0.017	0.018	0.016
0.8	40	4.111	0.790	0.761	0.617
	400	1.389	0.208	0.222	0.186
	40000	0.145	0.021	0.022	0.018
0.9	40	9.723	0.937	0.828	0.664
	400	3.291	0.228	0.237	0.196
	40000	0.344	0.022	0.023	0.019
0.95	40	27.345	1.051	0.873	0.701
	400	7.262	0.251	0.247	0.202
	40000	0.747	0.023	0.024	0.020

Table 6: Estimated parameters in MIDAS regressions for new registrations of motor vehicles

Variable	Full MIDAS		Without MA		Without MA and Y_{T-1}	
Intercept	-2692	(4100)	-2319	(3769)	4312	(4738)
Y_{T-1}	0.672	(0.162)	0.685	(0.140)		
$X_{3,T}$	1.057	(0.049)	1.055	(0.049)	1.034	(0.065)
$X_{2,T}$	0.983	(0.179)	0.999	(0.176)	1.160	(0.233)
$X_{1,T}$	1.237	(0.131)	1.216	(0.116)	0.944	(0.136)
$X_{3,T-1}$	-0.844	(0.183)	-0.868	(0.140)	-0.306	(0.106)
$X_{2,T-1}$	-0.528	(0.295)	-0.533	(0.282)	0.572	(0.226)
$X_{1,T-1}$	-0.696	(0.199)	-0.712	(0.172)	0.060	(0.090)
MA	0.071	(0.282)				
RMSE	2838.5		2840.7		3897.9	

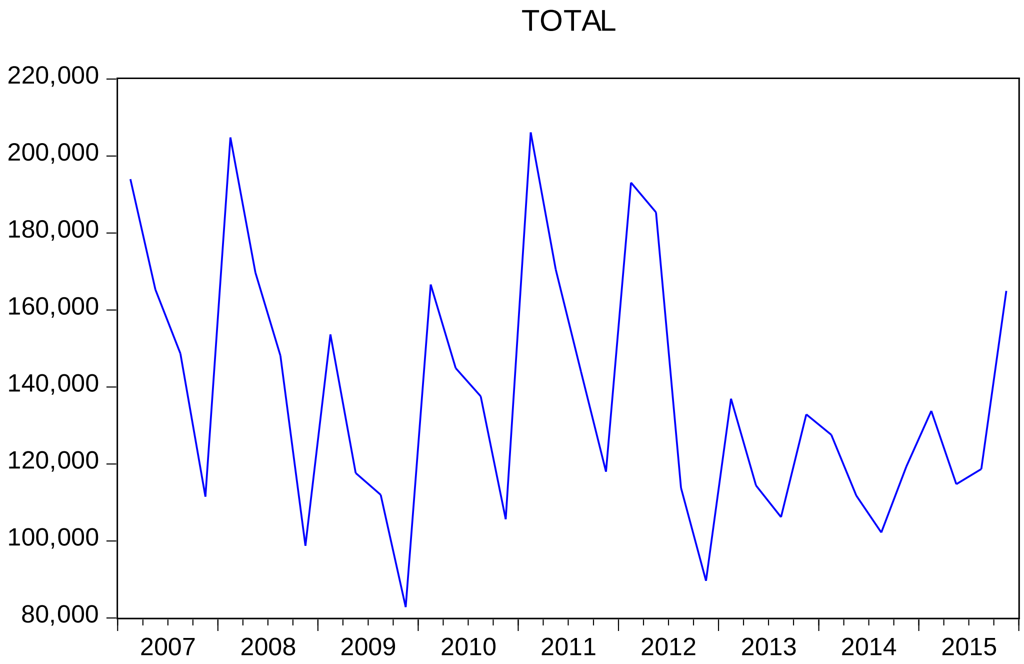

Figure 1: Total new registrations all vehicles in the Netherlands, 2007Q1-2015Q4

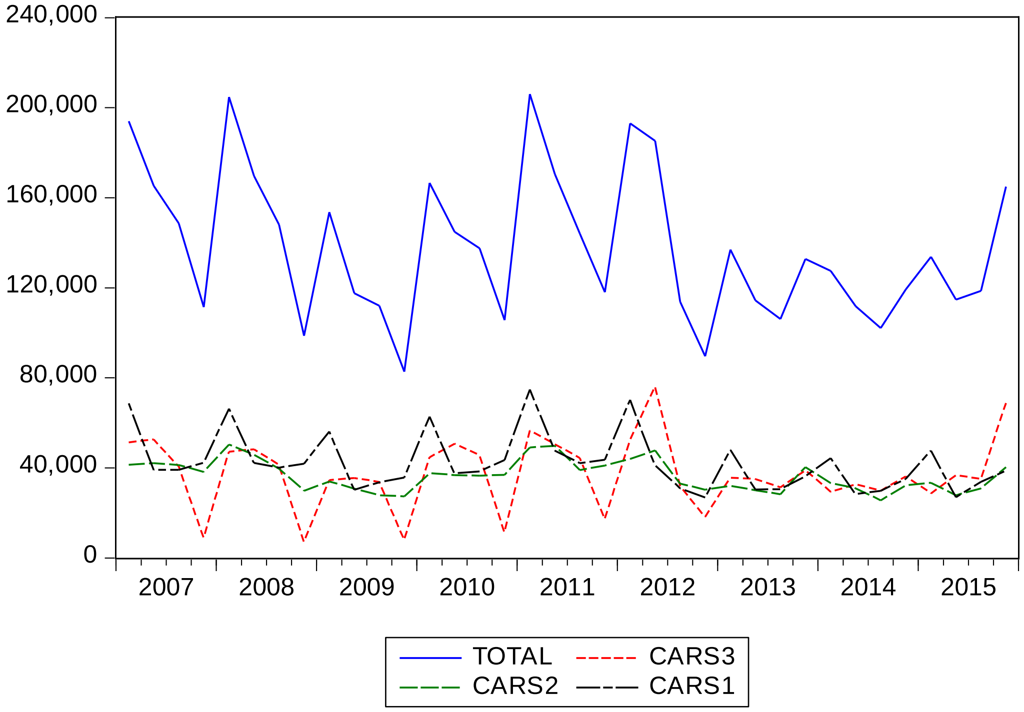

Figure 2: Total new registrations of all vehicles in the Netherlands, 2007Q1-2015Q4, and new registrations of new passenger cars in months 3, 2 and 1 of the quarters

References

Andreou, E., E. Ghysels and A. Kourtellos (2010), Regression models with mixed sampling frequencies, *Journal of Econometrics*, 158, 246-261.

Clements, M. and A.B. Galvao (2008), Macroeconomic forecasting with mixed-frequency data, *Journal of Business and Economic Statistics*, 26, 546-554

Franses, P.H. (1994), A multivariate approach to modelling univariate seasonal time series, *Journal of Econometrics*, 63, 55-63

Franses, P.H. and R. Paap (2004), *Periodic Time Series Models*, Oxford: Oxford University Press.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004), The MIDAS touch: Mixed data sampling regression models, CIRANO Working paper 2004s-20.

Ghysels, E., A. Sinko, and R. Valkanov (2007), Midas regressions: further results and new directions, *Econometric Reviews*, 26, 53-90.

Lütkepohl, H. (1986), Forecasting temporally aggregated vector ARMA processes, *Journal of Forecasting*, 5, 85-95.

Marcellino, M. (1999), Some consequences of temporal aggregation in empirical analysis, *Journal of Business and Economic Statistics*, 17, 129-136.