IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN: 2278-487X, p-ISSN: 2319-7668. Volume 22, Issue 10. Ser. IV (October 2020), PP 28-44
www.iosrjournals.org

Mixed Data Sampling Modelling (MIDAS): Application to the forecasting of French economicgrowth rates

Paul ROSELE CHIM[1]
Hisseine Saad MAHAMAT[2]
BETA EMADD BIO MINEA UR 7485 University of French Guiana
October - 2020

Abstract
The short-termanalysis of a country'seconomy in a context of globalisation and interdependenceis a verydelicate and complexexercise for the management of economic and monetarypolicy. Indeed, managers and politicaldecision-makersscrutinize the economic conditions of the moment, make anticipations, and adapttheirgovernanceaccordingly. Thus, the evolution of the parametersthat influence the rate of growth of the Gross Domestic Product (GDP) fuels passions and animatesdebates.
Time seriesfrom the real and financialeconomy do not have the samecharacteristics, both in terms of their sampling frequency and theirpredictive contribution. This raises questions about the use of thesedata:
-Which temporal aggregationis the mostrelevant?
-Whatindicatorsshouldbeconsidered?
-What model can beconstructedfor a givenestimate, at what temporal frequency?
The purpose of ourstudyis to answerthese questions by evokingfundamentalelements of econometric estimation in order to discern the problems and issues at stake. Weattempt a new model construction and applyit to the GDP data of the French economy, the unemployment rate and the CAC 40 stock market index from the 1st quarter of 2010 to the 3rd quarter of 2012.

Date of Submission: 30-09-2020 Date of Acceptance: 13-10-2020

I. Introduction

The global financialcrisis, the sovereign debtcrisis and the recessions that are stillongoing in some countries, evenamong the mostdeveloped, are evidence of the difficulty of anticipatingeconomic fluctuations, even in the near future. However, short-termanalysis of the national economy in a context of globalisation and interdependenceis as complex as itis essential for the definition of contemporaryeconomic and monetarypolicy. Indeed, economists, politicians, bankers, journalists, citizens, employees and employers, consumers, producers and investors all scrutinizecurrent, anticipated, hoped-for, predicted or forecasteconomic conditions and adapttheirbehaviour, policies and decisionsaccordingly. For example, the quarterly publication of Gross Domestic Product (GDP) growth rate figures, whichrepresent the evolution of the overall value addedthat an economyproduces over a certain period of time, as defined by the national accounts, stirs passions and animatesdebates. Although GDP has been the subject of criticism, itisnowadays the preferredindicator of a country'seconomichealth and as suchis of prime interest to economists and forecasters. Econometricstudies (estimates) must thereforebebased on a coherentmechanismthatmeasurescurrentcyclical conditions and the cyclical and systemic component of the éléments it mobilises. The available data on which an estimation analysisisbased have never been more important. Statistics on industry, employment, opinion surveys, commodityprices, shares, stock market indices, bond indices quoted in quasi-continuous time, real estatemarketindicators, the number of registrations, the unemployment rate, are all explanatory variables and potentialestimators of a country'seconomicgrowth.

Growthisdefined as the rate of change of a country's real GDP. It isgenerallytakenfrom the national accounts and calculated by the statisticalinstitute of the country concerned. It istraditionallyannounced on a quarterly basis. In France, the figure published by INSEE is an accountingresultbased on data on consumption, investment, changes in inventories, exports and imports, representing the production of value addedduring the period. Its publication isdelayed in relation to the quarter in question and is, moreover, subject to successive revisions, not delivering a definitiveresultuntilseveralyearslater. In France, the growth figure isknown about a month and a halfafter the end of the quarter in question (e.g. mid-May for the 1st quarter). This time lag has, in

[1] HDR University of Paris 1-Panthéon Sorbonne
[2]DoctorTeacher-Researcher

particular, made it possible to identify the French recessionthatbegan in March 2008 onlyfromNovember 2008. This shows the need to accuratelyanticipate fluctuations over a very short period of time. It is no longer a question of forecasting the future period but rather the currentperiod. Indeed, the delays in publication and the successive revisions of economicseries have evenforcedforecasters to consider prospective analyses at intra-period horizons, for example:forecastinggrowthfrom the first quarter to January. Suchmodellingisdefined in such a way as to mobilise the contemporary information available. It shouldbenotedthat the data, from the threefamilieswe have describedabove, fromwhichwewish to construct a predictivemethodology are certainlynumerous and probably informative, but have verydifferent sampling frequencies. Their use thereforerequires the development of adapted multi-frequencymodels. However, thisparticular temporal pattern should not represent an obstacle to modelling but rather one of itsfundamentalcharacteristics. Adoptingthistemporalityis a real challenge for economists and a challenge for ourresearchwork.

However, we note that time seriesfrom the real and financialeconomy do not have the samecharacteristics, both in terms of their sampling frequency and theirpredictive contribution. This raises questions about the use of thesedata:which temporal aggregationis the most relevant? Whichindicatorsshouldbeconsidered?What model shouldbeconstructed for a givenestimate, at what temporal frequency?

First, wewillevoke the fundamental éléments thatalloweconomic estimation and wewilldiscern the real problems and stakes of the study

In a second step, wewill propose a new methodology for model construction thatseeksparsimony of the variables thatconstituteitwithempirical performance. Weapplyit to French GDP data, the unemployment rate and the CAC40 stock market index from Q1 2010 to Q3 2012.

1. On the fundamentals of economic estimation and the issues at stake

1.1 Multi-periodicitymodels

Numerousmacroeconomicseries are available for the conjoncturist, but not necessarily on the same sampling frequency (or periodicity). In particular, the national accountsor the growth index (GDP), whichmosteconomiststry to forecast, are onlyavailable on a quarterly basis, whereasmanycyclicalindicatorssuch as the industrial production index, householdconsumptionexpenditure, the unemployment rate or opinion surveys are available on a monthly basis. To manage thesetwoperiodicitiessimultaneously in a model, Mariano and Murasawa (2003) proposed a dynamic factor model, put in a state-spaceform, whichconsidersquarterlyseries as monthlyseriescontainingmissing values. The ideais to try to estimate a factor common to N variables, some of which are quarterly and some of which are monthly.

The model we propose is a classicallinearregressionexceptthat the incorporated variables are of differentfrequencies, i.e. our time series of interest (French GDP) isobserved at lowfrequency (quarterly) and the explanatory a contemporaryway. Weproceed as follows: to explain a quarterly data (for example the first quarter available at variables (the unemployment rate and the CAC40) are sampled at high frequency (monthly and daily). We first reason in the end of March), weconsider for example the last 4 data of a monthly real variable (March, February, January and December), and the last 6 months of data of a dailyfinancial variable (at least 20×6 = 120 daily data fromOctober to March). The first ideawouldbe to weighteach of these values by a coefficient thatwewouldestimate. This modeling is not feasible for a large-scaleproblem:using the twopreviousexplanatory variables wouldimply the estimation of at least 124 parameters. This is a recurringproblemwithfinitesamples. The modellingwe propose seeks to reconcile the mixing of sampling frequencies and the parsimonynecessary for its estimation.

1.2 The MIDAS regression

The modellingwe propose seeks to reconcile the mixing of sampling frequencieswith the parsimonyrequired for its estimation. Ghysels (2002) and hisco-authorsdeveloped the MIDAS (Mixed Data Sampling) regression model. MIDAS aims to accommodate time aggregationusing a specific class of time seriesmodelsthatinvolvesparsimony and flexibility. Derivedfrom the technique of staggereddelaymodels, this new econometrictoolisbased on both a regression structure and a weightingfunctionthatfollows the high frequency of delays of the explanatory variables.

In the samecontext as the equation:$\mathrm{Y_t} = \sum_{k=1}^{K-1} w_k x_{t-k}^k + \varepsilon_t$)so the MIDAS aims to explain$\mathrm{Y_t}$ using the delays of the explanatory variablex_t^k sampled at frequency tk, it can bewritten as follows:

$\mathrm{Y_t} = \beta_0 + \beta_1 m_k(\theta, \mathrm{L})\mathrm{X_t^k} + \varepsilon_t$ (1.2.1)

We notice that the MIDAS model combines the usuallinearregressioncharacteristicswith an aggregation structure defined by the functionm_k. Ghysels et al (2002) and Ghysels et al (2007) have shownthat the constant $\beta_0$and the coefficient $\beta_1$maycontainssomuchfull empiricalinterpretations. The termε_t represents the residuals. We will now focus on the $\mathrm{m_k}$function of our MIDAS model, itsspecification and estimation.

1.3 Almonfunction and the weighting system

The kernel function$m_k$isprecisewith respect to the parameterθ and the past values of X_t^k.
Wedefine the delayoperator as$L^k x_t^k = L^k x_{tk} = x_{t/_{k-K}}$. The number of delays K is exogenous; as discussed in previous sections, the choice of K can bestatisticallytested or empiricallyevaluated. The parameters$\{\beta_0, \beta_1 \text{and } \theta\}$ are estimated (the technicals of estimation willbediscussedbelow). However, it can benotedthat the presence of the coefficient β_1 implies that the function m_k provides normalizedweights for past K values ofx_t.
Wedefine:

$m_k(\theta, L) = \sum_{k=0}^{K-1} \frac{\varphi(k,\theta)}{\sum_{l=0}^{K-1}\varphi(l,\theta)} L^k$ The weightfunction(1.2.2)

This expression of the weightfunctionis the commonform of the MIDAS as it has been popularized over the last decade. Manyparameters of thisweightingfunction have been proposed as a function of the number of coefficients or the form of the function. Models for mixed-frequency data wererecentlyreviewed by Foroni and Marcellino (2013b). Note that the expression Almon, which combines equations (1.2.1) with the Almonform, is a special case of MIDAS which can bewritten as follows :
$\beta_1 m_k(\theta, L) = \sum_{k=0}^{K-1}\left(\sum_j \theta_k K^j\right) L^k$ (1.2.3)
The recent MIDAS literatureinitiated by Ghysels et al (2002) preferred the non-linear expression for the weightfunction, itmainly comprises twoforms: the delayed Beta and the exponentialdelayfunctionAlmon. These are definedbelow :

The normalizedprobabilitydensity beta functiondefined as follows:

$$\varphi(k,\theta) = \varphi_k(k,\theta_1,\theta_2) = \frac{\frac{k^{\theta_1 - 1}}{k}\left(1-\frac{k}{K}\right)^{\theta_2 - 1}\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} \qquad (1.2.4)$$

Where$\Gamma(\theta) = \int_0^\infty e^{-x} x^{\theta-1} dx$, the size of the polynomial p is defined with respect to both regression performance and parsimony. Let us note that the Beta form of the latent variable allows interestingcharacteristicsaccording to itsspecifications. For example, by limiting the size of the argument of function (1.2.2) to a single parameterθ_1 is able to impose the decrease of the weight values. This weighting system which incorporates a single hyper parameterθ_1 is of the form :

$\varphi(k,\theta) = \varphi_k(k,\theta_1) = \theta_1(1-k)^{\theta_1 - 1}$ (1.2.5)

In terms of economicinterpretation, assets sloping (decreasing) downwardsmaybe a desirablefeatureespecially in amulti-stage direct forecasting configuration.

Anotherpopular expression of the MIDAS weightfunctionisAlmon'sexponential as a lag, which can bewritten as follows:

$$\varphi(k,\theta) = \varphi_k(k,\theta_1,\dots,\theta_p) = exp\left(\sum_{j=1}^{p}\theta_j\, k^j\right) \qquad (1.2.6)$$

This formula isderivedfrom the Almonfunction in a straight direction. The use of the exponentialfunction forces the weight to be positive (Le Juge et al. 1985). Almon'sexponentialfunctionisspecified in the literature, withtwoparameters, that's to say(p = 2 in equation (1.4)) :

$$\varphi(k,\theta) = \varphi_k(k,\theta_1,\theta_2) = \exp(\theta_1 k + \theta_2 k^2)$$

Thesetwoformsprovide a flexible and parsimonious data-basedweighting system thatinvolves a small set of parameters and isthereforewellsuited to smallsamples. Table 1.2 presents the forms of Almon'sexponential latent weightfunction; the choice of thesetwoparameters$\theta = (\theta_1 + \theta_2)$.

	weights	lags	starts
1	nealmon	0:5	c(1, -1)
2	nealmon	0:6	c(1, -1)
3	nealmon	0:7	c(1, -1)
4	nealmon	0:8	c(1, -1)
5	nealmon	0:9	c(1, -1)
6	nealmon	0:10	c(1, -1)

7	almonp0:5		c(1, 0, 0)
8	almonp0:6		c(1, 0, 0)
9	almonp0:7		c(1, 0, 0)
10	almonp	0:8	c(1, 0, 0)
11	almonp	0:9	c(1, 0, 0)
12	almonp	0:10	c(1, 0, 0)

Weightdelaydefined by Almon'sexponential latent

Kvedaras and Zemlys (2012) proposed a test for assessing the statisticalacceptability of a functionalconstraintthatisimposed on the MIDAS regressionparameters.
Afterdetermining the number of lags, itisnow a matter of estimating the $\beta_i$coefficients of the model. To avoidhavingerroneousresultswhenusing OLS due to the multicollinearitybetween the variables, weopt for the Almonmethodwhichallows us to minimize the number of parameters to beestimated.
The distribution of the coefficients can takeseveralforms, the Almon lag methodallows us to identify lag profiles that fit differentrepresentations, henceitiswellknown and widelyused. This technique consists in imposing on the coefficients to belong to the same polynomial of degree m, such as :
$y_t = \sum_{j=0}^{m} \delta_j x_{t-j} + u_t$ (1.2.7)
Wecannotestimatethis model directlybecause of multicollinearityproblems. Wewill assume that the$m +$ 1 parameters are constrained by a polynomial in j, of degreen < m (in generaln = 2, 3), whichimplies :

$$\delta_j = \sum_{i=0}^{n} \gamma_i j^i = p(j)$$

We have changed the previous notifications of the settings to avoidoverlap. Wethen replace the δ_j parameterswiththeir value in the :
$y_t = \sum_{j=0}^{m} (\sum_{i=0}^{n} \gamma_i j^i) x_{t-j} + u_t = \sum_{i=0}^{n} \omega_{it} \gamma_i + u_t$(1.2.8)
With$\omega_{it} = \sum_{j=0}^{m} j^i x_{t-j}$
Wethereforeconstruct a set of new exogenous variables by simple transformation. If W is the matrix of thesetransformedexogenes, the model isnoted as follows:
$y = W\gamma + u$ (1.2.9)
And weestimate the $n + 1$parameters γ of this model by OLS. We can thenretrieve the delayparametersusing the polynomial.
How to determine the degree$m$ of the polynomial to beused?
Well, we are going to compare an unconstrainedregression model which has$n$ lags and thus $n + 1$ regressors to a model where the coefficients of the n lags are constrained by a polynomial of degree$m$. For this we use an F test.
Let us call SSE the sum of the squares of the residuals for the unconstrained model and SSE_m the sum of the squares of the residuals for the model with an Almon polynomial of degree$m$.
Thus, to test the restriction, wejustneed to compute :

$$F = \frac{(SSE_m - SSE)/(m-n)}{SSE/(T-m-1)}$$

This statisticisdistributedaccording to a law in $F(m-n, T-m-1)$.
When the value of the number of delaysisunknown, there are statisticalcriteria to defineit. To do thiswe can use severalmethods, wewilllimitourselves to the presentation of three of them, namely : Fisher's test, Akaike'smethod and Schwarz'smethod. (See appendices).
The use of staggereddelaymodelsisverywidespread in the economicfield, and thisis due to the existence of manyquantitiesthat can beexplained by exogenous variables spread over time. The best illustration isundoubtedly the financialmarket in whichproductprices are closelydependenton values takenpreviously.

1.4 The NLS MIDAS estimator

We assume that the white noise term$\varepsilon_t$isnormallydistributedwith a densitygiven by :
$f(\varepsilon_t) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{{\varepsilon_t}^2}{2\sigma^2}\right)$ (1.3.0)
Nowwe note $\emptyset$ the family of unknownparameters, that's to say$\emptyset = \{\beta_0, \beta_1, \theta, \sigma\}$and wedefine $X_t(\emptyset) = X_t(\emptyset, x_t) = \beta_0, +\beta_1 m_k(\theta, L) x_t^k + \varepsilon_t$. Assumingthat the sample size isT, for

anyt = 0, …, T the conditionalprobability distribution of $y_t$isgiven by :

$$f(y_t|x_t;\emptyset) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\sum_{t=1}^{T}\frac{y_t-(\beta_0+\beta_1 m_k(\theta,L))X_t}{2\sigma^2}\right)$$
$$=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\sum_{t=1}^{T}\frac{y_t-X_t(\emptyset)}{2\sigma^2}\right)(1.3.1)$$

The log-likelihoodfunction can bewritten as follows:

$Lnf(y|\emptyset) = \sum_{t=1}^{T} Lnf(y|x_t;\emptyset) = \frac{1}{2}Ln2\pi - \frac{T}{2}Ln\sigma^2 - \frac{T}{2\sigma^2}\sum_{t=1}^{T}\left(y_t - X_t(\emptyset)\right)^2$(1.3.2)

Whichismaximized in relation to $\emptyset$.

However, in the context of the non-linearregression model, we can notice that the maximization or minimizationproblems (such as the Newton-Raphsonalgorithm) are simplified by expressing$\widehat{\sigma^2}$ as a function of $\hat{\beta}$ and $\hat{\theta}$. This isachieved by solving the first-order condition for pour $\widehat{\sigma^2}$which is the solution:

$\widehat{\sigma^2} = \frac{1}{T}\sum_{t=1}^{T}\left(y_t - X_t(\widehat{\emptyset})\right)^2$(1.3.3)

Thus by maximizing, the log-likelihoodfunction leads to the redefinition of the unknownparametervector$\emptyset = \{\beta_0, \beta_1, \theta\} \equiv \{\beta, \theta\}$.

This probabilityismaximizedwhen the sum of the residual squares $S(\emptyset) = \left(y_t - X_t(\emptyset)\right)^2$isminimized :

$\widehat{\emptyset} = \arg\min_{\emptyset} S(\emptyset)$ (1.3.4)

Next, the differentiation$S(\emptyset)$,

$$\frac{\partial S(\emptyset)}{\partial\emptyset} = \frac{\partial\left(y_t - X_t(\emptyset)\right)^2}{\partial\emptyset} = -2\left(y_t - X_t(\emptyset)\right)^2\frac{\partial\left(y_t - X_t(\emptyset)\right)}{\partial\emptyset} \qquad (1.3.5)$$

Setting the partial derivativesequal to 0gives the equationsthatdetermine the regression coefficients. There is no closed-form solution to the non-ordinary least squares problem. We use numericalalgorithmsinstead of findingparametersthatminimize the value (1.2.9). Nevertheless, the non-ordinary least squares estimator has asymptoticproperties. Assumingthat the decreasingterm of $\nabla X_t(\emptyset) = \left[\frac{\partial X_t(\emptyset)}{\partial\emptyset}\right]$ exists, a $\widehat{\emptyset}$estimator of non-ordinary least squares MIDAS is asymptotically normal.

$$\sqrt{T\left(\widehat{\emptyset}-\emptyset\right)} \xrightarrow{d} \mathcal{N}(0, \sigma^2 E[\nabla X_t(\emptyset)\nabla X_t(\emptyset)']^{-1}) \qquad (1.3.6)$$

This resultwasrigorouslyproven by Jennrich (1969). Werefer to Judge et al (1985) for more details in non-linearstatisticalmodels. In fact, MIDAS regressionmodels are generallyestimatedusing standard iterativeoptimization. The non-linearspecification φ requiresnumericaloptimization to determine solutions (case of the Levenberg-Marquardt algorithm or anyother gradient descentmethod).

Andreou et al (2010) studied the asymptoticproperties of the nonlinear least squares estimator MIDAS. Theyproposed to decompose the conditionalmean of the MIDAS regression to assess the consequences of temporal aggregation. Following their techniques, wederivefrom the MIDAS equation a sum of twoterms:aweight-basedaggregationterm and a non-lineartermthataggregatesorderdifferences of the high-frequency process :

$y_t = \beta X_t^l + \beta X_t^{nl}(\theta) + u_t$(1.3.7)

Where the first termis spread aggregation as defined as follows :

$X_t^l = \sum_{k=0}^{K-1}\frac{1}{K}x_{t-k}$. The second$X_t^{nl}$isdefined as the differencebetween the weight of the structure and the MIDAS weights, and thereforedepends on the hyperparameter.

It has the followingform:

$X_t^{nl}(\theta) = m_k(\theta, L)x_t - X_t^l = \sum_{k=0}^{K-1}\left(\frac{\varphi(k,\theta)}{\sum_{l=0}^{K-1}\varphi(l,\theta)}\right)x_{t-k}$(1.3.8)

The non-linearity of the Term$X_t^{nl}(\theta)$is due to the non-linearweightingmethod of the MIDAS regression model according to the form of the functionφ.

2. Study of the growthdynamics of the French economy

2.1 Empirical Application

The periodunderreview in thisstudyis the aftermath of the 2008-2009 financialcrisis. This period, whichwashoped to be the period of economicrecoveryfrom the recession, finallysaw the dawn of a new Europeancrisis. Not all signals are green, but according to INSEE the French economyisfinally on the road to recovery. The INSEE does not go so far as to saythat the government'spolicyisbeginning to produceitseffects, but experts point out that the reductions in the burden of the responsibilitypact and the competitiveness and employmenttaxcreditshould enable companies to recovertheirmargins, with the key to a resumption of investment by the end of 2016. This is a crucial point for business investment. It was the weaklink in the recovery. Moreover, INSEE forecaststhat in the comingyears, investmentcouldbecome the driving force behindgrowththat has so far been driven by householdconsumption. It isprecisely over this 2010-2012 periodthatwewillassessourmethodologyusing data on the three main parameters central to the national economy: GDP, the unemployment rate, and the CAC40 stock market index. For this estimation exercise, threemodelsbased on a MIDAS methodology are envisaged. Theysuccessively use real monthly variables and dailyfinancial variables. Theywillthus enable us to identify the anticipatoryfactors in each of thesesectors. The specification of thesemodelsis as follows:

The first model $midas^{M}$ considersonly the monthly variable of the so-called real economy (unemployment rate):

$$PIB^{T}_{t+h/t} = \alpha + \beta \, \mathrm{midas}^{M}(\theta)X^{M}_{t}$$

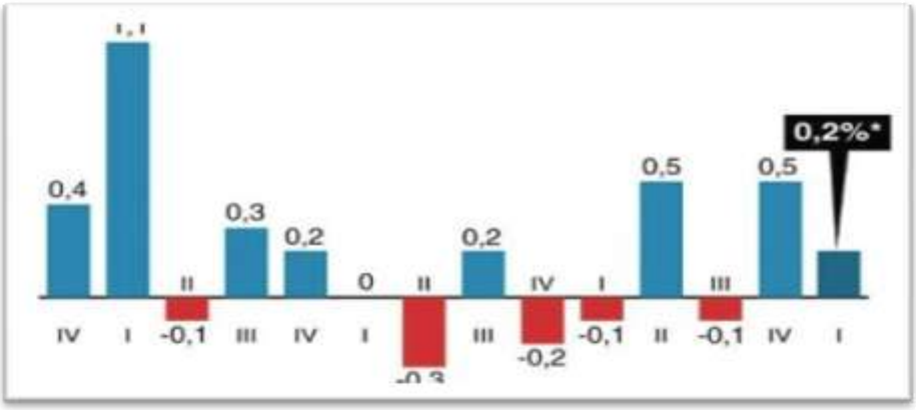
The second model $midas^{J}$ focuses on dailyfinancialvolatilities (CAC40 stock market indices) :

$$PIB^{T}_{t+h/t} = \alpha + \gamma \mathrm{midas}^{J}(\omega)X^{J}_{t}$$

Finally, the third model $midas^{MJ}$ intends to mix the twopreviousmodels by incorporatingmonthly real economicindicators and dailyfinancialvolatilitiessuch as :

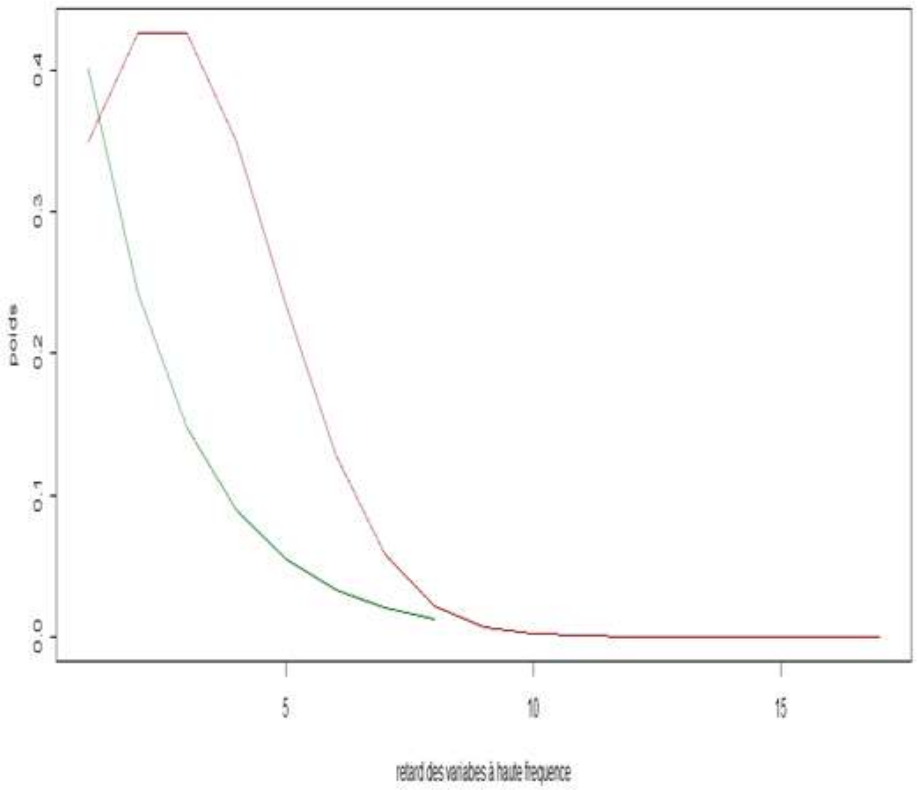
$$PIB^{T}_{t+h/t} = \alpha + \beta \mathrm{midas}^{M}(\theta)X^{M}_{t} + \gamma \mathrm{midas}^{J}(\omega)X^{J}_{t}$$

It shouldalsobenotedthat the size of the monthly or dailydatabasesis first a factorial principal component analysismethod. This so-called FAMIDAS (Factor Augmented MIDASreducedusing) modellingwasproposed by Marcellino and Schumacher (2010). The temporal nature isspecified by exponent of the variable (e.g. X^{M}_{t}) is the factor from the monthly real databaseconstructed by PCA thatrepresents the real economy). Finally, note that the estimation periods are definedrespectivelyfrom the first quarter of 2010 to the third quarter of 2012 (11 quarters).



French GDP quarterly trend Bank of France estimate for the 1st quarter 2014 (source: AFP)

Now suppose thatwe have (only) observations of Y; X; and z representing GDP, the unemployment rate, and the CAC40 stock index, respectively, which are stored as vectors, matrices, or time series. Our intention is to estimate MIDAS regressionmodels as in the equationabove:



Pacing of the impact of the explanatory variables, Red for the variable x, and Green for z.
It isinteresting to note that the impact of variable x can berepresentedusing the basic MIDAS characteristics, while the impact of z cannotbe possible to present.
In a first part wewillperform an estimation withoutrestricting the parameters as in U-MIDAS, as on the ordinary least squares (OLS) methodwhichgives us the followingresult :

Parameter :

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.9694327	0.1131838	17.400	< 2e-16	***
trend	0.1000072	0.0008035	124.467	< 2e-16	***
x1	0.5268124	0.0595920	8.840	2.99e-16	***
x2	0.3782006	0.0578522	6.537	4.24e-0	***
x3	0.1879689	0.0714227	2.632	0.009090	**
x4	-0.0052409	0.0631790	-0.083	0.933963	
x5	0.1504419	0.0671782	2.239	0.026118	*
x6	0.0104345	0.0644802	0.162	0.871591	
x7	0.0698753	0.0804935	0.868	0.386284	
x8	0.1463317	0.0729738	2.005	0.046149	*
z1	0.3671055	0.0664017	5.529	9.03e-08	***
z2	0.3502401	0.0598152	5.855	1.70e-08	***
z3	0.4514656	0.0617884	7.307	4.88e-12	***
z4	0.3733747	0.0579062	6.448	6.99e-10	***
z5	0.3609667	0.0677851	5.325	2.47e-07	***
z6	0.2155748	0.0589119	3.659	0.000316	***
z7	0.0648163	0.0608248	1.066	0.287752	
z8	0.0665581	0.0567170	1.174	0.241847	

z9	-0.0014853	0.0689694	-0.022	0.982837
z10	0.0466486	0.0802425	0.581	0.561598
z11	0.0384882	0.0761128	0.506	0.613588
z12	-0.0077722	0.0574461	-0.135	0.892501
z13	-0.0283221	0.0569727	-0.497	0.619598
z14	-0.0375062	0.0615205	-0.610	0.542715
z15	0.0297271	0.0587018	0.506	0.613072
z16	0.0184075	0.0588906	0.313	0.754900
z17	-0.0546460	0.0677653	-0.806	0.420875

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Residual standard error: 0.9383 on 222 degrees of freedom
- Multiple R-squared: 0.9855, Adjusted R-squared: 0.9838
- F-statistic: 579.3 on 26 and 222 DF, p-value: < 2.2e-16

Now, we are going to check if the explanatory variables used in our model, have the expectedsigns, and highlight their importance in the French growth (GDP).

Table (1) presents an estimation of OLSIM without instruction as in the OLS method, whiletakingintoaccount the frequencymultiplicity of the variables, the process has been estimated in sampleunder the monthly and daily lag basis respectively.

Moreover, the first coefficients of ourexplanatory variables are statisticallysignificant at the threshold of 1%, 5% and 10%, the resultsappearrelativelyinsensitive to the number of lags becausedespite the high standard deviation of error the criticalprobabilityassociatedwith the hypothesis test issatisfactory, thenwe have a coefficient of determination (R^2) which is also very high. In the light of our empirical developments, it emerges that the daily stock market index and the monthlyunemployment rate are explanatoryelements of France'squarterly GDP.

2.2 The Almonexponential polynomial withparameterconstraintusing the least non ordinary square (NLS)

As notedabove, the role of the Almon Polynomial in the MIDAS construction is to derivedelay profiles thatadapt to differentrepresentations. Thus, weperform the MIDAS estimation by integrating the Almonexponential polynomial withparameterconstraints (as in the nealmonfunction) using the sameprocedure as the least non ordinary square (NLS). We have the following output :

Parameter:

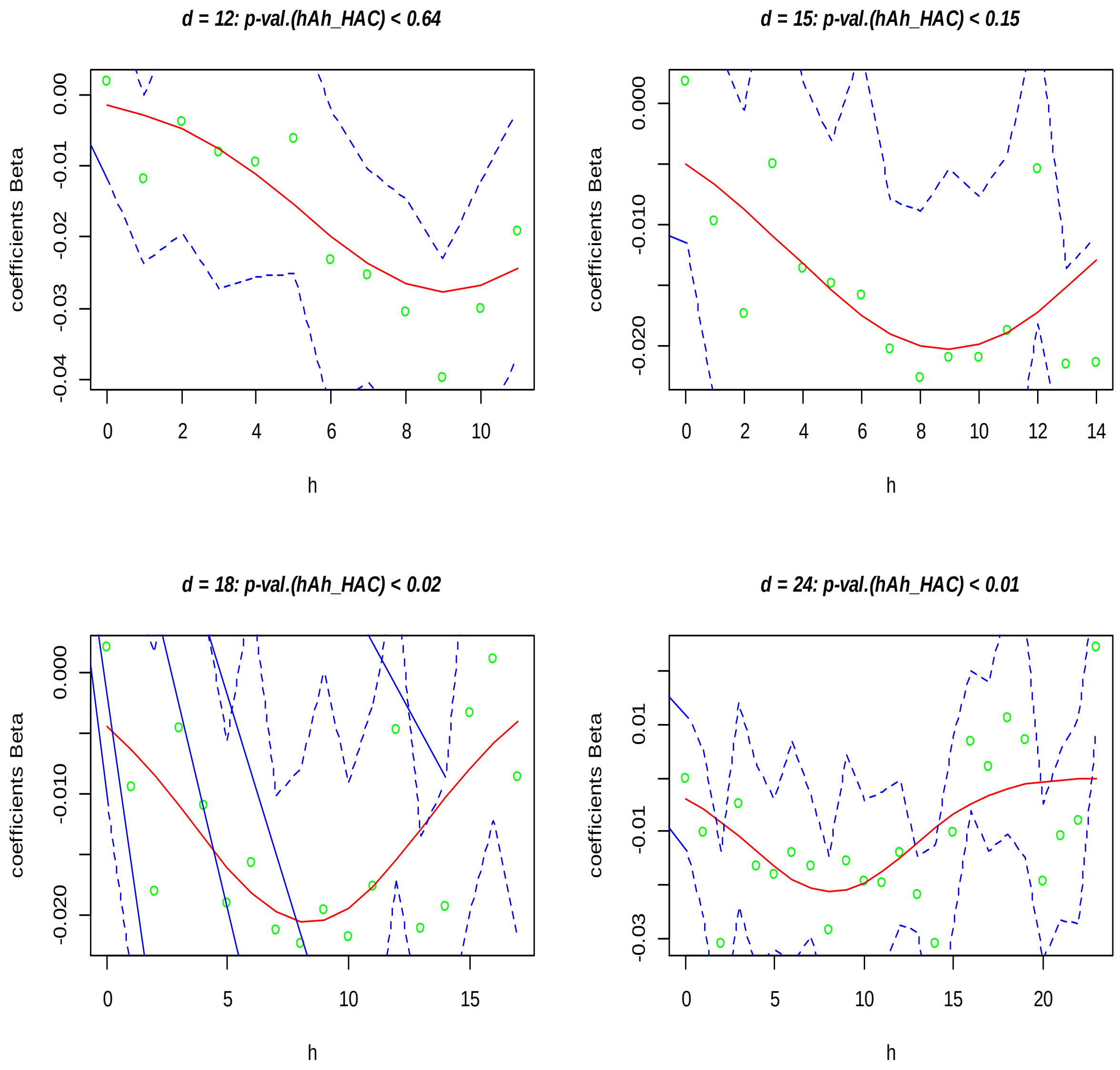
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.1502319	0.1292044	16.642	< 2e-16 ***
trend	0.0989839	0.0008573	115.460	< 2e-16 ***
x1	1.1408568	0.1720956	6.629	2.18e-10 ***
x2	-0.3308109	0.0828864	-3.991	8.72e-05 ***
z1	1.9339972	0.1937037	9.984	< 2e-16 ***
z2	0.8514051	0.3180375	2.677	0.00794 **
z3	-0.1530920	0.0495091	-3.092	0.00222 **

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*
Residual standard error: 0.9316 on 242 degrees of freedom

As we can see the syntax of the MIDAS functionissimilar to NLS. The delaysincluded and the functional restrictions used can beindividual for each variable. The table gives the estimatedparameters, their standard deviation and their t value. Weseethatonlysignificant coefficients are selected. Indeed, the statisticisvery high and exceeds the threshold value of a 5% test by a widemargin:Wethereforereject the hypothesis of global nullity. The hypothesis of homoscedasticityisthusverystronglyrejected. Wenowturn to the heteroscedasticity test whichaims at verifyingwhether the square of the residuals can beexplained by the explanatory variables of the model. If thisis the case, thereisheteroscedasticity. In thiscontext of the test, the nullhypothesisisthat all the coefficients of the regression of the squaredresiduals are null, in short, thereishomoscedasticity; the alternative hypothesisisthatthereisheteroscedasticity. Thus, in our case wereject the nullhypothesis (« t-value » <β), we can conclude to the presence of heteroskedasticity. This isexactly the test thatinterests us in this case. Sinceourregressionis non-linear, itseems to us that the heteroskedasticity test isessentiallyuseful for understanding the structure of the data.

2.3 Tests for the adequacy of restrictions

Given an estimated MIDAS regression model, weusedtwooptimizationmethods (techniques) to improve convergence, usingfunctionssuch as hAh.test and hAhr.test. Sinceourserie (Y_t) seriesisstationary and co-integratedwith the explanatory variables, bothmethods can beapplieddirectly, a special transformation has to beapplied as in the example of Bilinskas and Zemlys (2013). Bothmethods are alsousefulwhen the process errors are independently and identicallydistributed, then for the robustness of the test.



The hAhr.testmethod deals withcomputational technique explicitly for modelswith a large number of delays and werestrict the number of delays to bechosen. The Akaike Information Criterion (AIC) selected the model with 10, 14, 15 and 20 lags, for numbers of 12, 15, 18 and 24 significant coefficients respectively.

2.4 Model selection

Thus, wewillstudy the following on severalfunctionalconstraints, for example, the normalized ("nealmon") or non-normalized ("almon") exponentialpolynomials of Almondelay, or withpolynomials of degree 2 or 3, thus are adapted to a MIDAS regression model of variable y on x and z. Here, for each variable, weightsdefine the possible restrictions to betakenintoaccount and a list first gives the appropriatestarting values to implicitlydefine the number of hyper-parameters by a function. The potentialdelay structures are given by the degrees of high frequencydelays. Then the set of potentialmodelsisdefined as all the different possible combinations of functions and structure offsets with a corresponding set of initial values.

	weights	lags	starts
1	nealmon	1:2	c(1, -1)
2	nealmon	1:3	c(1, -1)
3	nbeta	1:2	c(0.5, 0.5, 0.5)
4	nbeta	1:3	c(0.5, 0.5, 0.5)

Weightvectordefines the restrictions for the Beta coefficients

Takingintoaccount the possible sets of specificities for each variable as definedabove, the estimation of the set of modelsiscarried out. So what can welearnfromtheseresults?

They are multiple. From a quantitative point of view, first of all, the results for the French economyappearsatisfactoryoverall. Growth in France isintrinsicallysmall (driven by relativelyunchangingconsumption (whichremainsunchanged)), and seems to be more easilyenvisaged by our model. This relative stability in recessive phases as well as in periods of economicupsurge, and its good estimation by our model, can beexplained by the factthat French growthisprimarilyguided by its long-termaverage, whichis the general trend of the economy, modelled by the coefficient in ourequations. Our results are in line with the consensus in the literatureonthissubject (Barhoumi et al. (2012).

2.5 Forecasting

Let us nowconsider the MIDAS regression model thatwe have justdescribed, for forecasting;itisthereforeappropriate to first define the scope of thisforecast. If, for example, at the beginning of February, wewish to forecast the growth of the second quarter (which ends at the end of June), itis a 5-month forecast (h = 5/3 in quarter) based on the data available at the end of January. Keeping the same notations, the equationcorresponding to a set of contemporaneous information a prediction at the time horizon t+1 is as follows :

Let us write the model (2.1) for the period t+1

$y_{t+1} = \alpha' y_{t,0} + \beta(l)' x_{t+1,0} + \varepsilon_{t+1}$(4.1)

Where$y_{t,0} = \left(y_t, \dots, y_{t-p+1}\right)' and \alpha = \left(\alpha_1, \alpha_2, \dots, \alpha_p\right)'$ are vectors of parameters of the autoregressive terms. This representation (at the horizon of one period) is well adapted for the conditional forecast of y_{t+1}; the only condition is that the information on the explanatory variables is available. In the absence of such information, the prediction of $x_{t+1,0}$ would also be valid for a joint process of $\{y_t, x_{t,0}\}$ which might be difficult to specify and estimate correctly, given the presence of data with mixed frequencies. Or there is also a direct approach to predictionthatcouldbeapplied in the MIDAS framework. Given a set of information available at a time t defined by : $I_{t,0} = \{y_{t,j}, x_{t,j}\}_{j=0}^{\infty}$ where :

$$y_{t,j} = \left(y_{t-j}, \dots, y_{t-j-p+1}\right)'$$
$$x_{t,j} = \left(x_{tm\,0}^{(0)}, \dots, x_{tm\,1}^{1}, \dots, x_{tm\,h}^{(h)}\right)'$$

A ℓ-period of directforecasting

$\bar{y}_{t+\ell} = E\left(y_{t+\ell} | I_{t,0}\right) = \alpha_\ell{}' y_{t,0} + \beta_\ell(L)' x_{t,0}, \quad \ell \in \mathbb{N}$ (4.2)

Can bebased on a model linked to a correspondingconditional expectation

$$y_{t+\ell} = \alpha_\ell{}' y_{t,0} + \beta_\ell(L)' x_{t,0} + \varepsilon_{\ell,t}, \qquad E\left(\varepsilon_{\ell,t} | I_{t,0}\right),$$

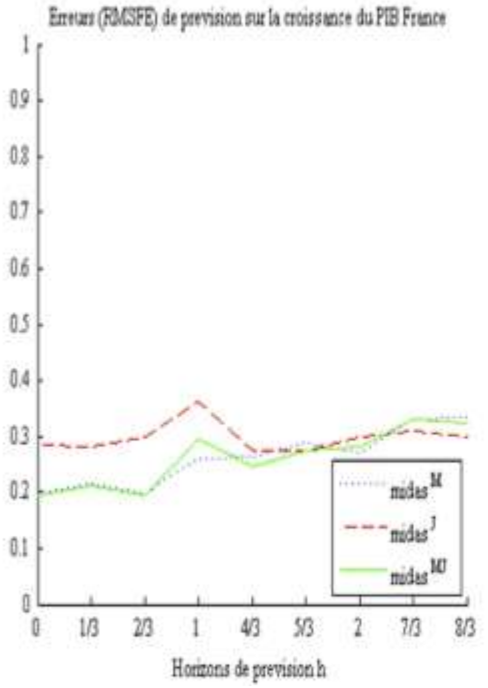
Where α_ℓ and $\beta_\ell(L)$ are specific parameters of the respective period. In principle, these conditional expectations have a particular form of representation with some restrictions on the original delay polynomials of the coefficients. In the general case, the appropriate restrictions for each l will have a differentform. For thisforecastingexercise, threemodelsbased on a MIDAS methodology are considered. Theysuccessivelyinvolvemonthly real and dailyfinancial variables, allowing us to identify the anticipatoryfactors in each of thesesectors.

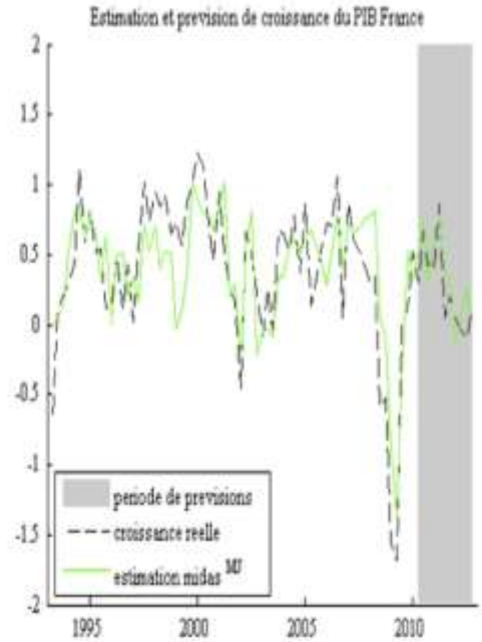
	Scheme	MSE	MAPE	MASE
1	EW	1.804712	4.434311	0.8068766
2	BICW	1.764362	4.392851	0.7979267
3	MSFE	1.803705	4.433331	0.8066651

4	DMSFE	1.802856	4.432504	0.8064866

Forecasterrorsaccording to the differentmethods

It shouldalsobenotedthat the size of the monthly or dailydatabasesis first reducedusing a factorial principal component analysismethod. This so-called FAMIDAS (Factor Augmented MIDAS) modellingwasproposed by Marcellino and Schumacher (2010). The temporal nature isspecified by exponent of the variable.





In analysingtheseresults, twoanticipatoryfacts are apparent from the above figures. First, we note that the best forecasts of French growth are given by real indicators (GDP). Indeed, for thiseconomy, the mixed model $MIDAS^{MJ}$ is the most efficient and seems to beguided by the real factor, model$MIDAS^{M}$, for short horizons (from h=4/3 that's to say. 4 monthsbefore the term). The financial variable considered in the $MIDAS^{J}$ model has a provenpredictive gain regardless of the forecast horizon. Theseresultsseem consistent with the macroeconomicanalysis of the financial structure of the French economy. Finally, we note that the multiple frequency modeling seems to perfectly fit the problem as we have consideredit of forecastingeconomicgrowth in the short term.

II. Conclusion

The purpose of ourstudywas to set out the context and issues involved in economic estimation. Differences in the sampling frequencies of macroeconomicindicatorsrestrict the efficient use of these data. In this respect, multi-frequency MIDAS (Mixed-Data Sampling) modelling has proveninstability to optimallyaggregate time series for economicestimates. The choice of estimators and the mixing of frequenciesappear to be the essential elements in estimatinggrowth. The MIDAS model technique is of particularinterest in thiscontextbecause of itsparsimony and empirical performance. Moreover, the identification of real and financialfactors has made it possible to model the nature and characteristics of the French economy in the contextstudied. The resultsobtained in thiscontext over the period show that the projections made withourmodellingprovide a relevant and preciseindicator of the evolution of contemporary French growth. However, itappearsthat certain aspects, notablyconcerning temporal aggregation in forecasting and estimation methods, maybesubject to furtherstudy and research.

References

[1]. Andreou E., Ghysels E., KourtellosA.(2010). "Regression Models With Mixed Sampling Frequencies", Journal of Econometrics, 158, pp. 246-261.
[2]. Andreou E., Ghysels E., Kourtellos A. (2011). "Forecasting with mixed-frequency data" In C M, D Hendry (eds.), Oxford Handbook of Economic Forecasting, pp. 225-245.
[3]. Andreou E., Ghysels E., Kourtellos A. (2013). "Should macroeconomic forecasters look at daily_nancial data? " Journal of Business and EconomicStatistics, 31, pp. 240-251.
[4]. Armesto M., Engemann K., Owyang M. (2010). "Forecasting with mixed frequencies." Federal Reserve Bank of St. Louis Review, 92, pp. 521{536}.
[5]. Bai J., Ghysels E., Wright J. (2012). "State Space Models and MIDAS Regressions." EconometricReviews (forthcoming).
[6]. Banerjee A., and Marcellino M. (2006). "Are there any reliable leading indicators for US inflation and GDP growth ?", International Journal of Forecasting, 22(1) pp.137–151.
[7]. Barhoumi K., Darné O., Ferrara L., and Pluyaud B. (2012). "Monthly GDP forecasting using bridge models: Application for the French economy". Bulletin of EconomicResearch, forthcoming.
[8]. Bilinskas B, Kvedaras V, Zemlys V (2013). \Testing the functional constraints on parameters in cointegrated MIDAS regressions." Workingpaper, availableupon a request.
[9]. Bilinskas B, Zemlys V (2013). Testing the Functional Constraints on Parameters in Regression Models with Cointegrated Variables of Di_erent Frequency." Submitted, availableupon a request.
[10]. Breitung J., Roling C., Elengikal S. (2013). "The statistical content and empirical testing of the MIDAS restrictions", Working paper,URLhttp://www.ect.uni bonn.de/mitarbeiter/joerg-breitung/npmidas.
[11]. BreuchT., andPagan A. R. 1979), "A Simple Test for Heteroskedasticity and Random Coefficient Variation," Econometrica, 47, pp.1287–1294.
[12]. Chauvet M., Senyuz Z., and Yoldas E. (2012). "What does financial volatility tell us about macroeconomic fluctuations ?" Workingpaper, Federal Reserve Board.
[13]. Claessens S., Kose M. A., and Terrones M. E. (2012). "How do business and financial cycles interact ?" Journal of International Economics, 87(1), pp.178–190.
[14]. Clements M., Galv~ao A. (2008). "Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US output growth.", Journal of Business and EconomicStatistics, 26, pp. 546-554.
[15]. Corsi F., (2009). A simple approximate long-memory model of realized volatility." Journal of Financial Econometrics, 7, pp. 174-196.
[16]. Ferrara, L. and Marsilli, C. (2013). Financial variables as leading indicators of GDP growth : Evidence from a MIDAS approach during the Great Recession. AppliedEconomicsLetters, 20(3), pp.233–237.
[17]. Ferrara L., Marsilli C., and Ortega J.-P. (2013). "Forecasting US growth during the Great Recession : Is the financial volatility the missing ingredient ?"
[18]. Foroni C., and Marcellino M. (2012). "A comparison of mixed approaches for modelling euro area macroeconomic variables", Technical report, EUI.
[19]. Ghysels E., Santa-Clara P., and Valkanov R. (2004). "The MIDAS touch: Mixed data sampling regression models"
[20]. Ghysels E., Sinko A., and Valkanov R. (2007). "MIDAS regressions : Further results and new directions", EconometricReviews, 26(1), pp. 53–90.
[21]. Lubrano M. (2008). « Modèles Econométriques Dynamiques à une équation », Chapitre I, pp. 5-7, janv.

[22]. Mahamat H. 2017. « Estimation des données financières à haute fréquence : Une approche par le modèle Scor-GARCH, "Thèse de doctorat, Université de Montpellier.
[23]. Mahamat H. 2017. «Simultaneous estimations of the parameters regression with Realized-GARCH-Errors", CFE-CMStatistic, Senate House, University of London, UK,
[24]. Mahamat H. 2018. "Modeling moments of order three and four of distribution of yealds, Review of Socio-Economic Perspectives, Vol. 3. Issue: 2/ December.
[25]. Marcellino M., and Schumacher C. (2010). "Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP", Oxford Bulletin of Economics and Statistics, 72(4), pp. 518–550.
[26]. Mazzoni, T. 2010. « Fast Analytic Option Valuation with GARCH », Journal of Derivatives, 18 : 18-40.
[27]. RoseléChim P. 2017. « statisticalanalysis as a principal component over a long period of time », Paper for the International Conferenceon Management and Governance, UniversityTituMaresco, Bucarest Roumania.
[28]. RoseléChim P., Panhüys B. (2018) « The digitization of the economy and the new dynamics of industrial firms », ASM-RW-18-603, Volume 2, Issue 3, December 5, DOI 555689 (2018) / ASM Annals of Social Sciences & Management Studies, Juniper Publishers, California, USA.
[29]. RoseléChim P. (2018) «Numerical economy and new industrial firm power », Volume 1, Issue 5, October 3, DOI 555573 (2018) / ASM Annals of Social Sciences & Management Studies, Juniper Publishers, California, USA.
[30]. RoseléChim P., and Radjou N. 2020. « MarketEquilibriumModels », workingpapers, BETA MINEA UR 7485, University of French Guyana, Cayenne, Guyane Française.
[31]. RoseléChim P. (2020) « French Guiana, Patterns of Growth and Development », ASM-RW-20-736, Annals of Social Sciences & Management Studies, Volume 5, Issue 4, july 13, DOI 1019080 / ASM 2020. 05. 555669, Juniper Publishers, California, USA.
[32]. Tibshirani R. (1996). "Regression Shrinkage and Selection via the Lasso", Journal of the Royal Statistical Society, 58(1), pp. 267–288.

Annexes

Staggereddelaymodels

Economictheorycommonly assumes the existence of effects spread over time betweendifferenteconomicquantities, and ignoringthis and beingsatisfiedwithonlyinstantaneous variables couldbemisleading in decisionmaking. Hence the interest in studyingmodelsthattakeintoaccount the concept of time in establishingrelationshipsbetween the variables understudy. There are several types of modelsthatallow the notion of time lag to beincluded in the analysis, in whatfollowswewilllimitourselves to the presentation of one of thesemodels, in this case the staggered lag model.

A staggered-delay model isnoted:

$y_t = \mu + \sum_{j=1}^{n} \beta_i x_{t-i} + u_t = \mu + \beta(L)x_t + u_t$ (1)

The coefficientsβ_i are the delay coefficients. Theydetermine how y_t will respond to a change in x_t. Sinceu_t is assumed to be Gaussian white noise, there is no particular statistical problem in estimating the coefficients of this model because the usual assumptions of least squares are satisfied and in particular the independencebetween the regressors and the errorterm. However a time seriesevolvesslowlybecause of memory effectssothat the different lags of the variable x_t will tend to be correlated with each other. We will therefore encounter a problem of multi-colinearity which will hamper the precision in the estimation of the regression coefficients. A particular structure willbeimposed on the shape of the lag coefficients to reduce the number of parameters to beestimated. The multi-colinearityproblemissolved by introducingadditional information. Several structures are possible:

First of all, the Almon polynomial thatwedevelop in the following section, as itis the basis of ourstudy model (MIDAS).

Almon (1965): the coefficients are constrained by a polynomial of degree$n$ lessthan the number of lags, usually2 or 3. Wewill have

$$\beta_i = \sum_{j=1}^{n} \gamma_j i^j$$

- rational staggereddelays: the structure of delaysisdetermined by the ratio of twodelaypolynomials:

$$y_t = \mu + \frac{B(L)}{A(L)} x_t + u_t$$

The article by Griliches (1967) is of interest. This type of modeling isextremely flexible since the simplest case with$B(L) = \beta_0 + \beta_1 L$ and $A(L) = 1 - \alpha L$ alreadyallows a widevariety of configurations for the delay structure.

-geometric or Koyckdelays. This is a special case of the previous one whereB(L) = 1 et$A(L) = 1 - \alpha L$. The delay coefficients decreaseexponentiallywith the length of the delay.

Let us examine the latter case in detail. In the initial staggered-delay model, a special structure isimposed on the coefficients with :

$\beta_i = \delta \alpha^i$avec$|\alpha|$<1 (2)

The values of β_i decrease very quickly over time. Also it is not very restrictive to assume an infinitenumber of delays. It can evenbeveryconvenient for calculations.

Least squares estimation isunthinkablebecause of the presence of a laggedendogenous variable, whichmakes the dependent variables correlatedwith the errorterm. Otherprocedures must beused.

In general, the effect of the exogenous variable isassumed to becomeweaker over time :

$$\beta_0 > \beta_1 > \beta_2 > \cdots > \beta_i$$

The writingalreadypresented can befurthersimplified by consideringD as the offset operatorsuch as :

$$D^i x_t = x_{t-i}$$

We'llhave:

$$y_t = \sum_{i=1}^{n} \beta_i x_{t-i} + \mu + u_t = \sum_{i=1}^{n} \beta_i D^i x_t + \mu + u_t = \left[\sum_{i=1}^{n} \beta_i D^i x_t\right] x_t + \mu + u_t \quad (3)$$

So $y_t = B(D)x_t + \mu + u_t$

With$B(D) = \beta_0 + \beta_1 D^1 + \beta_2 D^2 + \cdots + \beta_i D^i$

The number of delays$i$, can befinite or infinite, however the sum of the coefficients $\beta_i$tends towards a finitelimit. As an example:

ForD = 1, we have$B(1) = \beta_0 + \beta_1 + \beta_2 + \ldots + \beta_i$, this polynomial allows to measure the impact of the explanatory variable x_t of a quantityΔ_t on the variabley_t. The coefficients $\beta_i$represent the instantaneous multipliers and their sum the cumulative multiplier. The estimation of the parameters of the model raises a certain difficulty:

The problem of collinearitybetween the exogenous variables can beas the estimation of the coefficients, thisis all te hmore true as the number of lags is important.
This iswhatwillbe the subject of the nextdevelopment.

2. Calculation of delay coefficients
The total effectisthereforeobtained as the ratio between the sum of the coefficients of $B(L)$ and the sum of the coefficients of $A(L)$,that's to saywithoutcalculating the sequence of delay coefficients. But wemaysometimesneedit.
Let us define:
$(L) = \frac{B(L)}{A(L)} = \delta_0 + \delta_1 L + \delta_2 L^2 + \cdots$ (4)
This sequence can becalculatedsimply by a classical polynomial division operation. One can alsoproceed by identification by means of the formula:
$B(L) = D(L)A(L)$(5)
This isdonerecursively by identifying the terms on bothsides. Wewill start by treating an examplewhere the twopolynomials are of degree 1 beforegiving the general formula. We start from :

$$D(L) = \frac{\beta_0 + \beta_1 L}{1 - \alpha L} = \delta_0 + \delta_1 L + \delta_2 L^2 + \cdots$$

Fromwherewe're shooting
$B_0 + \beta_1 L = (1 - \alpha L)(\ \delta_0 + \delta_1 L + \delta_2 L^2 + \cdots)$

$$= \delta_0 + (\delta_1 - \alpha\delta_0)L + (\delta_2 - \alpha\delta_1)L^2 + \cdots$$

The identification of the powers of L of the twomemberswillgive the followingequations:
$\delta_0 = \beta_0$, $\delta_1 - \alpha\delta_0 = \beta_1$, $\delta_2 - \alpha\delta_1 = 0$
Hence the condition initiale$\delta_0 = \beta_0$, and the solution of the recurrencebecomes

$$\delta_1 = \beta_1 - \alpha\beta_0$$
$$\delta_2 = \alpha(\beta_1 - \alpha\beta_0)$$
$$\delta_3 = \alpha^2(\beta_1 - \alpha\beta_0)$$

Wededuce the relationship:$\delta_j = \alpha^{j-1}(\beta_1 - \alpha\beta_0)$.
This sequenceiseasilygeneralized for any $A(L)$ and $B(L)$. As in all the cases we will have
$\alpha_0 = 1$, itresultsthatwewillalways have the same condition initiale$\delta_0 = \beta_0$. Then comes the nextrecurrence:

$$\delta_j = \sum_{i=1}^{\min(j,r)} \alpha_i \delta_{j-i} + \beta_j \qquad \text{Si } 1 \leq j \leq 8$$
$$\delta_j = \sum_{i=1}^{\min(j,r)} \alpha_i \delta_{j-i} \qquad \text{Si j} > 8$$

3. Fisher'stest:

The Fisher test allows us to test the hypothesis of the nullity of the regression coefficients for lags greaterthan h. The hypotheses are formulated as followswhentestingdownwards a value of h between 0 < h < M.

$$H_0^1 : M - 1 \rightarrow aM = 0$$
$$H_0^2 : M - 2 \rightarrow aM - 1 = 0$$
…………………..
$$H_0^i : M - i \rightarrow aM - i + 1 = 0$$

The alternative assumptionsare:

$$H_1^1 : h = M \rightarrow aM \neq 0$$
$$H_1^2 : h = M - 1 \rightarrow aM - 1 \neq 0$$
……………………
$$H_1^i : h = M - i + 1 \rightarrow aM - i + 1 \neq 0$$

Each of the hypothèses issubject to the classicFisher's test, sowe'llhave:

$$F^* = (SCR_{M-i} - SCR_{M-i+1})/1/SCR_{M-i+1}/(n - M + i + 3)$$

Compared to the tabulated F at 1 and n-M+i-3, as soon as for a giventhreshold the calculated F ishigherthan the tabulated F, wereject the hypothèse H_0^i_and the procedure is finished. The value of the delay is thus equal to :

$$h = M - i + 1$$

In order to be able to carry out this test, the SCT must remain constant from one estimate to the other, thisindicatesthat the differentmodels must beestimatedwith an identicalnumber of observations which corresponds to the number of observations available for the largest lag, each lag causing the loss of one data.

4. The Akaikemethod :
The value of the number of lags h is the parameterwhichminimizes the so-calledAkaikefunctionwhichisgiven by :

$$\mathrm{AIC(h) = Ln(SCR_h/n) + 2h/n}$$

with $\mathrm{SCR_h}$: the sum of the residual squares for the h-delayed model. And n, the number of observations.

5. Schwarz'smethod:
It's a methodvery close to Akaike's, the value of his the one thatminimizes the followingfunction:

$$\mathrm{SC(h) = Ln(SCR_h/n) + (h - Ln(h)/n)}$$

The obviousproblemwith a staggered lag model isthat, becausex$_t$willoftenbehighlycorrelated to x_{tj1}, x_{tj2} and so on, the least-squares estimates of the coefficients will tend to bequiteimprecise. Manyways to manipulatethisproblem have been proposed, and wewilldiscussthem in whatfollows.
-Unrestrictedtesting of the parameters (as in U-MIDAS) and usingOLSmodel: y ~ trend + mls(x, 0:7, 4) + mls(z, 0:16, 12)

(Intercept)	trend	x1	x2	x3	x4
2.158494	0.098872	0.273647	0.240064	0.322580	0.140980
x5	x6	x7	x8	z1	z2
-0.027430	-0.011907	0.164761	-0.025537	0.317276	0.495292
z3	z4	z5	z6	z7	z8
0.395732	0.489775	0.200686	0.100535	-0.070955	0.105309
z9	z10	z11	z12	z13	z14
-0.084617	-0.025289	-0.001775	0.114194	0.083435	0.062434
z15	z16	z17			
-0.053978	0.080609	0.146170			

- Descriptive residuestatistics

Min	1Q	Median	3Q	Max
-2.2651	-0.6489	0.1073	0.6780	2.7707

Descriptive statisticsafter MIDAS estimation according to NLS.

The output of the usedoptimizationfunctionisunder the inspection of the MIDAS optimization output element.
- Optimizationwith the Nelder, Mead and plinearmethod
It is possible to re-estimate the NLS problemwithanotherequationusing the final solution of the previousequation as starting values. For example, itisknown, that the default algorithm in NLS is sensitive to starting values. So first we can use the standard Nelder-Mead equation to find the "mostfeasible" starting values, and then use the NLS to get the final result:
Here, we observe the Nelder-Mead methodevaluating the costfunction 60 times. The optimizationfunctionsindicate the state of convergence of the numerical constant optimizationmethod, indicating [0] successful convergence. This code isreported as part of the convergence of the MIDAS output.

(Intercept)	trend	x1	x2	z1	z2	z3
2.15023	0.09898	1.14086	-0.33081	1.93400	0.85141	-0.15309

- Optimizationaccording to the plinearmethod

(Intercept)	trend	x1	x2	z1	z2	z3
2.16436	0.09888	0.56455	-0.14399	1.97665	0.82049	-0.14808

-- Optimizationaccording to the Ofunction " NLS " method

(Intercept)	trend	x1	x2	z1	z2	z3
2.16436	0.09888	0.56455	-0.14399	1.97665	0.82049	-0.14808

- Wewant to use the optimizationalgorithm of Nelder and Mead, whichis the default option in the Optimfunction, wewill have the following output :

(Intercept)	trend	x1	x2	z1	z2	z3
2.16436	0.09888	0.56455	-0.14399	1.97665	0.82049	-0.14808

If wewant to use the Golub-Pereyraalgorithm for partial linear least squares modelsimplemented in the NLS function.

(Intercept)	trend	x1	x2	z1	z2	z3
2.16436	0.09888	0.56455	-0.14399	1.97665	0.82049	-0.14808

- In order to improve convergence, itis possible to use gradient functionsdefined by the estimator. To retrieve a constraintvectorthatestimates$\hat{\theta}$ (and, therefore, also $\hat{f} = f_{\gamma/\gamma=\hat{\gamma}}$). The minimum corresponds to the vector θ $(\beta, respectively)$.

(Intercept)	trend	x1	x2	z1	z2	z3
2.15023189	0.09898395	1.14085681	-0.33081090	1.93399719	0.85140507	-0.15309199

- Where the first variable follows and aggregatesbased on the MIDAS restriction scheme. Note that the selection of other types "A" and "B" are linked by specificequationswith a largernumber of parameters (see Table 3), hence the list of starting values must beadjusted to account for the increase in the number of (potentially) unequal impact parameters.

It shouldalsobenotedthat, wheneverrestrictivelyconnectedaggregates are used, the number of periodsshouldbe a multiple of the reportingfrequency. For example, the currentspecificationdelay for variable z is not compatible withthisrequirement and cannotberepresentedacross (periodic) aggregates, but either MLS (z, 0: 11.12, amweights, nealmon, "C") or MLS (z, 0: 23.12, amweights, nealmon, "C") wouldbevalid expressions from a code implementation point of view.

	weights	lags	starts
1	nealmon	0:10	c(1, -1)
2	nealmon	0:11	c(1, -1)
3	nealmon	0:12	c(1, -1)
4	nealmon	0:13	c(1, -1)
5	nealmon	0:14	c(1, -1)
6	nealmon	0:15	c(1, -1)
7	nealmon	0:16	c(1, -1)
8	nealmon	0:17	c(1, -1)
9	nealmon	0:18	c(1, -1)
10	nealmon	0:19	c(1, -1)
11	nealmon	0:20	c(1, -1)
12	nealmon	0:10	c(1, -1, 0)
13	nealmon	0:11	c(1, -1, 0)
14	nealmon	0:12	c(1, -1, 0)
15	nealmon	0:13	c(1, -1, 0)
16	nealmon	0:14	c(1, -1, 0)
17	nealmon	0:15	c(1, -1, 0)
18	nealmon	0:16	c(1, -1, 0)
19	nealmon	0:17	c(1, -1, 0)
20	nealmon	0:18	c(1, -1, 0)
21	nealmon	0:19	c(1, -1, 0)

22	nealmon0:20	c(1, -1, 0)

Weightvectordefines the restrictions for the variable z

Selected model with AIC = 674.5565
Based on restricted MIDAS regression model
The p-value for the null hypothesis of the test hAh.test is 0.7413531
Parameters:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.0090716	0.1192771	16.844	< 2e-16	***
trend	0.0997984	0.0008025	124.354	< 2e-16	***
x1	0.7726653	0.0788237	9.802	< 2e-16	***
x2	-0.2634821	0.0449879	-5.857	1.55e-08	***
x3	0.0231120	0.0051073	4.525	9.49e-06	***
z1	2.2396888	0.1828210	12.251	< 2e-16	***
z2	0.3844866	0.1533595	2.507	0.012835	*
z3	-0.0703943	0.0203889	-3.453	0.000656	***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Paul ROSELE CHIM, and Hisseine Saad MAHAMAT. "Mixed Data Sampling Modelling (MIDAS): Application to the forecasting of French economicgrowthrates."*IOSR Journal of Business and Management (IOSR-JBM)*, 22(10), 2020, pp. 28-44.