Disaggregation methods based on MIDAS regression

Alain Guay[1]

Université du Québec à Montréal, CIRPÉE, CIREQ and CREDDI

Alain Maurin

Université des Antilles et de la Guyane and CREDDI

April, 2015

[1]Emails: guay.alain@uqam.ca and alain.maurin@univ-ag.fr. The authors thank the editor and an anonymous referee for comments that lead to a substantial improvement of the paper. This paper has also benefited from helpful remarks and discussions during presentations at the International Conference on Policy Modeling, EcoMod, Berlin, 2008. The first author gratefully acknowledges financial support of Fonds de recherche sur la société et la culture (FRQSC).

Abstract

The need to combine data from different frequencies plays an important role for many economic decision-makers and economists. The process, which consists in using higher frequency data to construct a higher frequency indicator from its lower frequency counterpart, is called temporal disaggregation. In this paper, we propose a new temporal disaggregation technique based on MIDAS regression using time series data sampled at different frequencies. We first propose a simple disaggregation procedure more flexible than the more traditional approaches, such as Chow-Lin (1971), and we extend the procedure to a dynamic setting. The proposed procedure is flexible enough to take into account seasonality or calendar effects. An extensive simulation study examines the performance of the new approach compared to alternative approaches.

Keywords: Temporal Disaggregation, MIDAS regression

JEL Class.: C32, E32

Introduction

The need to combine data from different frequencies plays an important role for many economic decision-makers and economists. The process, which consists in using higher frequency data to construct a higher frequency indicator from its lower frequency counterpart, is called temporal disaggregation. The importance of temporal disaggregation is well reflected by the efforts of national statistical institutes to provide such disaggregated indicators.

Temporal disaggregation has been extensively studied in the econometric literature. The Denton (1971) method, which relies principally on optimizing calculations to make the annual totals of infra-annual series tally with annual reference totals, was among the first of its kind and continues to be one of the most widely used benchmarking methods. Using a formalization leading to a solution of generalized least squares in an univariate framework, Chow and Lin (1971) also produced a method that has become widely used by national statistical institutes to construct quarterly national economic accounts. While the Chow and Lin approach is applicable to static univariate models, Santos Silva and Cardoso (2001) extend the Chow and Lin approach to univariate dynamic models. Temporal disaggregation procedures using multivariate structural times series models based on state space representation have also be proposed (see Moauro and Savio 2005, and Proietti 2006).

In the domain of forecasting, modeling techniques that associate data from different frequencies have also been proposed in order to calculate projections for economic aggregates. Shen (1996) used VAR and Bayesian VAR modelling to forecast quarterly variables in Taiwan. By developing a procedure capable of taking into account data issued from monthly series, he demonstrated that this significantly increases forecast accuracy. Chin and Miller (1996), also using VAR, proposed a quarterly series forecasting procedure based on the combination of forecasts from a monthly model and a quarterly model. Schumacher and Breitung (2008) propose a factor model using mixed-frequency data to construct GDP forecasts. Recently, Clements and Galvão (2008) examine how MIDAS (Mixed Data Sampling) approach, as introduced by Ghysels, Sinko and Valkanov (2006), can be adapted for the forecasting of US output growth (see also Clements and Galvão 2009).

Marcellino and Shumacher (2010) introduce a Factor-Midas approach which exploits estimated factors rather than single or small groups of economic indicators as regressors to forecast German GDP. Andreou, Ghysels and Kourtellos (2013) also use Factor-MIDAS to examine the usefulness of daily financial data to forecast macroeconomic series.[1]

In this paper, we propose a new temporal disaggregation procedure based on an adaptation of the MIDAS regression as introduced by Ghysels, Sinko and Valkanov (2006).[2] We think that a disaggregated procedure must be kept relatively simple. By mixing data from different frequencies, MIDAS regression is a well suited framework to obtain a disaggregation procedure which is simple but at the same time more flexible than existing procedures based on an univariate equation. This also allows us to propose a procedure using dynamic models. As a result, a MIDAS-based disaggregated procedure is more flexible than the current practice adopted by most statistical offices. In particular the approach has the advantage to attach flexible weights to explanatory variables. This can take into account seasonality and calendar components without imposing, for example, that seasonal pattern of the aggregate series is proportional to that of the indicators, an underlying assumption of standard existing procedures. This point is important because seasonality and calendar effects explain a relevant part of the fluctuation of aggregate economic time series. A simulation study shows that the proposed MIDAS procedure is competitive compared to existing methods in cases of constant weights for the explanatory variables, and outperforms other existing methods in presence of varying weights.

Our work is organized as follows. Section 1 provides an overview of MIDAS regressions. In Section 2, we present our temporal disaggregation procedure based on the MIDAS regression. In particular, we show how we can obtain a disaggregated series in a dynamic setting using MIDAS regression. In Section 3, results from an extensive simulation study are presented which compares the new procedure to traditional alternative procedures. Conclusions appear in the last section.

[1] See also Kuzin, Marcellino and Schumacher (2011) for a comparison of MIDAS vs mixed frequency VAR.

[2] Recent surveys on MIDAS regressions appear in Armesto, Engemann and Owyang (2010), Andreou, Ghysels and Kourtellos (2011) and Ghysels and Valkanov (2012)

1 MIDAS regression

A regression such as MIDAS allows for the relation between variables sampled at different frequencies to be examined. We shall begin by presenting a simple case. Let us examine the variable y_t sampled once between the period $t-1$ et t. This variable, for instance, can be sampled at a quarterly frequency, the index t would thus correspond to quarters. The explanatory variable $x_t^{(m)}$ is sampled m times over the same period. For example, if $x_t^{(m)}$ is sampled on a monthly basis, the index m is equal to three. We are seeking here to examine the relationship between y_t and $x_t^{(m)}$. The objective is to project the variable y_t on the historical samplings of $x_{t-j/m}^{(m)}$. The index $t-j/m$ shows that the variable is sampled more frequently than the variable y_t. A MIDAS-type regression is defined by the relation:

$$y_t = \beta_0 + \beta_1 B\left(L^{1/m};\theta\right) x_t^{(m)} + \varepsilon_t \tag{1}$$

where $t = 1, \ldots, T$, $B\left(L^{1/m};\theta\right) = \sum_{k=0}^{K} B\left(k;\theta\right) L^{k/m}$ and $L^{1/m}$ is a lag operator such as $L^{1/m}x_t^{(m)} = x_{t-1/m}^{(m)}$. The term $B\left(L^{1/m};\theta\right)$ depends on the lag operator $L^{1/m}$ and on a parameter vector of limited dimension θ. The introduction of the vector θ allows for the handling of cases where the number of lags of $x_t^{(m)}$ is considerable. For example, y_t can depend on 12-quarter lags of the explanatory variable $x_t^{(m)}$. Since the latter is sampled on a monthly basis, the number of lags is therefore equal to 36. One approach which would allow us to address the problem of over-parameterization associated with the high number of lags in the case of a regression model with data from different frequencies is to constrain the lag polynomial of the explanatory variable. The latter therefore depends on a limited number of parameters contained here in the vector θ. Ghysels, Sinko and Valkanov (2006) proposed several alternatives for constraining the form of this lag polynomial. In the specification (1), the parameter β_1 captures the aggregate effect of lags of $x_t^{(m)}$ on y_t. This parameter can be identified by normalizing the function $B\left(L^{1/m};\theta\right)$ so that its sum is equal to the unit.

Specifying the parameters of the polynomial $B\left(L^{1/m};\theta\right)$ is the major challenge for a MIDAS-type regression model. Ghysels, Sinko et Valkanov (2006) proposed different functional forms for this polynomial. One of these forms is associated with the Almon lag polynomial (Almon, 1965). An exponential variant can be defined as follows:

$$B(k;\theta)=\frac{e^{\theta_1 k+\theta_2 k^2+\ldots+\theta_p k^p}}{\sum_{k=1}^{K} e^{\theta_1 k+\theta_2 k^2+\ldots+\theta_p k^p}}.$$

This function is very flexible and can take on several configurations, hence its interest. The simple case where $\theta_1=\theta_2=\cdots=\theta_p=0$ corresponds to the case where the lag weight of each lag is the same. This specification corresponds to the often used practice which consists of using an average of monthly values to obtain a quarterly series. In general, the weights can decline more or less rapidly or take on the desired form according to the value of the parameters θ_i, for $i=1,\cdots,p$.

The second specification proposed by Ghysels, Sinko and Valkanov (2006) depends on two parameters θ_1,θ_2 and has the following form:

$$B(k;\theta_1,\theta_2)=\frac{f\left(\frac{k}{K},\theta_1;\theta_2\right)}{\sum_{k=1}^{K} f\left(\frac{k}{K},\theta_1;\theta_2\right)}$$

and

$$\begin{aligned} f\left(x,a;b\right) &= \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\ \Gamma(a) &= \int_0^\infty e^{-x}x^{a-1}dx. \end{aligned}$$

This specification is based on the function gamma which is often used in econometrics for its flexibility. The function $f\left(x,\theta_1;\theta_2\right)$ can take on several forms, according to the values of θ_1 and θ_2. For example, when $\theta_1=\theta_2=1$, the weights are equal.

The two functions presented above involve two important characteristics: (i) the coefficients are positive; (ii) their sum is equal to 1. The fact that the coefficients are positive allows for the volatility process to be assessed and the second characteristic allows for the identification of the parameter β_1.

As shown by Ghysels, Sinko and Valkanov (2006), the specification (1) can be generalized. Lags of the endogenous variable can be added as regressors to the explanatory variable $x_t^{(m)}$, we then obtain the following dynamic specification:

$$y_t = \mu + \alpha(L)y_{t-1} + \beta_1 B\left(L^{1/m};\theta\right)x_t^{(m)} + \varepsilon_t$$

where $\alpha(L) = \sum_{q=0}^{Q} \alpha_q L^q$. The model (1) can also be extended to a multivariate context. This gives:

$$Y_t = \mu + A(L)Y_{t-1} + \mathbf{B}\left(L^{1/m};\theta\right)X_t^{(m)} + \varepsilon_t$$

where Y_t, ε_t and X_t are now vectors with μ, $A(L)$ and $\mathbf{B}$ of compatible dimensions.

2 The MIDAS regression and temporal disaggregation

We now show how a disaggregated series can be obtained from a MIDAS-type regression. The approach presented here is therefore an alternative to the more traditional Chow and Lin (1971) and Litterman (1983) approaches. Let us consider the simple case where a y_t series is sampled annually and a $x_t^{(m)}$ series is sampled on a quarterly basis.[3] The relation of the non-sampled quarterly series $y_{t-i/m}^{(m)}$ according to the sampled variable $x_{t-i/m}$ is represented as:

$$y_{t-i/m}^{(m)} = \beta_0^{(m)} + \beta_1^{(m)}\omega_i x_{t-i/m}^{(m)} + \varepsilon_{t-i/m}^{(m)}. \tag{2}$$

for $t = 1, \ldots, T$, $i = 0, \ldots, m-1$ and $m = 4$ in this case. It can be noted here that the ω_i weight can vary according to the quarter but are constrained to be positive and to sum up to unity. In cases where the annual series is the sum of the non-sampled quarterly series, we then

[3]This presentation can easily be generalized with p explanatory variables.

have $y_t = \sum_{i=0}^{m-1} y_{t-i/m}^{(m)}$. Using the relation (2) we obtain:

$$y_t = \sum_{i=0}^{m-1} y_{t-i/m}^{(m)} = m\beta_0^{(m)} + \beta_1^{(m)} \left(\sum_{i=0}^{m-1} \omega_i x_{t-i/m}^{(m)} \right) + \sum_{i=0}^{m-1} \varepsilon_{t-i/m}^{(m)}.$$

This relation corresponds to a MIDAS regression having the following form

$$y_t = \beta_0 + \beta_1 B \left(L^{1/m}; \theta \right) x_t^{(m)} + \varepsilon_t. \tag{3}$$

with $\beta_0 = m\beta_0^{(m)}$, $\beta_1 = \beta_1^{(m)}$, $\sum_{i=0}^{m-1} \omega_i x_{t-i/m}^{(m)} = B \left(L^{1/m}; \theta \right) x_t^{(m)}$ and $\varepsilon_t = \sum_{i=0}^{m-1} \varepsilon_{t-i/m}^{(m)}$.

In this case, the problem of disaggregation consists of constructing a quarterly series $y_t^{(m)}$ using the annual series y_t and the regression (3). Under the hypothesis that the error term is a homoscedastic white noise, a disaggregated series corresponding to the non-sampled quarterly series in (2) can be obtained using the following transformation:

$$y_{t-i/m}^{(m)} = \frac{1}{m}\hat{\beta}_0 + \hat{\beta}_1 \hat{\omega}_i x_{t-i/m}^{(m)} + \frac{1}{m}\hat{\varepsilon}_t. \tag{4}$$

where $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\omega}_i$ are parameter estimators for the MIDAS regression (3) and $\hat{\varepsilon}_t$ the residual from this regression. More specifically, the estimates of quarterly weights ω_i in (2) are provided by $\hat{B} \left(L^{1/m}; \theta \right)$ of the MIDAS regression (3) and the estimates of $\varepsilon_{t-i/m}^{(m)}$ in (2) by the equally weighted residuals $\frac{1}{m}\hat{\varepsilon}_t$ for $i = 0, \ldots, m-1$.

Let us now compare the approach developed here using the MIDAS regression to the Chow-Lin approach. Under the latter, the weighting of the quarterly indicator variable in the relation (2) is assumed to be the same for all quarters, i.e. $\beta_1^{(m)} \omega_i = \beta_1^{(m)} \omega$ for all i imposing an invariant relation according to the quarter. In the case of the MIDAS regression, this weighting is estimated using the polynomial $B \left(L^{1/m}; \theta \right)$. The approach proposed here uses a wider range of information than Chow and Lin's more standard method. This additional flexibility should improve aggregate performance and allow, among other things, for the possible presence of seasonality or calendar effects. In particular, in presence of seasonality, the flexible weights of the MIDAS regression avoid to impose that seasonal pattern of the aggregate series is proportional to that of the indicators, an underlying assumption of the Chow-Lin procedure.

Chow and Lin's (1971) approach also allows to take the autocorrelation into account by correcting the efficient least squares estimator in order to obtain a generalized least squares estimator. This correction is achieved using a first order autoregressive process for the error term of the disaggregated series. Unlike the approach proposed here, this temporal dependence does not serve to improve the calculation of the disaggregated series, but simply to obtain an effective estimator of β_0 and β_1. Accordingly, the disaggregated series obtained by Chow and Lin (1971) for the representation (3) is given in matrix form by:

$$\hat{Y}^{(m)} = X^{(m)}\hat{\beta} + V\mathcal{C}\left(\mathcal{C}'V\mathcal{C}\right)^{-1}\mathcal{C}\hat{\varepsilon}^{(m)}.$$

where $\hat{Y}^{(m)}$ is a $Tm \times 1$ vector containing the disaggregated series $y^{(m)}_{t-i/m}$ for $t = 1, \ldots, T$ and $i = 0, \ldots, m-1$, $X^{(m)}$ is a matrix containing a constant and the disaggregated samples $x^{(m)}_{t-i/m}$, $\hat{\varepsilon}^{(m)}$ is the vector of the residuals $\hat{\varepsilon}^{(m)}_{t-i/m}$, V is the variance-covariance matrix that takes into account the presence of autocorrelation and heteroscedasticity in the disaggregated error terms and $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$. The matrix $\mathcal{C}$ with a dimension of $T \times Tm$ has the following written form in the case of annual aggregated series and quarterly disaggregated series:

$$\mathcal{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & . & . & . & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & . & . & . & 0 & 0 & 0 & 0 \\ . & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & . & . & . & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{5}$$

The observed aggregated series Y satisfies the following regression model:

$$Y = X\beta + \varepsilon$$

with $X = \mathcal{C}X^{(m)}$, $Y = \mathcal{C}\hat{Y}^{(m)}$ and $\varepsilon = \mathcal{C}\varepsilon^{(m)}$. Thus, the estimators of $\beta = (\beta_0, \beta_1)'$ are the generalized least squares estimators defined as follows:

$$\hat{\beta} = \left[X'(\mathcal{C}V\mathcal{C}')^{-1}X\right]^{-1} X'(\mathcal{C}V\mathcal{C}')^{-1}Y$$

and $\hat{\varepsilon} = \mathcal{C}\hat{\varepsilon}^{(m)} = Y - X\hat{\beta}$. Based on the Chow and Lin approach, other procedures with alternative dynamics have been proposed. For example, Fernandez (1981) assumes a unit root for the error term and, Litterman (1983) suppose an AR(2) process with a unit root for the error term.

The MIDAS approach can also take into account the presence of autocorrelation. This can be done by performing a two-step estimation procedure. Suppose a first order autocorrelation coefficient denoted by ϕ. The first step consists of estimating the parameters of interest by using the identity matrix as the weighting matrix. Using the first-step estimator of the weight parameters ω_i, an estimator of ϕ is obtained by minimizing the generalized least criteria for the given first-step estimator $\hat{\omega}_i^1$. In fact, this corresponds to the Chow-Lin approach using known fixed weights $\hat{\omega}_i^1$. This estimator of ϕ allows to construct the weighting matrix uses in the second-step which elements (i, j) of this weighting matrix denoted by $\Omega(\phi)$ are given by $\phi^{|i-j|}$.

As aforementioned, the MIDAS regression can be extended to a dynamic setting by introducing an autoregressive form. For instance, the lagged variables of the aggregated series can be included, which can improve the calculation of the disaggregated series. Let us examine MIDAS regression including a lag of the dependent variable. We obtain the following regression equation:

$$y_t = \beta_0 + \rho y_{t-1} + \beta_1 B\left(L^{1/m}; \theta\right) x_t^{(m)} + \varepsilon_t.$$

This specification can be rewritten in the following form:

$$y_t = \frac{\beta_0}{(1-\rho)} + \frac{\beta_1 B\left(L^{1/m}; \theta\right)}{(1-\rho L)} x_t^{(m)} + \frac{\varepsilon_t}{(1-\rho L)}.$$

We thereby obtain a polynomial having the form $B\left(L^{1/m}; \theta\right) \sum_{j=0}^{\infty} \rho^j L^j$. This polynomial is compatible with the seasonal effects of $x_t^{(m)}$ on $y_t^{(m)}$. Indeed, for our example, this polynomial has a moving average representation with periodical effects corresponding to quarters. This MIDAS regression is compatible with a relation between the disaggregated variables given by:

$$y_{t-i/m}^{(m)} = \beta_0^{(m)} + \rho y_{t-(m-i)/m}^{(m)} + \beta_1^{(m)} \omega_i x_{t-i/m}^{(m)} + \varepsilon_{t-i/m}^{(m)}.$$

We can easily think that it would be more pertinent to examine the relation between the disaggregated series based on a first-order autoregressive representation of the non-sampled disaggregated series as in Santos-Silva and Cardoso (2001). We are rather interested by the following equation:

$$y_{t-i/m}^{(m)} = \beta_0^{(m)} + \rho^{(m)} y_{t-(i-1)/m}^{(m)} + \beta_1 \omega_i x_{t-i/m}^{(m)} + \varepsilon_{t-i/m}^{(m)}. \tag{6}$$

where the variable $y_{t-i/m}^{(m)}$ depends on its non-sampled value from the preceding quarter. This specification can be rewritten in the following form:

$$y_{t-i/m}^{(m)} = \frac{\beta_0^{(m)}}{(1-\rho^{(m)})} + \frac{\beta_1 \omega_i}{(1-\rho^{(m)} L^{1/m})} x_{t-i/m}^{(m)} + \frac{\varepsilon_{t-i/m}^{(m)}}{(1-\rho^{(m)} L^{1/m})}.$$

which gives

$$y_{t-i/m}^{(m)} = \frac{\beta_0^{(m)}}{(1-\rho^{(m)})} + \beta_1 \sum_{j=0}^{\infty} \left(\rho^{(m)}\right)^j \omega_i x_{t-i/m-j/m}^{(m)} + \sum_{j=0}^{\infty} \left(\rho^{(m)}\right)^j \varepsilon_{t-i/m-j/m}^{(m)}. \tag{7}$$

We thereby obtain a polynomial depending on $L^{1/m}$ having the form $B\left(L^{1/m};\theta\right) \sum_{j=0}^{\infty} \left(\rho^{(m)}\right)^j L^{j/m}$. This specification represents a challenge in terms of estimation since the lagged variable is not sampled. In what follows, we adopt the strategy proposed by Santos-Silva and Cardoso (2001) but adapted to a MIDAS regression. This strategy has the great advantage that does not need to condition on the initial observations and allows to compute directly the disaggregated series.

In that respect, the equation (7) can be expressed as

$$y_{t-i/m}^{(m)} = \frac{\beta_0^{(m)}}{(1-\rho^{(m)})} + \beta_1 \sum_{j=0}^{mt-i-1} \left(\rho^{(m)}\right)^j \omega_i x_{t-i/m-j/m}^{(m)} + \left(\rho^{(m)}\right)^{mt-i} \mu + \nu_{t-i/m}. \tag{8}$$

where $\mu = \beta_1 \sum_{j=0}^{\infty} \left(\rho^{(m)}\right)^j \omega_i x_{-j/m}^{(m)}$ is the truncation remainder and $\nu_{t-i/m} = \rho^{(m)} \nu_{t-i/m-1/m} + \varepsilon_{t-i/m}$. By aggregating this relation, we obtain:

$$y_t = \sum_{i=0}^{m-1} y_{t-i/m}^{(m)} = m\frac{\beta_0^{(m)}}{(1-\rho^{(m)})} + \beta_1 \sum_{i=0}^{m-1} \sum_{j=0}^{(mt-(i+1))} \left(\rho^{(m)}\right)^j \omega_i x_{t-i/m-j/m}^{(m)} + \sum_{i=0}^{m-1} \left(\rho^{(m)}\right)^{t-i/m} \mu + \tilde{\nu}_t.$$

where $\tilde{\nu}_t = \sum_{i=0}^{m-1} \nu_{t-i/m}$.

The equation (8) can be rewritten as

$$y_{t-i/m}^{(m)} = \frac{\beta_0^{(m)}}{(1-\rho^{(m)})} + \beta_1 X_{t-i/m}(\rho, w_i) + \left(\rho^{(m)}\right)^{t-i/m} \mu + \nu_{t-i/m}.$$

where $X_{t-i/m}(\rho^{(m)}, w_i) = \sum_{j=0}^{mt-(i+1)} \left(\rho^{(m)}\right)^j \omega_i x_{t-i/m-j/m}^{(m)}$. By aggregating the relation, we get

$$y_t = m\frac{\beta_0^{(m)}}{(1-\rho^{(m)})} + \beta_1 \sum_{i=0}^{m-1} X_{t-i/m}(\rho^{(m)}, \omega_i) + \sum_{i=0}^{m-1} \left(\rho^{(m)}\right)^{t-i/m} \mu + \tilde{\nu}_t.$$

Estimates of β_0, β_1, μ and ω_i for $i = 1, \ldots, m-1$ can be obtained by a weighted nonlinear regression with the weighting matrix $\Omega^*(\rho^{(m)}) = \mathcal{C}\Omega(\rho^{(m)})\mathcal{C}'$ where the elements (i, j) of $\Omega(\rho^{(m)})$ are given by $\left(\rho^{(m)}\right)^{|i-j|}$ and the matrix $\mathcal{C}$ is defined in (5). The objective function to minimized is

$$\tilde{\nu}'(\delta)\Omega^*(\rho^{(m)})^{-1}\tilde{\nu}(\delta)$$

with $\tilde{\nu}(\delta) = (\tilde{\nu}_1(\delta), \tilde{\nu}_2(\delta), \ldots, \tilde{\nu}_T(\delta))'$ where T is the number of observations for the aggregated series and $\delta = (\rho, \beta_0, \beta_1, \omega_0, \ldots, \omega_{m-1}, \mu)'$. Since the weighting matrix also depends on a parameter to estimate, a two-step procedure can be implemented with the identity matrix as the weighting matrix at the first step. This two-step weighted nonlinear least squares allows to jointly estimate the parameters of interest ρ, β_0, β_1, μ and ω_i for $i = 1, \ldots, m-1$. According to equation (8), the parameter μ can be interpreted as the conditional expectation of y_0 given the past values of $X_{t-i/m}(\rho^{(m)}, \omega_i)$ providing an estimator of the initial condition for y_t. The method then allows the construction of the estimated disaggregated series using (8), namely

$$\hat{y}^{(m)}_{t-i/m} = \frac{\hat{\beta}^{(m)}_0}{(1-\hat{\rho}^{(m)})} + \hat{\beta}_1 \sum_{j=0}^{mt-i-1} \left(\hat{\rho}^{(m)}\right)^j \hat{\omega}_i x^{(m)}_{t-i/m-j/m} + \left(\hat{\rho}^{(m)}\right)^{mt-i} \hat{\mu} + \hat{\nu}_{t-i/m}$$

where the vector of the residuals $\hat{\nu}_{t-i/m}$ is given as $\Omega\left(\hat{\rho}^{(m)}\right) C' \left(C\Omega\left(\hat{\rho}^{(m)}\right) C'\right)^{-1} \tilde{\nu}'(\hat{\delta})$.

3 Simulation experiments

In this section we examine the small sample properties of the temporal disaggregation technique based on MIDAS regression and we compare the performance of the new proposed methods respective to more traditional ones. First, we evaluate the ability of the MIDAS regression to estimate correctly parameter values in finite sample. The first DGP investigated generates a disaggregated series according to equation (4), namely:

$$y^{(m)}_{t-i/m} = \frac{1}{m}\beta_0 + \beta_1 \omega_i x^{(m)}_{t-i/m} + \frac{1}{m}\varepsilon_t. \tag{9}$$

with $m = 4$. The aggregation of the series is such that:

$$y_t = \sum_{i=0}^{m-1} y^{(m)}_{t-i/m} = \beta_0 + \beta_1 \sum_{i=0}^{m-1} \omega_i x^{(m)}_{t-i/m} + \varepsilon_t.$$

We also allow the error term to be characterized by an AR(1) process, namely:

$$\varepsilon_t = \phi\varepsilon_{t-1} + u_t$$

in order the evaluate the performance of dynamic regression with MIDAS.

In the first set of experiments (denoted Case 1), we examine the small sample performance with equal weights for each quarter. In this case, there is no advantage to use the method proposed here, in comparison with the Chow-Lin method. These experiments allow to investigate to which extend MIDAS regression can create spurious dynamic in the dissaggregated series. We fix the following values to the parameters: $\beta_0 = 8$, $\beta_1 = 8$ and $\omega_i = .25$ for $i = 1, 2, 3, 4$. The variables $x_{t-i/m}$ are drawn from a normal centered to a value of two with a variance also equal to two.[4] The standard deviation of the AR(1) process ε_t is fixed to one such as the error term u_t is drawn from a normal distribution centered to zero with $(1 - \phi^2)$ as variance. We simulate a disaggregated series of 200 observations for an aggregated series of 50 observations. We run 1000 simulations and calculate the mean bias and the root mean square errors of the estimator for parameters β_0, β_1, ω_1, ω_2, ω_3 and ω_4. The performance of the MIDAS approach is examined for different values of ϕ fixed to 0, .5 and .9. Table 1a contains results for this first set of simulation experiments.

In the second set of experiments (Case 2), we fix the weights to different values. We consider that $\omega_1 = .4$, $\omega_2 = \omega_3 = .25$ and $\omega_4 = .1$. Finally, we use the same values for the autocorrelation parameter as in the first set of experiments. Results appear in Table 1b.

For both Case 1 and Case 2, estimates obtained with the MIDAS regression are extremely precise in terms of the bias and the root mean squared errors. The bias and the root mean square errors increase slightly with the autocorrelation coefficient for all parameters. Among the parameters, the constant parameter β_0 is the less precisely estimated especially for $\phi = .9$. Overall, the performance of the MIDAS regression is greatly satisfactory.

We now examine the performance of the MIDAS regression for the autoregressive form corresponding to equation (6), namely

$$y^{(m)}_{t-i/m} = \rho^{(m)} y^{(m)}_{t-(i-1)/m} + \beta_1 \omega_i x^{(m)}_{t-i/m} + \varepsilon^{(m)}_{t-i/m}. \tag{10}$$

[4] Results are robust to other values of these parameters.

with $m = 4$. The specification has no constant and the error term $\varepsilon_{t-i/m}^{(m)}$ follows a $N(0,1)$. The aggregation of the series is also:

$$y_t = \sum_{i=0}^{m-1} y_{t-i/m}^{(m)}.$$

The case with equal weights (Case 1) and different weights (Case 2) are also investigated with the same parameter values as in the preceding experiments with $\rho^{(m)} = .25, .5, .75$. In contrast to the experiments above, the autoregressive parameters is now estimated. Results appear respectively in Tables 2a and 2b. Again here, parameters are precisely estimated in both cases and the performance does not vary much across autoregressive parameters with the exception of the parameter β_1 which is more precisely estimated with a higher value of $\rho^{(m)}$.

We now study the ability of the MIDAS disaggregation approach to retrieve the true disaggregated series compared to more traditional approaches. To do so, we first simulate disaggregated series corresponding to equations (9) and (10) and we aggregate these series afterward. To evaluate the performance of the various disaggregation methods, we compute the correlation between the true disaggregated series and the ones estimated by the various disaggregation methods investigated. The correlation is calculated for the series in level and in difference and the corresponding 90 % confidence interval is reported. The comparison is done for the non-dynamic MIDAS, dynamic MIDAS methods (denoted MIDAS-AR1 in the tables), Chow-Lin (1971), Litterman (1983), Fernandez (1981) methods and the dynamic method proposed by Santos Silva and Cardoso (2001) (denoted SSC in the tables). For both equations (9) and (10), we consider Case 1 with equal weights and Case 2 with different weights fixed at the same values as previous experiments. For the case corresponding to equation (9), the parameters are fixed to the same values as previously, namely: $\beta_0 = 8$, $\beta_1 = 8$, except that $\phi = .5, .75, .9$. The variables $x_{t-i/m}$ are also drawn from a normal centered to a value of two with a variance also equal to two. The standard deviation of the AR(1) process ε_t is fixed to one such that the error term u_t is drawn from a normal distribution centered to zero with $(1 - \phi^2)$ as variance. For the case corresponding to equation (10), the parameters are also fixed to the same values as previously, namely: $\beta_1 = 8$ and $\rho^{(m)} = .5, .75, .9$. The variables

$x_{t-i/m}$ are also drawn from a normal centered to a value of two with a variance equal to 2 and the error term $\varepsilon_{t-i/m}^{(m)}$ is drawn from a $N(0,1)$.

The number of observations for the simulated disaggregated series is fixed to 100 which yields corresponding aggregated series of 25 observations. The number of simulations is equal to 1000. Results for the equation (9) appear in Tables 3a and Table 3b for equal weights and different weights respectively while results for the equation (10) appear in Tables 4a and 4b for equal weights and different weights respectively.

Table 3a shows that Chow-Lin, Litterman, Fernandez and MIDAS methods perform equally well to identify the disaggregated series. Correlations between predicted disaggregated series and the true ones are all included between .970 and .995. The Chow-Lin method and the MIDAS regression slightly dominate other methods but not significantly. This result is encouraging for the MIDAS based approach because this method does not impose fixed weights in contrast to Chow-Lin method. The two dynamic methods SSC and MIDAS-AR1 underperform for this specification with a slight advantage to MIDAS-AR1. This is not surprising because these two methods are misspecified with respect to the DGP corresponding to equation (9). For Case 2 (Table 3b), as expected, the MIDAS regression outperforms other methods with correlation coefficients around .85 to .88 for series in level and .82 to .83 for series in difference. For Chow-Lin, Litterman and Fernandez, correlation coefficients are between .80 and .82 for series in level and .75 to .76 for series in difference. Interestingly, the MIDAS-AR1 outperforms other methods excepts the MIDAS one. The possibility of flexible weights seems to offset the misspecification problem. In conclusion, the MIDAS method seems to well capture the presence of unequal weights in the disaggregated series. Finally, the results are not very sensitive to the value of ϕ.

We now turn to the dynamic specification (10). Results for equal weights are reported in Table 4a. As expected, both dynamic methods, SSC and MIDAS, dominate other methods. This is true especially for the correlation between the true and the predicted series in difference. For this DGP, both dynamic methods are well specified in contrast to the other ones. Also as expected, SSC slightly dominates MIDAS-AR1. Since SSC constraints the weights to be fixed, a true restricted version should perform better. For the case with unequal weights, Table 4b shows that MIDAS-

AR1 clearly outperforms other methods especially for the correlation in difference. For instance, with $\rho = .9$, correlation in difference is equal to .9505 when the disaggregated series is predicted by MIDAS-AR1 whereas this correlation is equal to .8012 with the alternative dynamic method SCC. The correlation is around .60 and .66 for other methods. In this case, there is an improvement in the correlation coefficients with higher value of the persistence parameter $\rho^{(m)}$. In conclusion, these simulation experiments show that there exists almost no cost to using MIDAS methods in cases with equal weights and a clear advantage when these weights can differ.

Conclusion

We propose a new temporal disaggregation procedure based on an adaptation of the MIDAS regression as introduced by Ghysels, Sinko and Valkanov (2006). The MIDAS framework allows us to propose a simple procedure using dynamic models. The novel MIDAS-based disaggregated procedure is more flexible than the current practice adopted by most statistical offices. In particular, the procedure does not assume that the seasonal pattern of the aggregate series is proportional to that in the indicator. A simulation study shows that the proposed MIDAS disaggregation procedure is competitive or outperforms existing alternative procedures.

In this paper, we have only considered a MIDAS regression with one regressor although the approach can accommodate higher dimensional system and other frequencies of observations. As future research, an extended Monte-Carlo study should be performed to evaluate if the simulation results obtained here are robust to higher dimension system and other frequencies of disaggregated series. It would be also interesting to extend the disaggregated methods based on MIDAS proposed in this paper to Factor-MIDAS models to take into account all potential disaggregated information to retrieve a more accurate disaggregated series. Finally, applications on real series are under process by both authors.

References

[1] Almon, S. (1965), " The distributed Lag between Capital Appropriations and Expenditures," *Econometrica*, 33, 178-196.

[2] Andreou E., E. Ghysels and A. Kourtellos (2011), Forecasting with Mixed-Frequency Data, in Clements, M. and D. Hendry, ed: Oxford Handbook of Economic Forecasting, pp. 225–245.

[3] Andreou E., E. Ghysels and A. Kourtellos (2013), Should Macroeconomic Forecaster Use Daily Financial Data and How?, Journal of Business & Economic Statistics, 31, 240–251.

[4] Armesto M.T., K.M. Engemann and M.T. Owyang (2010), Forecasting with Mixed Frequencies, Federal Reserve Bank of St-Louis, 92, 521–536.

[5] Chin D.M. and Miller P.J., 1996, Using monthly data to improve quarterly model forecasts, Federal Reserve Bank of Minneapolis Review, vol. 20, No 2, 16-33.

[6] Chow G.C. and Lin An-Loh (1971), Best linear unbiased interpolation, distribution and extrapolation of times series by related series, Review of Economic and Statistics, Vol. 53, 372-375.

[7] Clements, M.P. and A.B. Galvõ, (2008), "Macroeconomic Forecasting with Mixed Frequency Data: Forecasting US Output Growth", Journal of Business & Economic Statistics, 26, 546-554.

[8] Clements, M.P. and A.B. Galvõ, (2009), "Forecasting US Output Growth Using Leading Indicators: An Appraisal Using MIDAS Models", Journal of Applied Econometrics, 24, 1187-1206.

[9] Denton F.T., 1971, Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization, Journal of the American Statistical Association, Vol. 66, 92-102.

[10] Fernandez, R.B., 1981, A methodological note on the estimation of time series, Review of Economics and Statistics, 63, 471-476.

[11] Ghysels, E., A. Sinko and R. Valkanov (2006), "MIDAS Regressions: Further Results and New Directions", Econometric Review, Vol. 26, 53-90.

[12] Ghysels, E. and R. Valkanov (2012), "Forecasting Volatility with MIDAS", in L. Bauwens, C. Hafner and S. Laurent, ed: Handbook of Volatility Models and Their Applications, pp. 383–401. Wiley.

[13] Kuzin, V., M. Marcellino and C. Schumacher (2011), MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area, International Journal of Forecasting 27, 529-542.

[14] Litterman, R.B., 1983, A random walk, Markov model for the distribution of time series, Journal of Business & Economic Statistics, 63, 471-476.

[15] Marcellino, M. and C. Schumacher (2010), Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP, Oxford Bulletin of Economics and Statistics 72, 518-550.

[16] Moauro, F. and G. Savio, 2005, "Temporal Disaggregation Using Multivariate Structural Time Series", Econometrics Journal, vol. 8, 214–234.

[17] Proietti T., 2006, Temporal disaggregation by state space methods: dynamic regression methods revisited, Econometrics Journal, vol. 9, 357–372.

[18] Santos Silva J.M.C. and F.N. Cardoso, 2001, The Chow-Lin method using dynamic models, Economic Modelling, vol. 18, 269-280.

[19] Shen C. H., 1996, Forecasting macroeconomic variables using data of different periodicities, International journal of forecasting, No 12, 269-282.

[20] Schumacher, C. and J. Breitung (2008), Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data, International Journal of Forecasting 24, 368-398.

Table 1a: Bias and root mean square errors (rmse)

	Case 1					
	$\phi = 0$		$\phi = .5$		$\phi = .9$	
	bias	rmse	bias	rmse	bias	rmse
ω_1	.0019	.0005	.0026	.0010	.0025	.0016
ω_2	-.0014	.0005	-.0016	.0010	- .0006	.0017
ω_3	-.00002	.0005	-.0004	.0010	-.0018	.0016
ω_4	.0005	.0005	-.0006	.0010	-.0002	.0016
β_0	.0076	.7788	.0177	1.6314	.0051	3.7331
β_1	-.0025	.1722	-.0078	.3465	.0014	.5641

Table 1b: Bias and root mean square errors (rmse)

	Case 2					
	$\phi = 0$		$\phi = .5$		$\phi = .9$	
	bias	rmse	bias	rmse	bias	rmse
ω_1	.0023	.0006	.0036	.0012	.0037	.0019
ω_2	-.0014	.0005	-.0016	.0010	- .0006	.0017
ω_3	-.00002	.0005	-.0005	.0010	-.0018	.0016
ω_4	-.0009	.0006	-.0015	.0012	-.0012	.0018
β_0	.0076	.7788	.0172	1.6305	.0013	3.7148
β_1	-.0025	.1722	-.0075	.3463	.0034	.5600

Table 2a: Bias and root mean square errors (rmse) for AR(1)

	Case 1					
	$\rho^{(m)} = .25$		$\rho^{(m)} = .5$		$\rho^{(m)} = .75$	
	bias	rmse	bias	rmse	bias	rmse
ω_1	.00054	.0023	.00087	.0020	.0011	.0020
ω_2	.00066	.0007	.00046	.0007	-.000016	.0007
ω_3	-.00010	.0013	-.0012	.0017	-.00074	.0020
ω_4	-.00140	.0019	-.00029	.0021	.00077	.0020
ρ	-.0128	.0103	-.0058	.0037	-.0023	.00069
β_1	.1532	1.2872	.1061	1.0675	.0857	.9054

Table 2b: Bias and root mean square errors (rmse) for AR(1)

	Case 2					
	$\rho^{(m)} = .25$		$\rho^{(m)} = .5$		$\rho^{(m)} = .75$	
	bias	rmse	bias	rmse	bias	rmse
ω_1	-.00018	.0013	.00058	.0012	.00025	.0014
ω_2	.0029	.0012	.0016	.0012	.0011	.0013
ω_3	.00046	.0014	-.0018	.0015	-.00096	.0018
ω_4	-.0032	.0017	-.00042	.0020	-.00043	.0019
ρ	-.0329	.0238	-.0150	.0079	-.0024	.0012
β_1	.3457	2.6314	.2289	1.8631	.0705	1.1065

Table 3a:

Correlation between the true and the estimated disaggregated series

$\omega_1 = .25$, $\omega_2 = .25$, $\omega_3 = .25$, $\omega_4 = .25$.

	Case 1					
	$\phi = .5$		$\phi = .75$		$\phi = .9$	
	Level	Difference	Level	Difference	Level	Difference
Chow-Lin	.9717	.9721	.9854	.9866	.9943	.9949
	(.9590,.9809)	(.9604,.9811)	(.9783,.9905)	(.9811,.9910)	(.9914,.9963)	(.9928,.9966)
Litterman	.9698	.9715	.9843	.9864	.9939	.9948
	(.9550,.9800)	(.9595,.9806)	(.9751,.9900)	(.9805,.9908)	(.9904,.9962)	(.9926,.9965)
Fernandez	.9713	.9720	.9851	.9866	.9941	.9949
	(.9587,.9806)	(.9604,.9810)	(.9778,.9904)	(.9811,.9909)	(.9911,.9963)	(.9928,.9965)
SSC	.8367	.8598	.8109	.8344	.8127	.8355
	(.6188,.9608)	(.6742,.9606)	(.5884,.9584)	(.5824,.9597)	(.5907,.9624)	(.5695,.9668)
MIDAS	.9718	.9721	.9855	.9867	.9943	.9949
	(.9590,.9809)	(.9604,.9811)	(.9783,.9905)	(.9811,.9910)	(.9914,.9963)	(.9927,.9965)
MIDAS-AR1	.8804	.8648	.8857	.8671	.8946	.8747
	(.7661,.9503)	(.7114,.9545)	(.7649,.9595)	(.7118,.9638)	(.7784,.9692)	(.7206,.9733)

Table 3b:

Correlation between the true and the estimated disaggregated series

$\omega_1 = .4$, $\omega_2 = .25$, $\omega_3 = .25$, $\omega_4 = .1$.

	Case 2					
	$\phi = .5$		$\phi = .75$		$\phi = .9$	
	Level	Difference	Level	Difference	Level	Difference
Chow-Lin	.8057	.7536	.8161	.7600	.8214	.7633
	(.7385.8614)	(.6595,.8310)	(.7532,.8680)	(.6653,.8351)	(.7625,.8715)	(.6701,.8366)
Litterman	.8014	.7511	.8125	.7578	.8184	.7615
	(.7322,.8588)	(.6554,.8296)	(.7473,.8660)	(.6634,.8341)	(.7560,.8692)	(.6668,.8361)
Fernandez	.8049	.7532	.8153	.7596	.8207	.7630
	(.7381,.8610)	(.6576,.8310)	(.7526,.8678)	(.6655,.8345)	(.7609,.8711)	(.6695,.8365)
SSC	.6798	.6548	.6758	.6472	.6761	.6468
	(.5793,.7588)	(.5597,.7375)	(.5758,.7551)	(.5488,.7343)	(.5797,.7553)	(.5385,.7320)
MIDAS	.8597	.8274	.8727	.8386	.8771	.8398
	(.7513,.9760)	(.6728,.9772)	(.7643,.9856)	(.6829..9876)	(.7696,.9942)	(.6820,.9949)
MIDAS-AR1	.8281	.8293	.8306	.8266	.8352	.8304
	(.6871,.9411)	(.6474,.9449)	(.6872,.9481)	(.6305,.9523)	(.6918,.9530)	(.6440,.9546)

Table 4a:

Correlation between the true and the estimated disaggregated series

$\omega_1 = .25$, $\omega_2 = .25$, $\omega_3 = .25$, $\omega_4 = .25$.

	Case 1					
	$\rho^{(m)} = .5$		$\rho^{(m)} = .75$		$\rho^{(m)} = .9$	
	Level	Difference	Level	Difference	Level	Difference
Chow-Lin	.8968	.8201	.9074	.8185	.9758	.7915
	(.8538.9302)	(.7721,.8498)	(.8411,.9529)	(.7765,.8585)	(.9478,.9907)	(.7365,.8510)
Litterman	.9086	.8259	.9137	.8150	.9770	.7717
	(.8517,.9597)	(.7639,.8669)	(.8424,.9588)	(.7733,.8644)	(.9466,.9915)	(.5240,.8566)
Fernandez	.8933	.8200	.9052	.8194	.9752	.7946
	(.8566,.9300)	(.7672,.8501)	(.8404,.9524)	(.7752,.8584)	(.9447,.9907)	(.7378,.8535)
SSC	.9866	.9548	.9861	.9506	.9971	.9557
	(.9794,.9911)	(.9479,.9671)	(.9774,.9922)	(.9328,.9653)	(.9949,.9985)	(.9393,.9685)
MIDAS	.6524	.5211	.6435	.4708	.8356	.3726
	(.5311,.7932)	(.3103,.6653)	(.4465,.8728)	(.2367..8015)	(.6788,.9768)	(.1217,.7912)
MIDAS-AR1	.9790	.9185	.9784	.9190	.9956	.9276
	(.9626,.9875)	(.8525,.9539)	(.9608,.9897)	(.8515,.9564)	(.9915,.9980)	(.8700,.9615)

Table 4b:

Correlation between the true and the estimated disaggregated series

$\omega_1 = .4$, $\omega_2 = .25$, $\omega_3 = .25$, $\omega_4 = .1$.

	Case 2					
	$\rho^{(m)} = .5$		$\rho^{(m)} = .75$		$\rho^{(m)} = .9$	
	Level	Difference	Level	Difference	Level	Difference
Chow-Lin	.8363	.6694	.8455	.6726	.9502	.6618
	(.7262.9192)	(.5715,.7557)	(.7495,.9184)	(.5699,.7629)	(.8900,.9823)	(.5512,.7579)
Litterman	.8408	.6581	.8515	.6676	.9553	.6540
	(.7169,.9246)	(.5696,.7565)	(.7443,.9287)	(.5564,.7619)	(.8957,.9840)	(.5350,.7597)
Fernandez	.8334	.6700	.8423	.6722	.9482	.6608
	(.7089,.9202)	(.5686,.7603)	(.7474,.9164)	(.5680,.7617)	(.8873,.9820)	(.5516,.7575)
SSC	.9388	.7812	.9420	.7825	.9861	.8012
	(.9045,.9630)	(.7045,.8364)	(.9128,.9642)	(.7098,.8424)	(.9765,.9927)	(.7343,.8554)
MIDAS	.6741	.7097	.6814	.7108	.7979	.6128
	(.5618,.7584)	(.5860,.7894)	(.5824,.7705)	(.5872..8135)	(.6740,.9008)	(.4562,.7445)
MIDAS-AR1	.9819	.9481	.9818	.9470	.9956	.9506
	(.9665,.9913)	(.9039,.9729)	(.9666,.9911)	(.9035,.9728)	(.9916,.9981)	(.9080,.9743)