POOLED MEAN GROUP ESTIMATOR WITH MIDAS COVARIATES

Sven Schaffrath

CPA Ferrere/Department of Economic Analysis - Business Analytics, Montevideo, Uruguay

Abstract

In response to an always increasing availability of large $T$ and $N$ panels of data and taking into account the success of the different MI(xed) DA(ta) S(ampling) aggregation procedures, we propose a new version of Pesaran-Shin-Smith’s Pooled Mean Group Estimator that allows mixing different frequencies on both sides of the equations. We evaluate its performance, focusing in particular on the ability of the model to correctly estimate the shared long-run relationship. We further show that although mixing frequencies leads to a small-sample downward bias on the estimator, it remains consistent and the overall performance of the model is satisfactory, clearing the path for further research on the subject of Dynamic Panel Data models under MIDAS covariates.

The author would like to send a special message of gratitude to Jean-Pierre Urbain and Thomas Gotz for their invaluable support, help and guidance on this task.
1 Introduction

After its introduction by Ghysels et al. (2004a), Ghysels et al. (2004b) and Ghysels et al. (2007) among others, there has been an increasing interest in the (MI)xed (DA)ta (S)ampling time series regression method in both theoretical and applied economic and finance circles. While a standard time series regression involves variables that are sampled at the same frequency, researchers tend to face data sets where the sampling frequencies differ. This poses an additional challenge for the researcher. The common practice is to neglect any additional information concealed in the higher frequency and pre-filter the data by ad-hoc methods so that both sides of the equation have matching frequencies, which normally implies wasting information that could otherwise prove to be useful. MIDAS offers an alternative way. Basically, it introduces a simple, parsimonious and flexible class of time series models that avoids restricting the sampling intervals to be equal by allowing to project the dependent variable $y_t$ onto a history of lagged observations of $x^{(m)}_{t-\frac{j}{m}}$, where $m$ represents how many times faster the covariate is sampled with respect to $y_t$.

Nevertheless, the extension of MIDAS to Panels has been a somewhat unexplored terrain. Khalaf et al. (2012) presents one of the first thorough attempt on extending Ghysels’ approach on Dynamic Panel Data Models. In particular, they extended MIDAS to the Arellano-Bond model and observed very promising performances. Based on their success, we decided to take a step forward and explore how a heterogeneous dynamic panel would react to this modification. Specifically, we work with the Pesaran-Shin-Smith’s Pooled Mean Group Estimator, where the countries/individuals/industries or any other group the analyst focuses on, share a common long-run relationship, but the convergence speeds are allowed to differ among themselves. If the MIDAS-augmented model were to perform adequately, its improved power could be used to retest previously inconclusive relationships and to explore new terrains.

This paper develops the PMG-MIDAS Estimator, where MIDAS covariates are introduced in the standard Pooled Mean Group Estimator model. We test in particular how the model behaves under a slowly decaying Exponential Almon Lag function through Monte Carlo simulations. We find that the modified model performs quite well, although it requires special care on the setting of the initial values for the non-linear estimation procedures. We also find that MIDAS seems to induce an additional downward bias on the estimators for the long-run relationship and the short-run adjustment coefficients. Moreover, this seems to be an increasing function of $m$. Even further, we find that the MIDAS estimation procedure seems to interact with the heterogeneity, inducing further skewness whenever $N > 1$.

Additionally, we found some evidence of consistency. As the dimensions $N$ and $T$ of the data field

---

1 The reader may want to check Ghysels et al. (2004b) for an example on how MIDAS was used to reexplore some previously undecided issues.
increase, the size of the estimation errors decrease strongly and this pattern is observed for every studied \( m \) and estimator.

Lastly, we hypothesize about the nature of the interaction between the long-run relationship, the short-run adjustment rates and the high-to-low aggregation weights. We argue that since all the estimated weight functions tended to be flatter than expected -this is, lower weights on the more recent observations and higher weights on the older ones- part of the information that should have been captured by the first two is indeed captured through the weights belonging to the oldest observations and thus leading to this lower-than-expected estimates for both parameters.

The paper is structured as follows. Section 2 introduces the reader to MIDAS, explores its main attributes and presents some of the current alternatives regarding the different weighting functions that have been used in recent literature, Section 3 shortly reviews the general framework for Heterogeneous Panel Data Models and then immerses the reader into the Pooled Mean Group model, that is circumscribed in the Heterogeneous Dynamic Panel Data class of models and is the second pillar on which this paper is built. Section 4 presents our modified version of Pesaran-Shin-Smith’s model, while Section 5 displays the evaluation methodology and sets the parameters under which the simulation study is done. Finally, Section 6 concludes.

2 Mixed Data Sampling

A standard time series regression usually involves variables that are sampled at the same frequency, but empirical modelers tend to face data sets where the sampling frequencies differ. This poses an additional challenge for the researcher. The common practice is to neglect any additional information concealed in the higher frequency and pre-filter the data so that both sides of the equation have matching frequencies. Nevertheless, as Andreou et al. (2010b) correctly pointed out, there is no a priori reason to ignore the fact that the variables were originated from a mixed frequencies Data Generating Process (DGP) nor to assume an econometric model where the aggregation procedure is restricted only to equal weights, as it is the case with the usual practice of averaging.

The (MI)xed (DA)ta (S)ampling (MIDAS) regressions proposed in Ghysels et al. (2004a), Ghysels et al. (2004b) and Ghysels et al. (2007) offer an alternative to this standard practice. Basically, they introduce a simple, parsimonious and flexible class of time series models that avoid restricting the sampling intervals to be equal by allowing to project the dependent variable \( y_t \) onto a history of lagged observations of \( x_{t-\frac{m}{n}} \), where \( m \) represents how many times faster the covariate is sampled with respect to \( y_t \). Another element that needs to be highlighted is that MIDAS regressions are not autoregressive models, since the concept of autoregression relies on the idea of identical past sampling frequencies.
As mentioned before, MIDAS frees us of the need of following an ad-hoc pre-filtering technique by introducing a data-driven weighting scheme that only depends on very small set of hyperparameters. In this sense, the standard linear MIDAS regression model is

\[ y_t = \beta_0 + \beta_1 B(L^{1/m}; \kappa)x^{(m)}_t + \epsilon^{(m)}_t \]  

(2.1)

where \( B(L^{1/m}; \kappa) = \sum_{j=0}^{J} B(j; \kappa)L^{j/m} \), \( K \) is the length of the polynomial (possibly infinite), \( L^{1/m} \) is a lag operator such as \( L^{1/m}x^{(m)}_t = x^{(m)}_{t-1} \), and \( B(j; \kappa) \) is the weight function only dependent of the small set of hyperparameters \( \kappa \).

Ghysels et al. (2004a) show that since MIDAS regressions exploit all the available information in the finer sampled variable, whereas the pre-filtering does not, MIDAS is always going to lead to improved efficiency. This is clearly related to the fact that an ad-hoc pre-filtering of the series could lead to the potential loss of information and thus render the relationship between variables harder to detect (see Ghysels et al. (2004b), Ghysels et al. (2006) and Ghysels et al. (2007) for empirical examples).

One key element behind MIDAS regressions is the use of tightly parameterized weight functions that restrict the potential parameter proliferation that could be observed otherwise. Therefore, a deeper analysis of some of the available alternatives is required. The following sections address this issue.

### 2.1 Polynomial Specifications

As it was highlighted previously, the key feature behind MIDAS success is the proposition of a tightly parameterized and very flexible functional form that captures the optimal data-driven convex combination of the \( K \) last values of \( x^{(m)}_t \). Current literature (see Ghysels et al. (2007) for a thorough analysis on the subject) introduces several different options, but in practice the Exponential Almon Lag, which is directly related to the Almon Lag used in distributed lag literature (see Almon (1965)), is the most popular alternative amongst researchers. We also introduce the Beta Lag which is based on the Beta function, commonly used in Bayesian econometrics to impose flexible, but still parsimonious prior distributions.

Finally, we present an additional alternative which would allow us to circumvent the numerically intensive non-linear optimization procedures behind any MIDAS regressions. In standard MIDAS regressions this would not pose a major complication, but for the case of the MIDAS augmented Heterogeneous Dynamic Panel that we are trying to explore, the lack of analytical constraints may hinder our ability of obtaining successful results.
2.1.1 Exponential Almon Lag and Beta Lag

The Exponential Almon Lag given by Equation (2.2) is probably one of the simplest representations available for MIDAS weight function $B(j; \kappa)$ and it is related to the Almon Lag literature introduced by Almon [1965]. Although nothing limits $Q$ to be high, the function achieves a very satisfactory level of flexibility just by setting $Q = 2$ and thus ensuring an adequate level of parsimony. Furthermore, it nests the pre-filtering technique of simple arithmetic average. Therefore, whenever it happens to be optimal, simple averages can be obtained by setting $\kappa_1 = \ldots = \kappa_Q = 0$. The function can easily approximate processes with slow and fast declines or increases, as well as hump shaped weighting schemes. Figure 1 presents some examples of shapes the function may take with only two hyperparameters. Moreover, the only condition required to ensure declining weights is that $\kappa_2 \leq 0$.

$$B(j; \kappa) = \frac{e^{\kappa_1 j + \ldots + \kappa_Q j^Q}}{\sum_{j=1}^{K} e^{\kappa_1 j + \ldots + \kappa_Q j^Q}} \tag{2.2}$$

Another element to take into account is that once the functional form is specified, there is no need to concern about lag length selection, since the decline rate determines how many lags are included in the estimation and the former is purely data-driven and simultaneously done with the estimation of the "main" model.

![Figure 1: Exponential Almon polynomial MIDAS weights with different values for $\kappa_1$ and $\kappa_2$. $\kappa_2$ is defined as non-positive to ensure declining weights. Nevertheless, although they are not shown in the figure, increasing weights or simple average are still included in the Exponential Almon Lag family of functions.](image-url)
The second standard alternative present in current literature is the Beta Lag, shown in Equation 2.3, which shares some features with the previous alternative. Although it is usually introduced in the literature as a possible alternative for the Exponential Almon Lag, it has been scarcely used. Some of the exceptions available in current literature are Ghysels (2012), Rodriguez and Puggioni (2010) and Alper et al. (2012). As it was the case with the Exponential Almon Lag, the Beta Lag allows for a very flexible range of functional forms and still nests the simple average of values when \( \kappa_1 = \kappa_2 = 1 \). Furthermore, as long as \( \kappa_2 > 1 \) declining weights are ensured. Faster declining rates may be obtained with an increase in \( \kappa_2 \).

As it was mentioned before, the Beta Lag inherits its flexibility from the Beta distribution (Equation 2.4) which is commonly used in the context of Bayesian econometrics to impose flexible, but parsimonious prior distributions. As it was the case with the Exponential Almon Lag, the effective number of lags included is automatically given by the decline rate and therefore optimal in terms of the observed data. One last element to take into account is that since weights are constrained to be positive and add up to one, the aggregated value of the regressor is not only a convex combination of the lagged values of \( x_t^{(m)} \), but also restricted to a weighted average of them. This is a very desirable feature when the methodology is used to exploit the information available in high frequency data in the context of volatility modelling, as in Ghysels et al. (2004b) and Ghysels et al. (2007) among others.

\[
B(j; \kappa) = \frac{f(\frac{j}{\kappa}, \kappa_1, \kappa_2)}{\sum_{j=1}^{\kappa} f(\frac{j}{\kappa}, \kappa_1, \kappa_2)}
\]  

(2.3)

where:

\[
f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}
\]  

(2.4)

and

\[
\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}dx
\]  

(2.5)

### 2.1.2 Step Functions

The advantage of the standard MIDAS framework is that it allows for a simple, parsimonious parametric representation of the weighting scheme, as it was highlighted when we introduced the Exponential Almon Lag and Beta Lag functions. On the other hand, such approaches require non-linear estimation methods that could be computationally very intensive. This is could be of particular importance for our model.
Figure 2: Beta polynomial MIDAS weights with different values for $\kappa_1$ and $\kappa_2$. $\kappa_2$ is defined as non-positive to ensure declining weights. Slow decay: $\kappa_1 = 1, \kappa_2 = 4$. Fast decay: $\kappa_1 = 1, \kappa_2 = 20$. Hump shaped: $\kappa_1 = 1.6, \kappa_2 = 7.5$.

since as it is going to be shown in the following sections, the model’s estimation procedure is already non-linear and computationally demanding. Another element to take into account, is that even under non-linearity, the two standard weight functions presented above seldomly have a close form solution and require some level of "brute force" approach. Nevertheless, the alternative introduced by Forsberg and Ghysels (2007) and inspired by the HAR model presented by Corsi (2009) and extended by Andersen et al. (2007) would allow us to circumvent this limitation at the cost of some level of parsimony.

Following Ghysels et al. (2007), we define the MIDAS regression with step functions as:

$$y_t = \beta_0 + \sum_{i=1}^{M} \beta_i X_t(Q_i, m) + \varepsilon_t$$

(2.6)

where $X_t(Q, m) = \sum_{j=0}^{Q} x_{t-j/m}$ and $Q_1 < ... < Q_M$. Since $x_{t-j/m}$ appears in every partial sum, its total impact is going to be given by $\sum_{i=1}^{Q_M} \beta_i$. The total effect of $x_{t-j/m}$ for $K_1 < j < K_2$ is determined by $\sum_{i=2}^{Q_M} \beta_i$ and so on. Therefore, the distributed lag pattern is now approximated by a set of step functions, instead of the more parsimonious alternatives introduced previously. Another element to take into account is that, unless we restrict it to be so, there is no guarantee that the weights are going to be positive nor add up to one as it was the previous case.

Nevertheless, if the use of a standard MIDAS approach in the context of our model proved to be
unsuccessful or the computational requirements to steep, this particular approach may reveal itself as a potential fix for the issue.

3 Random Coefficients Panel Data Models

As in Hsiao and Pesaran (2004), let us start considering the usual linear regression model given by

\[ y = \beta'X + u \]  \hspace{1cm} (3.1)

where \( y \) is the dependent variable, \( \beta \) is a \( K \times 1 \) vector of coefficients and \( X \) is \( K \times 1 \) vector of covariates. As usual, \( u \) represents the effects of all the other variables that affect \( y \) and are not included in \( X \). The most common assumption is that this vector behaves like a random variable and is uncorrelated with the set of regressors \( X \). This being said, under a panel data framework, the emphasis is placed over the individual outcomes, highlighting the need for a deeper analysis of this set of "all other variables" included in the error term. The effect of these excluded (and potentially unobservable) factors may be individual specific and/or time varying, implying that there may be some additional information to be modelled and exploited.

The standard approach of a variable intercept assumes that the heterogeneity across units and/or time is due to the effect of omitted variables that are either (i) individual-specific and time-invariant, such as gender, skills and race among others for individuals, and geographical locations and institutional framework for countries, regions or industries, or (ii) time-specific and individual-invariant as it would be the case with elements that would affect every analyzed unit in a similar way. Among the latter, we include variables that may convey information about the general economic context that affect all individuals in the same way, but allowing for changing conditions. Some examples may be interest rates, prices, exchange rates and overall economic sentiment.

The problem with the aforementioned approach is that as it is given, the interaction between the individual and/or the time specific differences and the explanatory variables is ruled out. A more general formulation could be the one given by Equation 3.2

\[ y_{i,t} = \beta_{i,t}'x_{i,t} + u_{i,t} \]

\[ = \sum_{j=1}^{K} \beta_{j,i,t}x_{j,i,t} + u_{i,t} \]  \hspace{1cm} (3.2)

where \( i = 1, ..., N \) and \( t = 1, ..., T \). Equation 3.2 represent the most general formulation for a linear panel data model and it implies that every individual has a specific set of parameters that are also time
variant. Nevertheless, as it is clear, such an approach is very uninformative, lacks explanatory power (it is, at best, purely descriptive), may not be used for prediction nor is it estimable, due to the fact that the number of parameters involved is higher than the available observations.

In order to turn such a model into an estimable and informative model, it is critical to impose further structure on the parameters of interest. One way to do so is to assume that

$$\beta_{k,i,t} = \beta_k + \alpha_{k,i} + \lambda_{k,t}$$

(3.3)

where $\alpha_{k,i}$ and $\lambda_{k,t}$ are two random variable with well defined moments. This is what is commonly named "random coefficient" model, which allows us to significantly reduce the number of parameters to be estimated, while still allowing for parameter heterogeneity among individuals and/or the time dimension.

This "random coefficient" model may be thought as the general version of a dynamic random coefficient model, which in turn contains the dynamic heterogeneous panel with homogeneous long-run relationships introduced in Pesaran et al. (1997) and Pesaran et al. (1999) that are the main focus of this paper and the following sections.

3.1 The Pooled Mean Estimator - Pesaran-Shin-Smith Framework

Concomitantly with an increasing availability of large $N$ and $T"data fields", term coined by Quah (1990), there has been an upsurge in the interest for analyzing long-run relationships between groups, such as firms, industries, regions and countries. These groups usually show some level of heterogeneity, but simultaneously share a good level of commonality that could potentially transpire to similar specifications and parameters.

When these data fields are available, the common practice for identifying and quantifying the common patterns of response consists basically of two alternatives. On one hand, we have a sort of two staged estimation procedure where one equation is estimated for every unit separately and then once every model has been estimated, its parameter’s distribution and its moments are examined. The standard technique involves estimating the averages of the parameters and its referred as the Mean Group Estimator. Pesaran and Smith (1995) show that even under $N$ independent estimations, the aforementioned average will deliver a consistent estimate of the parameters’ real means. The main issue with this approach is that since the information conveyed in the panel dimension is not being exploited, the estimator will not be able to take into account that some of the parameters could actually be the same across units, meaning that all parameters, intercepts, short-run coefficients, long-run coefficients and variances will be allowed to differ across groups, even when that is not the case, thus leading to a decrease in the
efficiency of the estimator.

The second alternative includes the traditional pooled estimators, such as the fixed and random effect estimators, where the intercepts are allowed to differ, but all the slopes are assumed to be identical across individuals. As discussed in Pesaran and Smith (1995), one of the main obstacles for this approach is that unless the slopes of the dynamic panel data model are in fact identical, these standard estimation procedures may produce inconsistent and possibly very misleading estimates of the average values of the parameters. Since the aforementioned assumption seems to be in most cases excessively strong and that identical slopes across groups is often rejected by the available battery of test, the relevance of an intermediate path is self-evident.

Pesaran et al. (1997) introduce an alternative where some degree of heterogeneity in the slopes is allowed, but a good level of long-run commonality is assumed, avoiding the very restrictive assumption of identical slopes and the potentially excessive generalization of totally unrelated parameters. Specifically, Pesaran et al. (1997) consider a model where intercepts, short-run coefficients and error variances are allowed to vary freely, but constrains the long-run coefficient to be the same across groups. This proposal is of particular interest in several plausible economic applications, where there may be good reasons to assume that in the long-run common technologies, budget constrains and arbitrage conditions will influence the groups in similar ways, leading in turn to relatively close reaction patterns. This being said, there is no strong reason to assume that the short-run dynamics and speed of convergence must coincide.

3.1.1 The Model

Let us start assuming that a data field of dimensions $t = 1, ..., T$ and $i = 1, ..., N$ is readily available with which we intend to estimate an $ARDL(p, q, ..., q)$ model of the form

$$y_{i,t} = \sum_{j=1}^{p} \lambda_{i,j} y_{i,t-j} + \sum_{j=0}^{q} \delta_{i,j}' x_{i,t-j} + \gamma_{i}' d_{t} + \varepsilon_{i,t} \quad (3.4)$$

where $x_{i,t}(k \times 1)$ and $d_{t}(s \times 1)$ are vectors of explanatory variables, the former being free to vary across time and groups and the latter restrained to include nothing but regressors that differ only across time periods. As is it to be expected $T$ has to be large enough to allow us to correctly estimate the model for each group, but needs not to coincide through the different units as it is also the case with the order of the lag polynomial of each regressor. Nevertheless, just for reasons linked to simplicity and clarity of exposition we will work with equal $T$ and $q$. 
As in Pesaran et al. (1997), we focus on re-parameterizing 3.4 as an error-correction model of the form

\[ \Delta y_{i,t} = \phi_i y_{i,t-1} + \beta_i x_{i,t} + \sum_{j=1}^{p-1} \lambda_{i,j}^* \Delta y_{i,t-j} + \sum_{j=0}^{q-1} \delta_{i,j}^* \Delta x_{i,t-j} + \gamma_i d_t + \varepsilon_{i,t} \]  

(3.5)

where \( \phi_i = -(1 - \sum_{j=1}^{p} \lambda_{i,j}) \), \( \beta_i = \sum_{j=0}^{q} \delta_{i,j}^* \), \( \lambda_{i,j}^* = -\sum_{m=j+1}^{p} \lambda_{i,m} \), \( j = 1, ..., p - 1 \), and \( \delta_{i,j}^* = -\sum_{m=j+1}^{q} \delta_{i,m} \), \( j = 1, ..., q - 1 \), \( i = 1, ..., N \).

Furthermore, after stacking over the time dimension for every group, 3.5 may be rewritten as

\[ \Delta y_i = \phi_i y_{i,-1} + X_i \beta_i + \sum_{j=1}^{p-1} \lambda_{i,j}^* \Delta y_{i,-1} + \sum_{j=0}^{q-1} \Delta X_{i,-j} \delta_{i,j}^* + D \gamma_i + \varepsilon_i \]  

for \( i = 1, ..., N \),

(3.6)

where \( y_i \) is a \( T \times 1 \) vector of dependent variables, \( X_i \) is a \( T \times k \) matrix of time/group-variant regressors and \( D_i \) is a \( T \times s \) matrix of fixed regressors such as intercepts, time trends and/or other time-varying explanatory variables. Moreover, \( y_{i,-j} \) and \( X_{i,-j} \) are the \( j \) period lagged versions of \( y_i \) and \( X_i \).

Three basic assumptions are to be made.

**Assumption 3.1** (iid-ness of error terms) The disturbances \( \varepsilon_{i,t} \), \( i = 1, ..., N \), \( t = 1, ..., T \) in 3.6 are independently and normally distributed across \( i \) and \( t \), with zero means and variances \( \sigma_i^2 \) and finite fourth moments. Furthermore, \( \varepsilon_{i,t} \) are independently distributed of regressors \( x_{i,t} \) and \( d_t \).

Independence of error terms across groups could be achieved by modelling explicitly any cross dependence between units, through the inclusion of common factors as an additional element in \( d_t \) and thus removing it from the residuals. In Pesaran and Smith (1995) for example, this procedure is followed by including a regressor that was defined as the sum over groups of one element in \( x_{i,t} \), but this was possible only because the common factor was itself of an observable nature. If this common factor were to be unobservable, it may be possible to write the data as deviations from the yearly means, thus reducing the impact of the common yearly specificities, although its complete elimination would depend on the slope coefficients being identical across groups. Additionally, as it was just discussed, the time dependence may also be modelled just by increasing \( p \) and \( q \) in the ARDL\((p,q,...,q)\) model. Lastly, a more general specification of 3.1 may drop the normality requirement, but we will keep it in order to follow a likelihood approach where a distributional assumption is also required.

**Assumption 3.2** (dynamic stability of the model) The roots of the polynomial

\[ \sum_{j=1}^{p} \lambda_{i,j} z^j = 1 \]  

for \( i = 1, ..., N \)
lie all outside the unit circle, implying that the ARDL\((p,q,\ldots,q)\) model given by 3.4 is stable.

Assumption 3.2 ensures that \(\phi_i < 0\) and the existence of a long-run relationship between \(y_{i,t}\) and \(x_{i,t}\) of the type
\[
y_{i,t} = -\left(\frac{\beta_i}{\phi_i}\right)x_{i,t} + \eta_{i,t},
\]
for each \(i = 1, \ldots, N\), where \(\eta_{i,t}\) represents a stationary process.

**Assumption 3.3** (long-run homogeneity) The long-run coefficients on \(X_i\), defined by \(\theta_i = -\frac{\beta_i}{\phi_i}\), are the same across groups. Namely,
\[
\theta_i = \theta, \quad (3.7)
\]
for each \(i = 1, \ldots, N\).

Combining assumptions 3.2 and 3.3 and equation 3.6 we could re-write the latter as an error-correction model of the form
\[
\Delta y_i = \phi_i \xi_i(\theta) + W_i \kappa_i + \varepsilon_i, \quad i = 1, \ldots, N, \quad (3.8)
\]
where
\[
\xi_i(\theta) = y_i,-1 - X_i \theta, \quad i = 1, \ldots, N, \quad (3.9)
\]
is the error correction component and \(W_i = (\Delta y_{i,-1}, \ldots, \Delta y_{i,-p+1}, \Delta X_i, \ldots, \Delta X_{i,-q+1}, D)\) and \(\kappa_i = (\lambda_i^*, \ldots, \lambda_{i,p-1}^*, \delta_i^*, \ldots, \delta_{i,q-1}^*, \gamma_i^*)'\).

Since we will focus only in the long-run parameters we prefer to work with the concentrated log-likelihood function of the model. Following assumption 3.1 the log-likelihood function follows
\[
\ell_T(\psi) = -\frac{T}{2} \sum_{i=1}^{N} \ln 2\pi \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (\Delta y_i - \phi_i \xi_i(\theta))' H_i (\Delta y_i - \phi_i \xi_i(\theta)), \quad (3.10)
\]
where \(\psi = (\theta', \phi', \sigma')\), \(H_i = I_T - W_i (W_i' W_i)^{-1} W_i'\) and \(I_T\) is an identity matrix of size \(T\).

Provided the standard identification conditions are met and the regressors are not cointegrated among themselves\(^2\) the estimators presented in the following section will be consistent and asymptotically normal.

---

\(^2\)See assumptions 2.4 and 2.5 in Pesaran et al. (1997) for further detail.
3.1.2 The Pooled Mean Group Estimator

Maximizing $\ell_T(\psi)$ with respect to $\psi$ will lead to the computation of the optimal estimators for $\theta$ and the specific short-run adjustment coefficients, $\phi_i$. Pesaran et al. (1997) name these functions as the "pooled mean group" estimators in order to highlight, on one hand, the commonality implicit in the pooling procedure that is condensed in the long-run coefficient $\theta$ and, on the other, the fact that the error-correction coefficients and the rest of the short-run dynamics parameters are estimated using group-wide averages.

The maximization procedure leads to the following estimators

$$
\hat{\theta} = -\left\{ \sum_{i=1}^{N} \frac{\phi^2_i}{\hat{\sigma}^2_i} X_i' H_i X_i \right\}^{-1} \left\{ \sum_{i=1}^{N} \frac{\phi_i}{\hat{\sigma}^2_i} X_i' H_i \left( \Delta y_i - \hat{\phi}_i y_{i,-1} \right) \right\}, \tag{3.11}
$$

$$
\hat{\phi}_i = \left( \hat{\xi}_i' H_i \hat{\xi}_i \right)^{-1} \hat{\xi}_i' H_i \Delta y_{i,i}, i = 1, ..., N, \tag{3.12}
$$

$$
\hat{\sigma}^2_i = T^{-1} \left( \Delta y_i - \hat{\phi}_i \hat{\xi}_i \right)' H_i \left( \Delta y_i - \hat{\phi}_i \hat{\xi}_i \right), i = 1, ..., N \tag{3.13}
$$

Although the model may be estimated under stationary and non-stationary regressors, we focus only on the stationary case. For a thorough discussion on the non-stationary case we refer the reader to section 3.2 in Pesaran et al. (1997).

As mentioned previously, under assumptions [3.1-3.3] the fairly standard regularity conditions implied by assumptions 2.4 and 2.5 in Pesaran et al. (1997) and stationarity of regressors $X_i$, the estimators provided by equations [3.11] and [3.12] will be consistent. Furthermore, under a fixed $N$ and $T \to \infty$, the Pooled Mean Group estimator will have the following asymptotic distribution:

$$
\sqrt{T} (\hat{\psi} - \psi) \sim N\{ 0, J^{-1}(\psi) \}, \tag{3.14}
$$

where $J^{-1}(\psi)$ is the information matrix defined by A.37 in Pesaran et al. (1997).

4 The Pooled Mean Estimator with MIDAS Covariates

The main goal of this paper is to further develop the Pooled Mean Estimator introduced by Pesaran et al. (1997) by allowing covariates to be sampled at higher frequencies than the ones used to sample the dependent variable. The prime contribution of such model is that it should lead to an increase in the
efficiency of the estimators by exploiting all the information available to the analyst. Nevertheless, since MIDAS estimators under Panels are still under scrutiny, its use may also carry some unexpected costs and effects. Therefore, section 5 analyzes the behaviour and performance of the model and proposed estimation procedure through the implementation of a Monte Carlo simulation study.

As discussed in section 2, there are several paths that lead to mixing frequencies. Our study will focus on the standard Exponential Almon Lag function. The main advantage of such function is that it would allow us to represent the \( m \) weights with only two parameters, \( \kappa_1 \) and \( \kappa_2 \). Nevertheless, this simplicity may come at a cost. Functions like the Exponential Almon Lag and the Beta Lag hinder our ability to obtain close-form solutions for the hyperparameters. In more standard MIDAS frameworks the lack of close-form solutions may not pose a significant difficulty, but in the context of PMG-MIDAS model, where the original model was already non-linear with cross-equation restrictions present, the lack of an evident solution for the set of hyperparameters \( \kappa \) may indeed create some additional problems. Two possible paths may be followed in order to mitigate the effect of this issue. The first path involves the straightforward estimation of the weights under an unconstrained framework as in Foroni et al. (2011). The problem with this alternative is that it may go against our basic premise of parsimony. For example, regressing a yearly variable on only one monthly regressor would imply that we should estimate 12 unconstrained weights, while if we were to follow an Exponential Almon Lag approach it would only require the estimation of the two aforementioned hyperparameters. An intermediate path was suggested in section 2.1.2, where we introduced the step function alternative, which is a weighting function that would allow us to maintain a certain level of parsimony, while still leading to close-form solutions for the weights.

The modified version of Pesaran’s model is introduced in section 4.1, the estimators and the estimation procedure is presented in section 4.2 and section 4.2.1 discusses shortly about how an unconstrained weight estimator would look like and when it could prove itself useful.

4.1 The Model

As in section 3.1.1, let us start assuming that we want to regress \( y \), a \( T \times N \) matrix, on a variable that is sampled \( m \) times more frequently in every group\(^4\). Additionally, let us assume that there is good reason to believe that this two variables are linked by a long-run relationship that may be common for all groups.

\(^3\)Whenever the actual order of the function is indeed 2. Although this is the standard case, nothing restricts \( Q \) to be higher than 2.

\(^4\)There is no a priori reason, except for simplicity, for not considering group-specific \( m \), but in the context of this paper we will restrain ourselves to that specific case.
As in section 3.1.1 we would like to estimate an ARDL($p, q, ..., q, m$)

$$y_{i,t} = \sum_{j=1}^{p} \lambda_{i,j} y_{i,t-j} + \sum_{j=0}^{q} \delta_{i,j} \tilde{x}_{i,t-j} + \gamma_{i} d_{t} + \varepsilon_{i,t}, \quad (4.1)$$

where $m$ denotes how many times faster the explanatory variables are sampled with respect to the dependent variable, $\tilde{x}_{i,t} = \sum_{j=0}^{m-1} B(\kappa; j) x_{i,t-\frac{j}{m}}$ and the rest of the parameters coincide with the definitions in sections 2 and 3.1.1.

Following the same line of thought of re-parameterizing and stacking, we arrive to the MIDAS counterpart of equation 3.6

$$\Delta y_{i} = \phi_{i} y_{i,-1} + \tilde{X}_{i} \beta_{i} + \sum_{j=1}^{p-1} \lambda_{i,j}^{\ast} \Delta y_{i,-1} + \sum_{j=0}^{q-1} \Delta \tilde{X}_{i,-j} \delta_{i,j}^{\ast} + D \gamma_{i} + \varepsilon_{i} \text{ for } i = 1, ..., N, \quad (4.2)$$

As before we hold mixed-frequencies versions of assumptions 3.1-3.3, which may be explicitly stated as follows.

**Assumption 4.1** The disturbances $\varepsilon_{i,t}^{(m)}$, $i = 1, ..., N$, $t = 1, ..., T$ in (4.2) are independently and normally distributed across $i$ and $t$, with zero means and variances $\sigma_{i}^{2}$ and finite fourth moments. Furthermore, $\varepsilon_{i,t}^{(m)}$ are independently distributed of regressors $x_{i,t}$ and $d_{t}$.

**Assumption 4.2** The roots of the polynomial

$$\sum_{j=1}^{p} \lambda_{i,j} z^{j} = 1 \text{ for } i = 1, ..., N$$

lie all outside the unit circle, implying that the ARDL($p, q, ..., q, m$) model given by (4.1) is stable.

**Assumption 4.3** The long-run coefficients on $\tilde{X}_{i}$, defined by $\theta_{i} = -\beta_{i} / \phi_{i}$ are the same across groups. Namely,

$$\theta_{i} = \theta, \quad (4.3)$$

for each $i = 1, ..., N$.

As before, assumptions 4.2 and 4.3 imply (i) the existence of a long-run relationship for each group and (ii) that these relationships are common across the units. In other words, all groups are assumed to converge to a long-run equilibrium, independently of the level of heterogeneity present in the adjustment rates.
Combining the aforementioned assumptions and equation 4.2 we get

\[ \Delta y_i = \phi_i \xi_i(\theta, \kappa) + W_i(\kappa) \Pi + \varepsilon_i, \quad i = 1, \ldots, N, \]

(4.4)

where

\[ \xi_i(\theta, \kappa) = y_{i,-1} - \bar{X}(\kappa), \quad i = 1, \ldots, N, \]

(4.5)

is the new MIDAS counterpart of the error-correction component introduced in \(3.9\), and \( W_i(\kappa) = (\Delta y_{i,-1}, \ldots, \Delta y_{i,-p+1}, \Delta \bar{X}(\kappa), \ldots, \Delta \bar{X}(\kappa), \cdots) \)

\[ \Pi_i = (\lambda_i^{*1}, \ldots, \lambda_i^{*p-1}, \delta_i^{*0}, \ldots, \delta_i^{*q-1}, \gamma_i) \quad 5 \]

The concentrated log-likelihood also looks very familiar, the inclusion of the new set of hyperparameters being the only difference with respect to the original equation 3.10.

\[ \ell_T(\psi) = \frac{-T}{2} \sum_{i=1}^{N} \ln 2\pi \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{N} 1 \sigma_i^2 (\Delta y_i - \phi_i \xi_i(\theta, \kappa))^\prime H_i (\Delta y_i - \phi_i \xi_i(\theta, \kappa)), \]

(4.6)

where \( \psi = (\theta', \phi', \sigma', \kappa), H_i = I_T - W_i(W_i'W_i)^{-1}W_i' \) and \( I_T \) is an identity matrix of size \( T \).

### 4.2 The Estimators

Optimizing the log-likelihood function in equation 4.6 w.r.t. the different elements of \( \psi \) leads to close-form solutions for the estimators of \( \theta, \phi \) and \( \sigma \) but not for \( \kappa \) where we may only get an implicit function at best. As mentioned before, this is a common issue for MIDAS estimators and leads to the need for numerical optimization methods. In the context of simpler applications, this "brute force" approach has limited effects over the results, but in this context, where the original model was already non-linear and shows cross-equation restrictions, the effects may not be easily tractable beforehand.

In consequence, we also present in section 4.2.1 an alternative set of estimators inspired by the U-MIDAS estimators introduced by Foroni et al. (2011), where we disregard the hyperparameters and focus solely on the estimation of a set \( \omega \) of \( m \) weights.

The estimators that have clear, concise, close-form solutions are

\[ \hat{\theta} = -\left\{ \sum_{i=1}^{N} \frac{\phi_i^2}{\sigma_i^2} \bar{X}_i' \bar{H}_i \bar{X}_i \right\}^{-1} \left\{ \sum_{i=1}^{N} \frac{\phi_i}{\sigma_i^2} \bar{X}_i' \bar{H}_i (\Delta y_i - \hat{\phi}_i y_{i,-1}) \right\} \]

(4.7)

\[ \hat{\phi}_i = \left\{ \hat{\xi}_i' \bar{H}_i \hat{\xi}_i \right\}^{-1} \left\{ \hat{\xi}_i' \bar{H}_i (\Delta y_i) \right\} \text{ for } i = 1, \ldots, N. \]  

(4.8)
\[ \hat{\sigma}_i^2 = \frac{1}{T} \hat{\epsilon}_i' \hat{H}_i \hat{\epsilon}_i \text{ for } i = 1, \ldots, N. \]  

(4.9)

where the matrices branded with a tilde are the result of the MIDAS aggregation procedure\footnote{Since \( \xi_i \) and \( \varepsilon_i \) are functions of \( \tilde{X}_i \) they also depend on the aggregation procedure, but just for the sake of simplicity we prefer omitting this particular branding.}. The derivations are displayed in section A.2 of the Appendix.

Given the non-linear nature of the model and the cross-dependence of the parameters it is clear that the estimation procedure has to be based on numerical optimization methods, such as "back-substitution" algorithms or the standard Newton-Raphson algorithm. Independently of the method chosen, we would recommend a special care on the set-up of the initial values and a constrained optimization imposing the close-form solutions for the estimators as restrictions to ensure a convergence of improved quality.

### 4.2.1 The U-MIDAS Case

As implied in the previous pages of this paper, the lack of a close-form solution for the hyperparameters leads to the non-existence of a parsimonious close-form solution for the weights. Nevertheless, whenever \( m \) and the number of regressors involved in the estimation procedure are small, the cost of focusing directly on the particular weights is not excessively onerous and therefore solutions as the proposed by Foroni et al. (2011) are feasible.

As the reader may have noticed, since \( \theta \) is estimated regressing \( \Delta y \) on functions of \( \tilde{X} \), which is in turn the result of the aggregation procedure, an inadequate estimation of the weights \( \omega \) will lead to biases and possibly inconsistencies in the estimation of \( \theta \). This is of particular importance since \( \theta \) represents the long-run relationship, probably the most important parameter for the researcher applying a technique of this type.

In order to highlight this point, let us consider again the term \( \tilde{X}, \theta \) in equation 4.3 and define \( X_{i,.}^{(m)} = \{x_{i,1-\frac{1}{m}}, \ldots, x_{i,T-\frac{1}{m}}\}' \). In other words, \( X_{i,.}^{(m)} \) is no more than the collection of the high frequency elements \( \frac{j}{m} \)-high-frequency-periods behind for every \( t = 1, \ldots, T \). Combining the latter definition with the fact that \( \tilde{x}_{t,j} = \sum_{j=0}^{m-1} B(\kappa; j) x_{i,t-\frac{j}{m}}^{(m)} \) and letting \( \omega_j = B(\kappa; j) \), we can re-express the term in question as

\[
\tilde{X}_i \theta = \sum_{j=0}^{m-1} \omega_j X_{i,.}^{(m)} \theta^{*} = X_{i,.}^{(m)} \theta^{*}
\]  

(4.10)
where \( \theta^* = \{ \omega_0 \theta, ..., \omega_{m-1} \theta \} \). This result highlights the relationship between the original PMG model and the PMG-(U-)MIDAS. Our version is a very close cousin of the original one. In the current U-MIDAS approach, each formerly aggregated variable, \( \tilde{X}_i \), is now separated in \( m \) different regressors collecting only the \( t - \frac{j}{m} \) element for every \( t = 1, ..., T \).

It should be noted that this decomposition in \( m \) different regressors is also valid for the cases where a weighting function is being estimated explicitly. The main difference is given by the fact that the parameter estimation will still focus on \( \kappa \) and \( \theta \) separately and not on \( \theta^* \). Nevertheless, close-form solutions for \( \omega_j \) are available when we disregard the estimation of the different parsimonious weighting functions and specifically, the estimation of \( \kappa \). The estimators for \( j \) ranging from 0 to \( m - 1 \) are

\[
\hat{\omega}_j = - \left\{ \frac{N \sum_{i=1}^N \hat{\phi}_i \hat{X}^{(m)}_i \hat{H}_i X^{(m)}_i \hat{\theta}}{\sum_{i=1}^N \frac{\hat{\phi}_i^2 \hat{\theta}'}{\sigma_i^2} X^{(m)}_i X^{(m)}_i} \right\}^{-1} \left\{ \frac{N \sum_{i=1}^N \hat{\phi}_i \hat{\theta}' X^{(m)}_i \hat{H}_i \hat{\Psi}_i}{\sum_{i=1}^N \frac{\hat{\phi}_i^2 \hat{\theta}'}{\sigma_i^2} X^{(m)}_i X^{(m)}_i} \right\} (4.11)
\]

where \( \hat{\Psi}_i = [\Delta y_i - \hat{\phi}_i (y_{i,-1} - \hat{\phi}_i \tilde{X}^{(m)*}_{i,-\frac{1}{m}})] \).

5 Simulation Study

Once the model is set up, we proceed to its evaluation. As the standard practice suggest, we evaluate its performance by Monte Carlo simulations, where a given number of data fields are generated using an adequate DGP and then the estimation procedure is applied. Once the estimates are gathered, we may analyze their distribution and compare them to the real value of the parameter. Section 5.1 presents the procedure followed and sets the main parameters of interest. Section 5.2 introduces our main results about the general evaluation of the PMG-MIDAS model under an Exponental Almon Lag weighting scheme.

5.1 General Setting

The simulations are based on equation \[4.1\] and on the DGP for the high frequency variable provided by

\[
x^{(m)}_{i,t} = \alpha_i + \rho_i x^{(m)}_{i,t-\frac{1}{m}} + \eta_{i,t},
\]

where for simplicity \( \alpha_i \) is assumed to be equal to zero for every \( i \), \( \rho_i = 0.8 \) and \( \eta_{i,t} \sim N(0,1) \). In Khalaf et al. (2012), they preferred to set a DGP for \( x^{(m)}_{i,t} \) where the partial autocorrelation is set to be with respect to \( x^{(m)}_{i,(t-1)-\frac{1}{m}} \), instead of with respect to \( x^{(m)}_{i,t-\frac{1}{m}} \). The difference is that the former ensures the autocorrelation to be independent of the weighting functions and associated parameters, while our
choice lacks of such an advantage. On the other hand, we think that our DGP may better reflect real conditions and real data. In any case, we think that the impact of such choice should not affect greatly the final results.

Furthermore, while Khalaf et al. (2012) simulate different sets of combinations for $\kappa$, we restrict ourselves to a slowly decaying Exponential Almon Lag function given by $\kappa = \{0.1, -0.2\}$. This answers mainly to computation-time needs. The PMG model is already very onerous in terms of computational requirements, so when we add on top of it the need to numerically fit $\kappa$ its time consumption grows exponentially.

Once $x_{i,t}^{(m)}$ and $\tilde{X}_i$ are available we proceed to generate $y_i$ by following expression 4.1 and setting $\sigma_i^2 = 1, \mu_i = 0$ and $\varepsilon_{i,t} \sim N(\mu_i, \sigma_i^2)$. A special note on how the slope heterogeneity is achieved should be made. In particular, we started by assuming $\theta = \{2.5\}$ and $\lambda_i \sim \text{Unif}(0.4, 0.8)$. Then, we generated the different $\beta_i$ following $\theta_i = \frac{\beta_i}{\beta_i} = \frac{\beta_i}{1-\lambda_i} = \theta \Rightarrow \beta_i = (1-\lambda_i)\theta$.

We mentioned before that we found that our method is particularly sensitive to the set of initial values. We started trying a standard grid search approach were we chose the set that lead to the maximum value for the initial likelihood. The problem with the latter was that it tended to find optimal combinations of values that were too far away of the real set of parameters. We observed for example that the weighting functions tended to be skewed to the oldest observations instead of returning the expected slowly decaying function. In response to that, $\hat{\theta}_0$ compensated this by choosing the lowest possible values.

Although we did not test it properly, we argue that this could be a consequence of a very flat likelihood function around the neighbourhood of the real set of parameters and the existence of a very steep slope that favours very high levels for $\hat{\kappa}_{2,0}$ and very low $\hat{\theta}_0$. This was mostly observed for the first set of values delivered by the grid search procedure. In response, we decided to increase the robustness of our initial values selection procedure by selecting the median of the 15 best combinations suggested by the grid search procedure. Luckily, this seems to work very well, although there is no particular reason for choosing 15 solutions instead of 10 or 30.

Next, we estimate the model by a iterative constrained optimization procedure, assuming $\sigma_i > 0$ and $\phi_i < 0$ and restricting the solutions to be close to the analytical solutions when available.

This procedure was repeated for several configurations of $m$, $N$ and $T$. For $m$, we chose three different cases, $m = \{1, 3, 10\}$. The first case should provide a benchmark to compare the effects of aggregation on the PMG model, while the second and third cases focus on a low and a high number of intra-period observations of the regressors. To test the effects of sample size over the estimation of the parameters, we tried for every $m$, $N = \{1, 10, 25, 50\}$ and $T = \{25, 50, 100, 200\}$. We favored long panels, because we were mainly interested on the dynamic dimension of the model. Next section presents our main findings.
and a general evaluation of our model. Section 6 concludes and highlights some potential lines of further research.

5.2 Results

Figures 3 to 5 compare the estimated empirical density functions of the deviations of each simulated estimator with respect to the null hypothesis. The figures show a panel of graphs in which the three different densities (one for each \( m \)) are compared between each other. Since the horizontal axis share the same range, visual comparison of the density functions is allowed and very informative. Following, Tables 1 to 3 present some interesting descriptive statistics on the distribution of the aforementioned deviations. Lastly, Table 4 compares the implied weight functions for several cases with respect to the actual weighting scheme.

Firstly, as it was to be expected, the estimators seem to remain consistent. Both \( \hat{\theta} \) and \( \hat{\phi}_i \) show, for a given \( m \) decreasing standard deviations as \( N \) and \( T \) increase. For example, \( \sigma_{\epsilon_{\hat{\theta}}} \) goes from 0.721 to 0.017 for the cases of \( m = 1, N = 1, T = 25 \) and \( m = 1, N = 50, T = 200 \) respectively. The former observation is confirmed by a concomitant decrease of the same style in the interquartile range implied in Tables 1 and 2. Furthermore, this pattern repeats itself for every \( m \) and for every analyzed estimator.

Nevertheless, this gain in precision may hide some consequences of the aggregation under heterogeneous short-run adjustment speeds, which leads to our second observation. MIDAS seems to exacerbate the downward bias present in the standard PMG model. For example, when comparing the average of \( \epsilon_{\hat{\theta}, N=1, T=25} \) under the different values of \( m \), we observe that the means start at -0.069 for \( m = 1 \), then decrease to -0.072 for \( m = 3 \) and end up in -0.114 for \( m = 10 \). Figure 3 suggests that this is directly linked to the introduction of a downward skewing force on the densities. Furthermore, the skewness is greatly empowered when \( N > 1 \), suggesting that there may be some interaction between the estimation of \( \kappa \) and the heterogeneity in the slopes. The same may be said about \( \epsilon_{\hat{\phi}_i} \), where the same type of behaviour is observed (as implied in Figure 4 and Table 2).

As it was to be expected, the bias in the estimates of the long-run relationship and the group-specific short-run adjustment coefficients, lead to positively skewed set of empirical densities of \( \epsilon_{\hat{\omega}_i} \).

Lastly, Table 4 studies what happens with the estimation of the hyperparameters, which play a central role in our framework. The top half of the table displays the mean estimates for \( \kappa_1 \) and \( \kappa_2 \) and the implied \( \omega_j \). In favour of succinctness, we focused on the first three weights and added up the rest. The bottom half shows the difference between the implied estimated weight and the actual value. We focus solely on the case of \( m = 10 \), since the aforementioned bias in \( \hat{\phi}_i \) and \( \hat{\theta} \) seems to be an increasing function of \( m \) and the effects over the weights should be clearer in that case than in the other. The real weights
are 51.7%, 31.4%, 12.7% and 4.2% for $\omega_1$, $\omega_2$, $\omega_3$ and $\sum_{j=4}^{10} \omega_j$ respectively. The first three estimated weights are always lower for the six cases presented in Table 4, but the rest are always higher, fact that implies that our estimated weight functions tend to be flatter than it should be or at least, somewhat more skewed to the oldest data. The latter is especially the case when $N > 1$ and the slope-heterogeneity starts disrupting the estimates. Nevertheless, there is some evidence in Table 4 that the size of the error diminishes as the data-field increases its dimension. For example, while the added estimated weight of the last 6 high-frequency observations was 8.9 percentage points higher than it should be when $N = 50$ and $T = 100$, when $T = 200$ the deviation decreases to 7.9 percentage points. The same pattern is observed (in absolute values) for almost all the cases displayed in the bottom half of the aforementioned table.

Although further research is required in order to understand more deeply the nature of the interaction between the weights and the other estimates, we hypothesize that some information on the dynamics and the time-dependence that should have been captured by $\hat{\theta}$ and $\hat{\phi}_i$ is imbued in the weight function. In other words, we suspect that some correlation with $y_{t-1}$ is actually captured through the "oldest" weights $\omega_j$, thus leading to lower-than-expected $\hat{\phi}_i$'s.
Figure 3: $\epsilon_\hat{\theta} = \hat{\theta} - \theta$ for $m = \{1, 3, 10\}$ (solid, dash and dash-dot lines respectively). 2,000 simulations, Epanechnikov kernel function.
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D</th>
<th>Median</th>
<th>25% Perc.</th>
<th>75% Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=1</td>
<td>T=25</td>
<td>-0.069</td>
<td>0.721</td>
<td>-0.103</td>
<td>0.243</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.037</td>
<td>0.396</td>
<td>-0.039</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.028</td>
<td>0.237</td>
<td>-0.019</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.010</td>
<td>0.160</td>
<td>0.000</td>
<td>-0.102</td>
</tr>
<tr>
<td>N=10</td>
<td>T=25</td>
<td>-0.042</td>
<td>0.160</td>
<td>-0.044</td>
<td>0.148</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.025</td>
<td>0.205</td>
<td>-0.015</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.009</td>
<td>0.057</td>
<td>0.000</td>
<td>-0.046</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.004</td>
<td>0.040</td>
<td>-0.005</td>
<td>0.029</td>
</tr>
<tr>
<td>N=25</td>
<td>T=25</td>
<td>-0.031</td>
<td>0.105</td>
<td>-0.035</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.015</td>
<td>0.056</td>
<td>-0.014</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.007</td>
<td>0.037</td>
<td>-0.007</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>0.005</td>
<td>0.022</td>
<td>0.000</td>
<td>-0.019</td>
</tr>
<tr>
<td>N=50</td>
<td>T=25</td>
<td>-0.033</td>
<td>0.086</td>
<td>-0.038</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.014</td>
<td>0.041</td>
<td>-0.013</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.007</td>
<td>0.026</td>
<td>-0.008</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>0.004</td>
<td>0.017</td>
<td>-0.004</td>
<td>0.016</td>
</tr>
<tr>
<td>N=100</td>
<td>T=25</td>
<td>-0.072</td>
<td>0.752</td>
<td>-0.115</td>
<td>0.480</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.037</td>
<td>0.452</td>
<td>-0.041</td>
<td>0.297</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.018</td>
<td>0.303</td>
<td>-0.013</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.022</td>
<td>0.193</td>
<td>-0.014</td>
<td>0.117</td>
</tr>
<tr>
<td>N=200</td>
<td>T=25</td>
<td>-0.279</td>
<td>0.351</td>
<td>-0.229</td>
<td>-0.393</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.170</td>
<td>0.156</td>
<td>-0.151</td>
<td>-0.244</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.134</td>
<td>0.086</td>
<td>-0.125</td>
<td>-0.186</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.120</td>
<td>0.059</td>
<td>-0.118</td>
<td>-0.160</td>
</tr>
<tr>
<td>N=300</td>
<td>T=25</td>
<td>-0.250</td>
<td>0.391</td>
<td>-0.192</td>
<td>-0.299</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.135</td>
<td>0.085</td>
<td>-0.131</td>
<td>-0.185</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.111</td>
<td>0.054</td>
<td>-0.109</td>
<td>-0.146</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.100</td>
<td>0.038</td>
<td>-0.099</td>
<td>-0.124</td>
</tr>
<tr>
<td>N=400</td>
<td>T=25</td>
<td>-0.175</td>
<td>0.216</td>
<td>-0.152</td>
<td>-0.224</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.117</td>
<td>0.062</td>
<td>-0.113</td>
<td>-0.159</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.097</td>
<td>0.043</td>
<td>-0.097</td>
<td>-0.125</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.085</td>
<td>0.030</td>
<td>-0.085</td>
<td>-0.105</td>
</tr>
<tr>
<td>N=500</td>
<td>T=25</td>
<td>-0.114</td>
<td>1.066</td>
<td>-0.136</td>
<td>-0.564</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.113</td>
<td>0.714</td>
<td>-0.036</td>
<td>-0.411</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.075</td>
<td>0.544</td>
<td>-0.018</td>
<td>-0.248</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.055</td>
<td>0.406</td>
<td>-0.011</td>
<td>-0.145</td>
</tr>
<tr>
<td>N=600</td>
<td>T=25</td>
<td>-0.539</td>
<td>0.737</td>
<td>-0.350</td>
<td>-0.819</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.450</td>
<td>0.634</td>
<td>-0.171</td>
<td>-0.651</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.365</td>
<td>0.854</td>
<td>-0.109</td>
<td>-0.440</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.225</td>
<td>0.461</td>
<td>-0.056</td>
<td>-0.242</td>
</tr>
<tr>
<td>N=700</td>
<td>T=25</td>
<td>-0.593</td>
<td>0.756</td>
<td>-0.251</td>
<td>-1.266</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.428</td>
<td>0.677</td>
<td>-0.084</td>
<td>-0.530</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.209</td>
<td>0.494</td>
<td>-0.024</td>
<td>-0.144</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.074</td>
<td>0.282</td>
<td>0.002</td>
<td>-0.072</td>
</tr>
<tr>
<td>N=800</td>
<td>T=25</td>
<td>-0.646</td>
<td>0.794</td>
<td>-0.193</td>
<td>-1.473</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.350</td>
<td>0.630</td>
<td>-0.042</td>
<td>-0.421</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.095</td>
<td>0.350</td>
<td>0.000</td>
<td>-0.048</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.007</td>
<td>0.123</td>
<td>0.011</td>
<td>-0.010</td>
</tr>
</tbody>
</table>

Table 1: Main descriptive statistics for $\epsilon_\theta = \hat{\theta} - \theta$ for $m = \{1, 3, 10\}$
Figure 4: $\hat{\phi}_i - \phi_i$ for $m = \{1, 3, 10\}$ (solid, dash and dash-dot lines respectively). 2,000 simulations, Epanechnikov kernel function.
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D</th>
<th>Median</th>
<th>25% Perc.</th>
<th>75% Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=1</td>
<td>T=25</td>
<td>-0.047</td>
<td>0.107</td>
<td>-0.035</td>
<td>-0.106</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.021</td>
<td>0.062</td>
<td>-0.016</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.011</td>
<td>0.039</td>
<td>-0.010</td>
<td>-0.036</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.004</td>
<td>0.027</td>
<td>-0.003</td>
<td>-0.021</td>
</tr>
<tr>
<td>N=10</td>
<td>T=25</td>
<td>-0.022</td>
<td>0.075</td>
<td>-0.019</td>
<td>-0.068</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.009</td>
<td>0.053</td>
<td>-0.008</td>
<td>-0.041</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.005</td>
<td>0.033</td>
<td>-0.004</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.002</td>
<td>0.023</td>
<td>-0.002</td>
<td>-0.017</td>
</tr>
<tr>
<td>N=25</td>
<td>T=25</td>
<td>-0.019</td>
<td>0.072</td>
<td>-0.017</td>
<td>-0.065</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.010</td>
<td>0.048</td>
<td>-0.008</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.005</td>
<td>0.033</td>
<td>-0.004</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.003</td>
<td>0.023</td>
<td>-0.002</td>
<td>-0.018</td>
</tr>
<tr>
<td>N=50</td>
<td>T=25</td>
<td>-0.019</td>
<td>0.072</td>
<td>-0.017</td>
<td>-0.065</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.009</td>
<td>0.047</td>
<td>-0.008</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.005</td>
<td>0.033</td>
<td>-0.004</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.003</td>
<td>0.023</td>
<td>-0.002</td>
<td>-0.018</td>
</tr>
<tr>
<td>m=3</td>
<td>T=25</td>
<td>-0.048</td>
<td>0.109</td>
<td>-0.038</td>
<td>-0.113</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.024</td>
<td>0.067</td>
<td>-0.020</td>
<td>-0.064</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.013</td>
<td>0.041</td>
<td>-0.010</td>
<td>-0.041</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.009</td>
<td>0.029</td>
<td>-0.007</td>
<td>-0.027</td>
</tr>
<tr>
<td>N=10</td>
<td>T=25</td>
<td>-0.055</td>
<td>0.091</td>
<td>-0.053</td>
<td>-0.108</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.043</td>
<td>0.054</td>
<td>-0.041</td>
<td>-0.077</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.037</td>
<td>0.037</td>
<td>-0.036</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.033</td>
<td>0.027</td>
<td>-0.032</td>
<td>-0.051</td>
</tr>
<tr>
<td>N=25</td>
<td>T=25</td>
<td>-0.049</td>
<td>0.090</td>
<td>-0.050</td>
<td>-0.100</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.039</td>
<td>0.050</td>
<td>-0.037</td>
<td>-0.071</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.033</td>
<td>0.035</td>
<td>-0.032</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.031</td>
<td>0.026</td>
<td>-0.030</td>
<td>-0.048</td>
</tr>
<tr>
<td>N=50</td>
<td>T=25</td>
<td>-0.045</td>
<td>0.077</td>
<td>-0.043</td>
<td>-0.091</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.036</td>
<td>0.048</td>
<td>-0.034</td>
<td>-0.066</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.031</td>
<td>0.034</td>
<td>-0.029</td>
<td>-0.053</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.028</td>
<td>0.025</td>
<td>-0.027</td>
<td>-0.044</td>
</tr>
<tr>
<td>m=10</td>
<td>T=25</td>
<td>-0.044</td>
<td>0.117</td>
<td>-0.037</td>
<td>-0.107</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.024</td>
<td>0.075</td>
<td>-0.018</td>
<td>-0.064</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.013</td>
<td>0.049</td>
<td>-0.010</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.010</td>
<td>0.033</td>
<td>-0.007</td>
<td>-0.029</td>
</tr>
<tr>
<td>N=10</td>
<td>T=25</td>
<td>-0.094</td>
<td>0.158</td>
<td>-0.069</td>
<td>-0.172</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.084</td>
<td>0.108</td>
<td>-0.063</td>
<td>-0.144</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.075</td>
<td>0.083</td>
<td>-0.059</td>
<td>-0.126</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.067</td>
<td>0.070</td>
<td>-0.050</td>
<td>-0.112</td>
</tr>
<tr>
<td>N=25</td>
<td>T=25</td>
<td>-0.090</td>
<td>0.158</td>
<td>-0.064</td>
<td>-0.163</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.072</td>
<td>0.098</td>
<td>-0.051</td>
<td>-0.122</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.056</td>
<td>0.068</td>
<td>-0.039</td>
<td>-0.093</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.044</td>
<td>0.053</td>
<td>-0.028</td>
<td>-0.074</td>
</tr>
<tr>
<td>N=50</td>
<td>T=25</td>
<td>-0.089</td>
<td>0.155</td>
<td>-0.062</td>
<td>-0.157</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.061</td>
<td>0.088</td>
<td>-0.042</td>
<td>-0.099</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.039</td>
<td>0.055</td>
<td>-0.027</td>
<td>-0.065</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.027</td>
<td>0.039</td>
<td>-0.017</td>
<td>-0.042</td>
</tr>
</tbody>
</table>

Table 2: Main descriptive statistics for $\hat{\phi}_i - \phi_i$ for $m = \{1, 3, 10\}$
Figure 5: $\hat{\sigma}_i^2 - \sigma_i^2$ for $m = \{1, 3, 10\}$ (solid, dash and dash-dot lines respectively). 2,000 simulations, Epanechnikov kernel function.
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>S.D</th>
<th>Median</th>
<th>25% Perc.</th>
<th>75% Perc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N=1)</td>
<td>T=25</td>
<td>-0.160</td>
<td>0.258</td>
<td>-0.192</td>
<td>-0.347</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.079</td>
<td>0.191</td>
<td>-0.096</td>
<td>-0.219</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.042</td>
<td>0.140</td>
<td>-0.051</td>
<td>-0.134</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.022</td>
<td>0.099</td>
<td>-0.027</td>
<td>-0.091</td>
</tr>
<tr>
<td>(m=1)</td>
<td>T=25</td>
<td>-0.125</td>
<td>0.270</td>
<td>-0.152</td>
<td>-0.317</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.059</td>
<td>0.218</td>
<td>-0.078</td>
<td>-0.204</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.032</td>
<td>0.140</td>
<td>-0.037</td>
<td>-0.130</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.016</td>
<td>0.100</td>
<td>-0.019</td>
<td>-0.085</td>
</tr>
<tr>
<td>(N=10)</td>
<td>T=25</td>
<td>-0.122</td>
<td>0.266</td>
<td>-0.149</td>
<td>-0.313</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.060</td>
<td>0.195</td>
<td>-0.074</td>
<td>-0.198</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.031</td>
<td>0.140</td>
<td>-0.038</td>
<td>-0.129</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.015</td>
<td>0.099</td>
<td>-0.019</td>
<td>-0.084</td>
</tr>
<tr>
<td>(N=25)</td>
<td>T=25</td>
<td>-0.208</td>
<td>0.263</td>
<td>-0.239</td>
<td>-0.394</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.100</td>
<td>0.191</td>
<td>-0.113</td>
<td>-0.242</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>-0.046</td>
<td>0.135</td>
<td>-0.051</td>
<td>-0.138</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>-0.021</td>
<td>0.100</td>
<td>-0.023</td>
<td>-0.093</td>
</tr>
<tr>
<td>(m=10)</td>
<td>T=25</td>
<td>0.021</td>
<td>0.384</td>
<td>-0.037</td>
<td>-0.224</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>0.043</td>
<td>0.238</td>
<td>0.021</td>
<td>-0.121</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>0.061</td>
<td>0.167</td>
<td>0.052</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>0.070</td>
<td>0.121</td>
<td>0.063</td>
<td>-0.015</td>
</tr>
<tr>
<td>(N=25)</td>
<td>T=25</td>
<td>-0.049</td>
<td>0.328</td>
<td>-0.088</td>
<td>-0.269</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.002</td>
<td>0.213</td>
<td>-0.019</td>
<td>-0.152</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>0.027</td>
<td>0.153</td>
<td>0.019</td>
<td>-0.080</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>0.039</td>
<td>0.111</td>
<td>0.034</td>
<td>-0.038</td>
</tr>
<tr>
<td>(N=50)</td>
<td>T=25</td>
<td>-0.107</td>
<td>0.627</td>
<td>-0.211</td>
<td>-0.392</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>-0.017</td>
<td>0.560</td>
<td>-0.113</td>
<td>-0.236</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>0.042</td>
<td>0.510</td>
<td>-0.039</td>
<td>-0.130</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>0.045</td>
<td>0.426</td>
<td>-0.007</td>
<td>-0.079</td>
</tr>
<tr>
<td>(m=\beta)</td>
<td>T=25</td>
<td>0.483</td>
<td>1.042</td>
<td>0.154</td>
<td>-0.125</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>0.525</td>
<td>0.944</td>
<td>0.205</td>
<td>-0.040</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>0.510</td>
<td>0.873</td>
<td>0.209</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>0.366</td>
<td>0.688</td>
<td>0.157</td>
<td>0.020</td>
</tr>
<tr>
<td>(N=50)</td>
<td>T=25</td>
<td>0.587</td>
<td>1.167</td>
<td>0.174</td>
<td>-0.130</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>0.497</td>
<td>0.992</td>
<td>0.142</td>
<td>-0.068</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>0.299</td>
<td>0.729</td>
<td>0.090</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>0.155</td>
<td>0.424</td>
<td>0.069</td>
<td>-0.021</td>
</tr>
<tr>
<td>(m=10)</td>
<td>T=25</td>
<td>0.664</td>
<td>1.242</td>
<td>0.200</td>
<td>-0.130</td>
</tr>
<tr>
<td></td>
<td>T=50</td>
<td>0.401</td>
<td>0.930</td>
<td>0.080</td>
<td>-0.100</td>
</tr>
<tr>
<td></td>
<td>T=100</td>
<td>0.147</td>
<td>0.530</td>
<td>0.037</td>
<td>-0.073</td>
</tr>
<tr>
<td></td>
<td>T=200</td>
<td>0.056</td>
<td>0.211</td>
<td>0.031</td>
<td>-0.045</td>
</tr>
</tbody>
</table>

Table 3: Main descriptive statistics for \(\hat{\phi}_i - \phi_i\) for \(m = \{1, 3, 10\}\)
Table 4: Weights $\hat{\omega}_j$ and hyperparameters $\hat{\kappa}_{N,T}$. Actual $\omega_j$ and $\kappa$ in italics. Bottom half shows mean deviation w.r.t. actuals.

6 Conclusion

The (MI)xed (DA)ta (S)ampling regression type of models were introduced almost a decade ago by Ghysels et al. (2004a), Ghysels et al. (2004b) and Ghysels et al. (2007) among other. Since then, it has attracted widespread interest of different economic and finance theorist and applied researchers. Its main feature is that it allows the econometrician to regress a dependent variable on a set of covariates that were sampled at higher frequencies without wasting any information. Not only that, it achieves such a goal with a simple, parsimonious and very flexible weighting function that is able to represent a myriad of different shapes.

Nevertheless, this flexibility comes to a cost. In particular, the nature of MIDAS implies that its extension to Panel Data Models is not straightforward and its consequences unforeseeable. Khalaf et al. (2012) was among the firsts that tried to extend MIDAS to a Dynamic Panel Data framework and was quite successful. Inspired by Khalaf et al. (2012) promising results, we decided to evaluate how a Dynamic Heterogeneous Panel Data model would react to MIDAS covariates. In particular, we decided to focus on the Pooled Mean Group estimator introduced by Pesaran et al. (1997) and Pesaran et al. (1999). This model is of particular interest because it allows the researcher to take a middle path when studying long-term responses, instead of the two extreme alternative assumptions. Previously, the mainstream researcher only had two options, either (i) he or she assumed that all the responses were identical amongst groups or (ii) he or she estimated $N$ totally unrelated equations. The Pesaran-Shin-Smith’s model makes a compromise between the two and assumes a shared long-run relationship and
individual-specific convergence rates, which seems a significantly more realistic assumption.

We evaluated the performance of our PMG-MIDAS model by developing and implementing a Monte Carlo simulation study. Unlike Khalaf et al. (2012) we focused just on a slowly decaying Exponential Almon Lag function, which we think is one of the most relevant shapes for applied researchers. Our findings were very promising and may lead to further research on the matter.

Firstly, we confirmed that as it is somewhat usual among non-linear estimation methods, our method depends strongly on a good set of initial values. Therefore, we propose a solution that seems to perform in a satisfactory manner. Secondly, we found that the estimation errors and the size of the bias were scaling down as the size of the data field increased, suggesting that the estimators remained consistent even under this modification. Thirdly, we observed that MIDAS seems to induce a higher than the original PMG downward bias on the estimates and that this bias is an increasing function of $m$. Furthermore, the skewness in the distribution of estimation errors seems to be exacerbated whenever $N > 1$ and heterogeneity comes into play, although the increased precision seems to compensate part of it.

A deeper understanding of the nature of the interaction between the long-run relationship, the short-run adjustment rates and the weight function is called for and could prove to be a promising line of research. We observed that our estimated weight functions tended to be flatter than expected, placing less weight in the recent observations, while weighting more the older ones. We suspect that some of the information that ideally should have been captured by the adjustment coefficients and the long-run relationship estimates is being captured by the $\omega_j$ linked to the highest high-frequency lags. Until further research is made, this will remain in the speculation realm.

This paper pretended to work as an additional stepping stone on the path started by Khalaf et al. (2012) on extending MIDAS to Dynamic Panel Data models. As such, further research is still needed and several questions remain to be answered. For example, is MIDAS worth the trouble for a model like this one? Is the cure worse than the disease? In other words, is the bias introduced by the MIDAS covariates less damaging than the effects of wasting information on ad-hoc pre-filtering methods? Is there a way to reduce the undesired effects of the aforementioned interaction? Moreover, bearing in mind that PMG allows I(1) variables, what would happen if we were to try such extension under MIDAS covariates? In conclusion, these new questions and others are on the open and will require some future attention.
References


A PMG with MIDAS covariates

A.1 PMG with MIDAS covariates - Concentrated Log-likelihood and previous definitions

\[
\ell_T(\psi) = -\frac{T}{2} \sum_{i=1}^{N} \ln 2\pi \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \tilde{\varepsilon}_i(\theta, \kappa)'H_i\tilde{\varepsilon}_i(\theta, \kappa) \tag{A.1}
\]

where

\[
\tilde{\varepsilon}_i(\theta, \kappa) = \Delta y_i - \phi_i \xi_i(\theta, \kappa) \text{ is the resulting error after concentrating the log-likelihood,}
\]

\[
\psi = \{\theta, \phi_1...\phi_N, \sigma_1^2...\sigma_N^2, \omega(\kappa_1, \kappa_2)\}, \omega = \{\omega_0,...,\omega_{M-1}\} \text{ and } \omega_j = B(\kappa_1, \kappa_2, j)
\]

\[
\varepsilon_i(\theta, \kappa) = y_{i,-1} - \tilde{X}_i(\kappa)\theta,
\]

\[
\tilde{x}(\kappa)_{i,t} = \sum_{j=0}^{m-1} B(\kappa_1, \kappa_2, j)x_{i,t-j}^{(m)}
\]

Let us maximize \(\ell_T\) w.r.t. \(\omega_j\).

Also, for clarity of exposition let

\[
\frac{\delta \tilde{X}_i}{\delta \omega_j} = X^{(m)}_{i,-\frac{j}{m}} \quad \tilde{X}^{(m)*}_{i,-\frac{j}{m}} = \tilde{X}_i - \omega_j X^{(m)}_{i,-\frac{j}{m}}
\]

where \(X^{(m)}_{i,-\frac{j}{m}} = \{x^{(m)}_{i,1-\frac{j}{m}}, ..., x^{(m)}_{i,T-\frac{j}{m}}\}'\) and \(Q_i = \tilde{\varepsilon}_i(\theta, \kappa)'H_i\tilde{\varepsilon}_i(\theta, \kappa)\).
A.2 PMG with MIDAS covariates - Derivations

A.2.1 Case for \( \hat{\theta} \)

\[
\frac{\delta \ell_T}{\delta \theta} = -\frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \delta Q_i = 0
\]

\[
= -\frac{1}{2} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left\{ \frac{\delta \bar{\varepsilon}_i}{\delta \theta} \right\}' \{2H_i \tilde{\varepsilon}_i \} = 0
\]

\[
= -\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left\{ \frac{\delta \bar{\varepsilon}_i}{\delta \xi_i} \frac{\delta \xi_i}{\delta \theta} \right\} H_i \tilde{\varepsilon}_i
\]

\[
= -\sum_{i=1}^{N} \frac{\phi_i}{\sigma_i^2} \bar{X}'H_i \varepsilon_i = 0
\]

\[
= -\sum_{i=1}^{N} \frac{\phi_i}{\sigma_i^2} \bar{X}'H_i (\Delta y_i - \phi_i \xi_i) = 0
\]

\[
= -\sum_{i=1}^{N} \frac{\phi_i}{\sigma_i^2} \bar{X}'H_i \Delta y_i + \sum_{i=1}^{N} \frac{\phi_i^2}{\sigma_i^2} \bar{X}'H_i \left( y_{i,-1} - \bar{X}_i \theta \right) = 0
\]

\[
= -\sum_{i=1}^{N} \frac{\phi_i}{\sigma_i^2} \bar{X}'H_i (\Delta y_i - \phi_i y_{i,-1}) - \theta \sum_{i=1}^{N} \frac{\phi_i^2}{\sigma_i^2} \bar{X}'H_i \bar{X}_i = 0
\]

\[
\Rightarrow \hat{\theta} = - \left\{ \sum_{i=1}^{N} \frac{\phi_i^2}{\sigma_i^2} \bar{X}'H_i \bar{X}_i \right\}^{-1} \left\{ \sum_{i=1}^{N} \frac{\phi_i}{\sigma_i^2} \bar{X}'H_i (\Delta y_i - \hat{\phi}_i y_{i,-1}) \right\} \quad (A.3)
\]

A.2.2 Case for \( \hat{\phi}_i \)

\[
\frac{\delta \ell_T}{\delta \phi_i} = -\frac{1}{2\sigma_i^2} \delta Q_i = 0
\]

\[
= -\frac{1}{2\sigma_i^2} \left\{ \frac{\delta \bar{\varepsilon}_i}{\delta \phi_i} \right\}' \left\{ \delta Q_i \right\} = 0
\]

\[
= \frac{1}{\sigma_i^2} \xi_i' H_i \bar{\varepsilon}_i = 0
\]

\[
= \xi_i' H_i \Delta y_i - \phi_i \xi_i' H_i \xi_i = 0
\]

\[
\Rightarrow \hat{\phi}_i = \left\{ \xi_i' H_i \xi_i \right\}^{-1} \left\{ \xi_i' H_i \Delta y_i \right\} \quad (A.5)
\]

for \( i = 1, \ldots, N. \)
A.2.3 Case for $\delta_i^2$

$$\frac{\delta \ell_T}{\delta \sigma_i^2} = -\frac{T}{2\sigma_i^2} + \frac{1}{2} \frac{1}{\sigma_i^2} \hat{\varepsilon}_i^T \mathbf{H}_i \hat{\varepsilon}_i = 0 \quad (A.6)$$

$$\Rightarrow \delta_i^2 = \frac{1}{T} \hat{\varepsilon}_i^T \mathbf{H}_i \hat{\varepsilon}_i \quad (A.7)$$

for $i = 1, ..., N$.

A.2.4 Case for $\hat{\omega}_j$. The U-MIDAS approach

$$\frac{\delta \ell_T}{\delta \omega_j} = -\frac{1}{2} \sum_{i=1}^{N} \left\{ \frac{\delta \varepsilon_i}{\delta \hat{X}_i} \frac{\delta \hat{X}_i}{\delta \omega_j} \right\}' \mathbf{H}_i \varepsilon_i = 0$$

$$= -\frac{1}{\sigma_j^2} \sum_{i=1}^{N} \left\{ \frac{\delta \hat{X}_i}{\delta \omega_j} \left( \frac{\delta \hat{X}_i}{\delta \omega_j} \right)' \right\} \mathbf{H}_i \varepsilon_i = 0$$

$$= -\frac{1}{\sigma_j^2} \sum_{i=1}^{N} \left\{ \frac{\delta \hat{X}_i}{\delta \omega_j} \left( \frac{\delta \hat{X}_i}{\delta \omega_j} \right)' \right\} \mathbf{H}_i = 0$$

$$= -\frac{1}{\sigma_j^2} \sum_{i=1}^{N} \left\{ \frac{\delta \hat{X}_i}{\delta \omega_j} \left( \frac{\delta \hat{X}_i}{\delta \omega_j} \right)' \right\} \mathbf{H}_i \Delta \mathbf{y}_i - \phi_i \left( \mathbf{y}_{i-1} - \hat{\mathbf{X}}_i \hat{\theta} \right) = 0$$

$$= -\frac{1}{\sigma_j^2} \sum_{i=1}^{N} \left\{ \frac{\delta \hat{X}_i}{\delta \omega_j} \left( \frac{\delta \hat{X}_i}{\delta \omega_j} \right)' \right\} \mathbf{H}_i \left( \Delta \mathbf{y}_i - \phi_i \mathbf{y}_{i-1} \right) - \frac{1}{\sigma_j^2} \sum_{i=1}^{N} \left\{ \frac{\delta \hat{X}_i}{\delta \omega_j} \left( \frac{\delta \hat{X}_i}{\delta \omega_j} \right)' \right\} \mathbf{H}_i \hat{X}_i \hat{\theta} = 0$$

$$= -\frac{1}{\sigma_j^2} \sum_{i=1}^{N} \left\{ \frac{\delta \hat{X}_i}{\delta \omega_j} \left( \frac{\delta \hat{X}_i}{\delta \omega_j} \right)' \right\} \mathbf{H}_i \left( \Delta \mathbf{y}_i - \phi_i \mathbf{y}_{i-1} - \phi_i \hat{X}_i^{(m)*} \hat{\theta} \right) - \omega_j \sum_{i=1}^{N} \left\{ \frac{\delta \hat{X}_i}{\delta \omega_j} \left( \frac{\delta \hat{X}_i}{\delta \omega_j} \right)' \right\} \mathbf{H}_i \hat{X}_i^{(m)} \hat{\theta} = 0 \quad (A.8)$$

$$\hat{\omega}_j = -\left\{ \sum_{i=1}^{N} \frac{\delta \hat{X}_i^{(m)*}}{\delta \omega_j} \mathbf{H}_i \hat{X}_i^{(m)} \hat{\theta} \right\}^{-1} \left\{ \sum_{i=1}^{N} \frac{\delta \hat{X}_i^{(m)*}}{\delta \omega_j} \mathbf{H}_i \hat{\Psi}_i \right\} \quad (A.9)$$

where $\hat{\Psi}_i = [\Delta \mathbf{y}_i - \phi_i \left( \mathbf{y}_{i-1} - \phi_i \hat{X}_i^{(m)*} \hat{\theta} \right)]$ for $j = 0, ..., m - 1$. 

35